Preprint ANL/MCS-P1588-0309.

Exploiting Weather Forecast Information in the Operation of Integrated Energy Systems

Victor M. Zavala,[†§*] **Emil M. Constantinescu,**[†]
Theodore Krause,[§] **and Mihai Anitescu**[†]

[†]Mathematics and Computer Science Division

[§]Chemical Sciences and Engineering Division

Argonne National Laboratory, Argonne, IL, 60439 United States

Abstract

In this work, we establish an on-line optimization framework to exploit detailed weather forecast information in the operation of integrated energy systems. We first discuss how the use of traditional *reactive* operation strategies that neglect the future evolution of the ambient conditions can result into high operating costs. To overcome this problem, we introduce a supervisory dynamic optimization strategy that can lead to more *proactive* and cost-effective operations. The strategy is based on the solution of a receding-horizon stochastic dynamic optimization problem. This permits the incorporation of economic objectives, statistical forecast information, and operational constraints in a systematic manner. To obtain the weather forecast information, we employ a state-of-the-art forecasting model initialized with real meteorological data. The statistical ambient information is obtained from a set of realizations generated by the weather model executed in an operational setting. We present proof-of-concept simulation studies to demonstrate that the proposed framework can lead to significant savings in operating costs.

1 Introduction

During the past several years, strong socioeconomic pressures have forced diverse industrial sectors to reassess the efficiency of current energy production and consumption facilities. In particular, increasing fossil fuel prices and carbon emission penalties will require the consideration of more efficient designs able to accommodate multiple energy sources and operating strategies able to maximize the utilization of such resources in a cost-optimal manner. The design and operation of these integrated energy systems

*Address all correspondence to this author. Argonne National Laboratory, Mathematics and Computer Science Division, 9700 S. Cass Avenue, Argonne, IL 60439, USA. email:vzavala@mcs.anl.gov

are complex tasks because of multiple physical phenomena arising in different units and because of their strong dependency on *exogenous* disturbances such as the ambient conditions, time-varying demands, and time-varying fuel and electricity prices. In this context, the availability of powerful simulation technologies able to predict and assess the performance of these systems under a wide variety of operating environments will become increasingly important.

Rigorous simulation models for myriad energy systems have been developed over the past few years and have been used extensively for off-line design and retrofitting tasks. The availability of these models has led to more systematic practices and, consequently, to more cost-effective systems. Examples of available simulation software are EnergyPlus and ADVISOR, developed at the NREL Laboratory to simulate the performance of building and hybrid vehicle systems [6; 14]; GCTool and PSAT, developed at Argonne National Laboratory to simulate hybrid vehicles and power train systems [8; 12]; and TRNSYS, developed at the University of Wisconsin to simulate a wide range of hybrid energy and building systems [11]. With this simulation technology at hand, several natural questions arise: Can we use these powerful models on-line to optimize the operation of energy systems? How can we integrate these simulators with available optimization technology? Can we handle highly uncertain and dynamic disturbances effectively?

The operation of industrial systems is normally decomposed in a hierarchical manner, as sketched in Figure 1. The high level is normally known as the *supervisory* or set-point optimization layer. At this level, the set-points are adjusted in order to optimize the system economic performance. In the context of energy systems, this level is also known as the centralized energy management system. The lower level is the *regulatory* control level that uses available actuators to track the set-points dictated by the supervisory level. Most state-of-the-art energy simulation packages provide closed-loop simulation capabilities that can be used to design and test different operating strategies. These capabilities are based mostly on proportional-integral-derivative (PID) and *logic-based* controllers. PID controllers are limited to regulation tasks, while logic-based controllers can be used for both regulation and set-point optimization. Logic-based controllers consist of a decision-making structure or tree designed to determine the controls as a function of the current outputs and exogenous disturbances. The decision-making structure and threshold values are tuned off-line using a simulation model in order to obtain a desired performance [21; 23]. The application of logic-based strategies is intuitive and can provide satisfactory results for regulatory control and basic optimization tasks. However, performing high-level supervisory optimization tasks can become cumbersome in large and tightly interconnected systems. The reason is that, as the amount of information and number of decision variables grow, the logic structure becomes more and more complex and tuning the associated threshold values becomes time-consuming and impractical. In addition, once the logic-based controller is tuned by using the simulation model on a variety of scenarios, the model will no longer be used on-line. Consequently, it is difficult to guarantee adequate performance under unexpected conditions, and retuning might be necessary. Moreover, handling economic objectives and operational limits can become complicated under this framework.

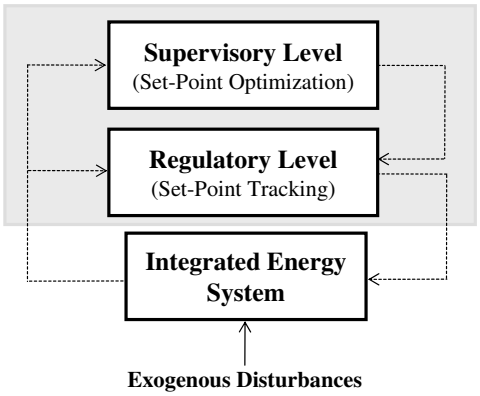


Figure 1: Schematic representation of operation hierarchy in industrial systems.

An alternative supervisory strategy is closed-loop real-time optimization (RTO) [15]. The idea is to use a *steady-state* rigorous model of the system and couple this to a large-scale optimization solver. The optimizer will determine the optimal output set-points that maximize the system profit using the current information of the exogenous disturbances. Note that, since the set-points need to satisfy the operational limits in the real system, a rigorous model is needed. The RTO output set-points are passed to a set of lower-level controllers that bring the system to the optimal steady-state. Once this is accomplished, the set-points are recomputed by RTO using the updated disturbance information. An advantage of this framework is that economic objectives and operational limits can be handled directly by the optimizer in a systematic manner. In addition, the rigorous model is always used and adapted on-line. Consequently, tuning tasks are significantly reduced. More important, it is always possible to guarantee that the system is at an optimal operating point. As can be seen, RTO offers significant advantages over logic-based strategies, especially in highly complex systems. Note also that, since time-varying factors such as energy prices or weather conditions can be seen as dynamic disturbances that move the system away from the optimal operating point, RTO can be seen as a closed-loop optimizer that rejects these disturbances and keeps track of the maximum system performance. As a result of these desirable economic adaptation features, RTO capabilities are now widely used in conjunction with chemical process simulators such as AspenPlus© and ROMeo©. This technology has generated millions of dollars in annual savings in diverse sectors of the chemical industry [25]. Nevertheless, an important limitation of RTO and logic-based strategies is that they are entirely *reactive*, in the sense that only current disturbance information is used to make decisions. This feature can limit their ability to handle highly dynamic disturbances efficiently. For instance, if the set-points are updated at a higher frequency than the controller settling time, erratic performance and instabilities can arise. These are important limitations in integrated energy systems where performance strongly depends on transient disturbances.

In this work, we introduce a dynamic real-time optimization (D-RTO) strategy to perform economic supervisory decisions in integrated energy systems. The idea is in principle

similar to RTO, but the key difference is that a rigorous *dynamic* model is used to compute future dynamic set-point trajectories [9]. In this context, the D-RTO strategy is a special variant of Nonlinear Model Predictive Control (NMPC). We show that this strategy permits a consistent handling of highly dynamic disturbances and can directly incorporate forecast information. This adds *proactiveness* to the control actions and reduces costs. In particular, we show that incorporating weather forecast information can translate into significant savings in energy systems. Nevertheless, we demonstrate that using accurate forecasts and uncertainty information is critical to achieve a reliable system performance. To obtain this information, we first propose to construct data-based autoregressive models using a Gaussian process modeling technique. While this strategy is useful to obtain quick estimates of certain weather conditions and related uncertainty information, it is limited to short-term forecasts and can give rise to inconsistent uncertainty bounds. Therefore, we explore the potential of using detailed weather models. From an operational point of view, these models are attractive because they can provide comprehensive information such as spatiotemporal fields of ambient temperature, solar radiation, and humidity. This information can be fully exploited by the rigorous model embedded within the D-RTO strategy. In current weather models, however, the uncertainty information is limited or in forms that are inconsistent with existing optimization technology. We therefore develop simplified uncertainty models for ambient variables that rely on model dynamics and require only few empirical assumptions. The weather model is driven in an operational setting with real data and thus provides realistic and attainable estimates on the uncertainty found in the meteorological fields. We argue that connecting these powerful weather prediction models with modeling and optimization capabilities has the potential of achieving unprecedented energy utilization efficiencies and cost reductions in diverse industrial and residential sectors. We present closed-loop simulation studies on a building system. We demonstrate that the proposed forecast-based framework is able to reduce operating costs by exploiting the use of weather forecast information, and by reacting in a timely manner to slow dynamic trends. We claim that the features can bring significant savings in more complex energy systems such as hybrid systems, power plants, energy parks, and integrated building systems.

The paper is organized as follows. In the next section, we establish the stochastic optimization framework and analyze the impact of adding forecast information in the economics of integrated energy systems. In addition, we present a strategy to obtain empirical forecast information using a Gaussian process modeling technique. In Section 3, we present techniques to obtain forecast uncertainty information directly from the WRF weather prediction system. In particular, we discuss the extraction of uncertainty information for spatiotemporal fields of the temperature and of the ground solar radiation through an ensemble-based approach. We close the paper with general conclusions and directions for future work.

2 Optimization Framework

In this section, we derive the basic components of the closed-loop dynamic optimization (D-RTO) framework, explain its advantages over steady-state optimization (RTO), and discuss extensions to consider stochastic disturbance information. We then illustrate the economic impact of folding weather forecast information into the operation of a photovoltaic-hydrogen hybrid energy system and a building system. We illustrate the use of the Gaussian process modeling technique to obtain on-line forecast information, and we explain how this information can be connected to the optimization framework.

2.1 Stochastic Dynamic Optimization

We begin by considering a differential-algebraic equation (DAE) model of the form

$$\frac{dz}{d\tau} = \mathbf{f}\left(z(\tau), y(\tau), u(\tau), \chi(\tau)\right) \tag{1a}$$

$$0 = \mathbf{g}(z(\tau), y(\tau), u(\tau), \chi(\tau)) \tag{1b}$$

$$z(0) = x_k, \tag{1c}$$

where τ is the *model* time dimension and t_k is the current time in the *real* system. Variables $z(\tau)$ are differential states, $y(\tau)$ are algebraic states, $u(\tau)$ are the controls or manipulated variables, and $\chi(\tau)$ are the exogenous disturbances. In this context, the term *exogenous* refers to the fact that the disturbances are not affected by the system variables (e.g., energy prices). The differential equations (1a) represent conservation equations (energy, mass, and momentum), while the algebraic equations (1b) represent consistency conditions and expressions to calculate physicochemical properties. The initial conditions at time t_k are given by the current state of the system x_k. Starting from this state and using a set of future control and disturbance trajectories, we can predict the evolution of the system. With these predictive capabilities, we can formulate a dynamic optimization problem of the form

$$\min_{u(\tau)} \int_{t_k}^{t_k+T} \varphi(z(\tau), y(\tau), u(\tau), \chi(\tau))d\tau \tag{2a}$$

$$\frac{dz}{d\tau} = \mathbf{f}\left(z(\tau), y(\tau), u(\tau), \chi(\tau)\right) \tag{2b}$$

$$0 = \mathbf{g}(z(\tau), y(\tau), u(\tau), \chi(\tau)) \tag{2c}$$

$$0 \geq \mathbf{h}(z(\tau), y(\tau), u(\tau), \chi(\tau)) \tag{2d}$$

$$z(t_k) = x_k, \quad \tau \in [t_k, t_k + T], \tag{2e}$$

where T is the length of the prediction or forecast horizon. The objective function (2a) represents the system operational costs accumulated over the future horizon (e.g., heating/cooling utilities). The inequality constraints (2d) are used to represent operational limits (e.g., temperature, pressure, and voltage levels). The dynamic optimization problem is infinite-dimensional because it depends on time, which is a continuous parameter.

This problem can be approximated by a finite-dimensional nonlinear programming (NLP) problem through discretization techniques. Note that any partial differential equation (PDE) model can also be represented in DAE form through discretization along the space dimensions.

To formulate the optimization problem at time t_k, we assume that we know the value of the current disturbance $\chi(t_k)$ but that the future disturbance trajectory is unknown. Nevertheless, we assume that the future disturbances belong to an uncertain space $\chi(\tau) \in \Omega_k, \tau \in [t_k, t_k + T]$ that can be approximated. To do so, we use *past* disturbance information $\chi(\tau), \tau \in [t_k - N, t_k]$ and a suitable forecast model. The forecast model can be either empirical or physics-based. In any case, we can assume that the model provides a predictive mean $\bar{\chi}(\tau)$ and that the forecast errors follow a normal or Gaussian distribution such that $\chi(\tau) = \mathcal{N}(\bar{\chi}(\tau), \mathbf{V}(\tau))$, where $\mathbf{V}(\tau)$ is the covariance matrix. The uncertain space adopts an ellipsoidal form

$$\Omega_k := \{\, z \,|\, (z - \bar{\chi}(\tau))^T \mathbf{V}^{-1}(\tau)(z - \bar{\chi}(\tau)) \leq \alpha\}, \tag{3}$$

where α represents an appropriate confidence level. This uncertainty region is sketched in Figure 2. Under these assumptions, all that is needed to represent the uncertain space is the predictive mean and the covariance matrix. However, the proposed structure of the uncertainty space is a modeling assumption and hence might not be accurate. Nevertheless, from a practical point of view, *what we seek is that the approximate space can encapsulate the true disturbance realization and that it has a physically meaningful structure.*

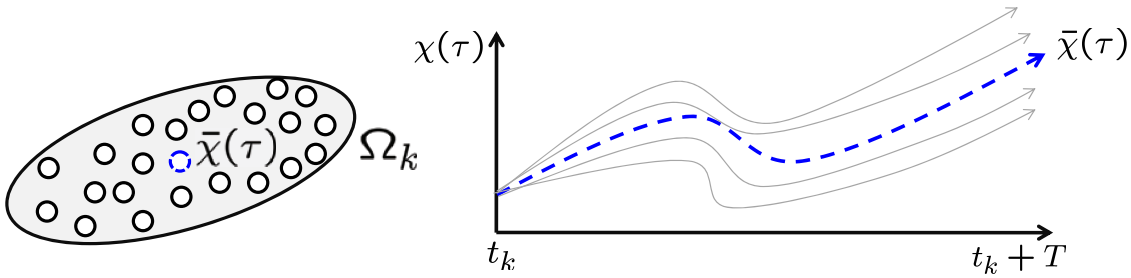


Figure 2: Schematic representation of ellipsoidal uncertainty region.

To exploit the entire statistical information at hand, we formulate a stochastic dynamic optimization problem of the form

$$\min_{u(\tau)} \mathop{\mathbf{E}}_{\chi(\tau)\in\Omega_k} \left[\int_{t_k}^{t_k+T} \varphi(z(\tau), y(\tau), u(\tau), \chi(\tau)) d\tau \right] \tag{4a}$$

$$\frac{dz}{d\tau} = \mathbf{f}\left(z(\tau), y(\tau), u(\tau), \chi(\tau)\right) \tag{4b}$$

$$0 = \mathbf{g}(z(\tau), y(\tau), u(\tau), \chi(\tau)) \tag{4c}$$

$$0 \geq \mathbf{h}(z(\tau), y(\tau), u(\tau), \chi(\tau)) \tag{4d}$$

$$z(t_k) = x_k, \quad \tau \in [t_k, t_k + T], \quad \chi(\tau) \in \Omega_k, \tag{4e}$$

where symbol $\mathbf{E}[\cdot]$ denotes the expectation operator. From the solution of this problem, we obtain the state and control trajectories $z^*(\tau), y^*(\tau), u^*(\tau), \tau \in [t_k, t_k + T]$ that we send to a lower-level controller as set-points. The controller will try to keep the system at the recommended target. At the next time step t_{k+1}, we obtain the updated state of the system x_{k+1} and the updated forecast disturbance information Ω_{k+1} that we use to solve the next stochastic problem (4). In this way, feedback is introduced. The resulting *closed-loop* D-RTO strategy is as follows.

1. **Obtain current state and forecast:** At time t_k, obtain current state x_k, disturbance mean $\bar{\chi}(\tau)$ and associated covariance matrix $\mathbf{V}(\tau), \tau \in [t_k, t_k + T]$.

2. **Compute set-points:** Solve stochastic optimization problem (4). Send optimal set-points $z^*(\tau), y^*(\tau), u^*(\tau), \tau \in [t_k, t_k + T]$ to low-level control layer.

3. **Update:** At $t_k + \Delta$, set $k \leftarrow k + 1$, and repeat process.

Here, Δ is the set-point update period. In Figure 3, we sketch this conceptual closed-loop optimization framework and its interaction with the low-level control layer and the forecasting capability. Note that, in principle, we can simplify the stochastic problem by using only the predictive mean $\bar{\chi}(\tau)$ without taking uncertainty into account. This is equivalent to solving the deterministic optimization problem

$$\min_{u(\tau)} \int_{t_k}^{t_k+T} \varphi(z(\tau), y(\tau), u(\tau), \mathbf{E}[\chi(\tau)])d\tau \tag{5a}$$

$$\frac{dz}{d\tau} = \mathbf{f}\left(z(\tau), y(\tau), u(\tau), \mathbf{E}[\chi(\tau)]\right) \tag{5b}$$

$$0 = \mathbf{g}(z(\tau), y(\tau), u(\tau), \mathbf{E}[\chi(\tau)]) \tag{5c}$$

$$0 \geq \mathbf{h}(z(\tau), y(\tau), u(\tau), \mathbf{E}[\chi(\tau)]) \tag{5d}$$

$$z(t_k) = x_k, \quad \tau \in [t_k, t_k + T], \tag{5e}$$

where $\bar{\chi}(\tau) = \mathbf{E}[\chi(\tau)]$. With this approach, however, we cannot guarantee satisfaction of the operational constraints. Note also that, in the presence of uncertainty, the cost function becomes a probability distribution because it depends on all the possible realizations of the disturbances. Therefore, optimizing a single instance of the cost function is meaningless. These inconsistencies are avoided by considering the stochastic formulation (4).

Note, however, that in problem (4) we have assumed that the mean of the objective distribution is an adequate measure of the performance of the system. However, this need not be the case. For instance, we could also choose the mean-risk approach of Markowitz where we seek to minimize simultaneously the mean and the variance of the cost distribution. In this stochastic optimization framework, the structure of the cost function becomes a design task because it is entirely problem dependent. Note also that the proposed stochastic problem assumes that no recourse exists in the future, as in a dynamic programming approach. This assumption is necessary in order to handle large-scale systems and inequality constraints, which becomes impractical in a dynamic programming setting.

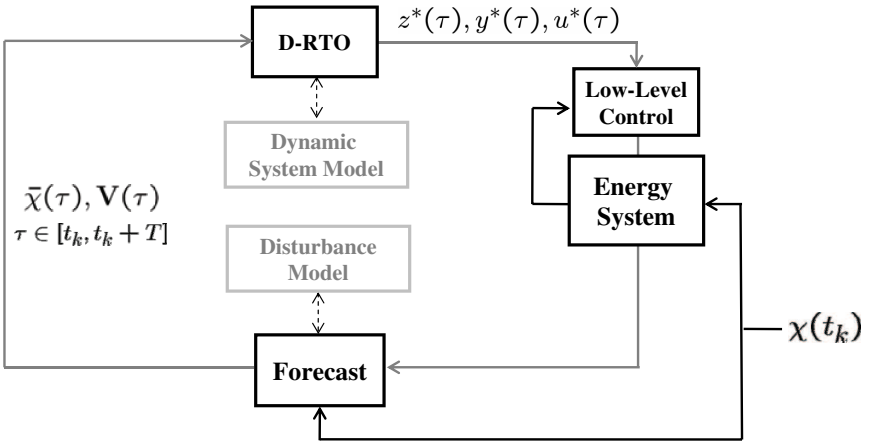


Figure 3: Structure of closed-loop stochastic optimization framework.

The stochastic dynamic formulation is far more computationally demanding than the deterministic dynamic optimization formulation. The reason is that the uncertain space Ω_k is also infinite dimensional. To solve the stochastic optimization problem, we use a sample-average approximation (SAA) approach. The idea is to obtain independent samples from the disturbance distribution to obtain a N_s realizations $\{\chi_1(\tau), \chi_2(\tau), ..., \chi_{N_s}(\tau)\}$. The samples are illustrated in Figure 2. With this, the approximate stochastic problem becomes

$$\min_{u(\tau)} \frac{1}{N_s} \sum_{j=1}^{N_s} \left[\int_{t_k}^{t_k+T} \varphi(z_j(\tau), y_j(\tau), u(\tau), \chi_j(\tau)) d\tau \right] \tag{6a}$$

$$\frac{dz_j}{d\tau} = \mathbf{f}\left(z_j(\tau), y_j(\tau), u(\tau), \chi_j(\tau)\right) \tag{6b}$$

$$0 = \mathbf{g}(z_j(\tau), y_j(\tau), u(\tau), \chi_j(\tau)) \tag{6c}$$

$$0 \geq \mathbf{h}(z_j(\tau), y_j(\tau), u(\tau), \chi_j(\tau)) \tag{6d}$$

$$z_j(t_k) = x_k, \quad \tau \in [t_k, t_k + T], \quad j = 1, ..., N_s. \tag{6e}$$

In this formulation, all the variables become a function of the particular disturbance realization except the controls, which are decision variables. One of the key advantages of the SAA approach is that it is straightforward to implement. Moreover, it is particularly suitable for large-scale systems because it gives rise to highly structured problems [13]. The theoretical properties of the SAA approach have been widely studied in the context of nonlinear programming. For instance, it has been shown that solutions of the SAA problem converge at an exponential rate to the solution of the stochastic counterpart [19]. Although no formal convergence results exist in the context of infinite-dimensional dynamic optimization problems, we can expect that the available convergence guarantees can be used under mild assumptions. Note also that, in the SAA approach, the forecast capability can send the disturbance samples directly to the D-RTO component instead of

the full covariance matrix. This can be useful in large-scale systems because the forecast capability can run in a centralized manner (consider a large weather model) and send the disturbance information to multiple D-RTO agents running on smaller, dedicated machines.

If we shrink the prediction horizon of the stochastic dynamic optimization problem (2) to zero, $T \to 0$, we recover the steady-state RTO problem:

$$\begin{aligned} \min_{u} \quad & \varphi(z, y, u, \chi(t_k)) && \text{(7a)} \\ 0 \;&=\; \mathbf{f}\left(z, y, u, \chi(t_k)\right) && \text{(7b)} \\ 0 \;&=\; \mathbf{g}(z, y, u, \chi(t_k)) && \text{(7c)} \\ 0 \;&\geq\; \mathbf{h}(z, y, u, \chi(t_k)). && \text{(7d)} \end{aligned}$$

This strategy finds the steady-state economic optimal operating point based only on the current disturbances, $\chi(t_k)$, which are known. With this, we no longer rely on any forecasting mechanism, and the problem reduces to a finite-dimensional NLP problem that is far less computationally expensive. While formulations (5) and (7) seem to have computational advantages, in the next section we show that strong economic penalties can be incurred by making these simplifications.

2.2 Economic Impact of Forecasting

In this section, we discuss some of the advantages of folding forecast information in operations. To do so, we present closed-loop D-RTO simulation studies on a building system. Our objective is to illustrate how the use of forecast information can add proactiveness to the D-RTO strategy and how this translates into lower operating costs.

Commercial buildings are energy-intensive facilities where considerable cost savings can be realized through optimal operating strategies. As an example, researchers have found that the thermal mass of a building can be used for temporal energy storage [2]. With this, one can optimize the temperature set-points trajectories during the day to shift the heating and cooling electricity demands to off-peak hours and thus reduce costs. For instance, a cooling strategy that has been used in commercial facilities consists in cooling down the building as much as possible at night when electricity is cheaper so as to reduce the amount of cooling needed during the day when electricity is more expensive [3]. Since the thermal response of the building can be slow (order of hours), this can be exploited to reduce the on-peak electricity demand the next day. However, we point out that an issue might arise in the implementation of these peak-shifting strategies: namely, the optimal timing at which it is decided to start the cooling at night directly depends on the ambient temperature expected the next day. In addition, special care needs to be taken to stay within the thermal comfort zone at all times. Motivated by these factors, in this case study we analyze the effect of adding forecast temperature information in a D-RTO strategy. We demonstrate that the D-RTO strategy can be easily generalized to consider simultaneous peak-shifting for both cooling and heating during the entire year and to consider multiple energy sources.

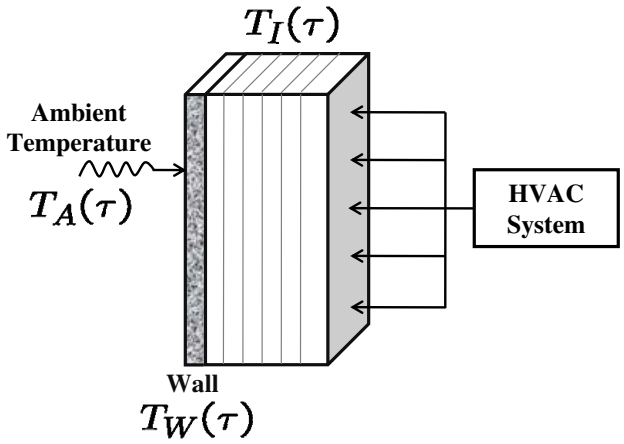


Figure 4: Schematic representation of building integration with heating, ventilation, and air-conditioning (HVAC) system.

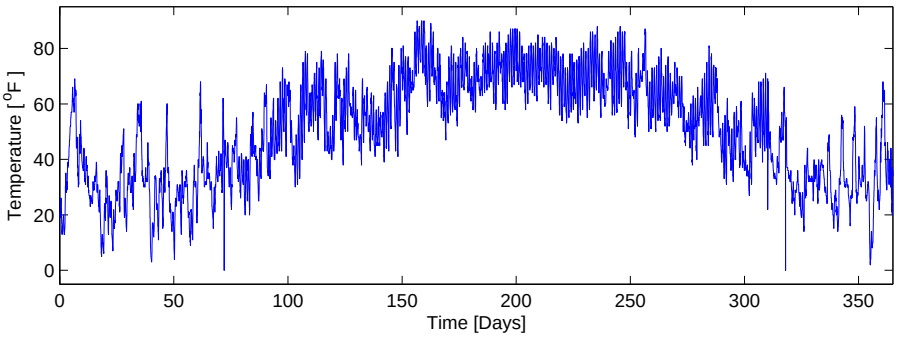


Figure 5: Ambient temperature in Pittsburgh PA, 2006.

The building system under consideration is sketched in Figure 4. We assume a total volume of 10,000 m^3 and a total surface area of 3,500 m^2. The building is equipped with a gas furnace, an electric heater, and an electric cooling system. Ambient temperature data at position 40 30'N/80 13'W in the Pittsburgh, PA, area for year 2006 were obtained from the National Weather Service Office [16]. The temperature profile is presented in Figure 5. Note that strong temperature variations arise at different time scales (daily and seasonal). The variability is stronger during the winter. The dynamic response of the building internal temperature is modeled by an ordinary differential equation; the building wall is modeled by a second-order PDE that accounts for conductive effects along the wall. The ambient temperature enters the model through a Neumann boundary condition at the wall external face. The basic heat-transfer model structure has been obtained from [4]. To analyze the effect of adding forecast information of the ambient temperature, we follow the approach described in the previous section. We first solve an open-loop dynamic optimization problem with perfect forecast information and a prediction horizon of one

year. The optimization problem has the following form

$$\min_{\varphi_c^{elec}(\tau),\varphi_h^{gas}(\tau),\varphi_h^{elec}(\tau)} \quad \int_{t_k}^{t_k+T} \left[C_{elec}(\tau)\varphi_c^{elec}(\tau) + C_{elec}(\tau)\varphi_h^{elec}(\tau) + C_{gas}\varphi_h^{gas}(\tau)\right] d\tau$$

$$C_I \cdot \frac{\partial T_I}{\partial \tau} = \varphi_h^{gas}(\tau) + \varphi_h^{elec}(\tau) - \varphi_c^{elec}(\tau) - S \cdot \alpha' \cdot (T_I(\tau) - T_W(\tau,0)) \quad (8a)$$

$$\frac{\partial T_W}{\partial \tau} = \beta \cdot \frac{\partial^2 T_W}{\partial x^2} \quad (8b)$$

$$0 = \alpha' \left(T_I(\tau) - T_W(\tau,0)\right) + \mathbf{k} \cdot \left.\frac{\partial T_W}{\partial x}\right|_{(\tau,0)} \quad (8c)$$

$$0 = \alpha'' \left(T_W(\tau,L) - T_A(\tau)\right) + \mathbf{k} \cdot \left.\frac{\partial T_W}{\partial x}\right|_{(\tau,L)} \quad (8d)$$

$$T_I^{min} \le T_I(\tau) \le T_I^{max} \quad (8e)$$

$$T_I(0) = T_I^k \quad (8f)$$

$$T_W(0,x) = T_W^k(x), \quad (8g)$$

where $T_A(\tau)$ is the ambient temperature, $T_I(\tau)$ is the internal temperature, and $T_W(\tau,x)$ is the wall temperature (all of them in ^{o}C). The controls are the gas heating power $\varphi_h^{gas}(\tau)$, the electric heating power $\varphi_h^{elect}(\tau)$, and the electric cooling power $\varphi_c^{elec}(\tau)$ (all of them in kcal/hr). The model parameters are summarized in Table 1. The base wall thickness is assumed to be 0.20 m. We assume an on-peak electricity price of 0.12 \$/kWh available from 9 a.m. to 10 p.m. The off-peak price is 0.04 \$/kWh. A demand rate of 16 \$/kW is charged for the monthly peak electricity demand. The natural gas price is fixed at 0.10 \$/kWh. Average prices were obtained from [22]. The thermal comfort zone is assumed to be 69-77^{o}F. The above PDE-constrained optimization problem is discretized by using a central difference scheme in the axial dimension and an implicit Euler scheme in time. The resulting NLP was implemented in AMPL and solved with the solver IPOPT [24].

From the solution of the open-loop dynamic optimization problem, we obtain the optimal cost and use it as a reference for the best economic performance of the system. The resulting minimum annual cost is \$28,672 (demand cost is approximately 60% of total cost). The one-year forecast problem contains 96,613 constraints and 26,349 degrees of freedom and can be solved in 25 iterations and 30 CPU-seconds. All numerical calculations are performed on a personal computer with 4 GB of memory and a Duo-Core Intel processor running at 2.1 GHz. We then solve closed-loop D-RTO problems over the entire year with prediction horizons of 1, 3, 6, 9, 12, 16, and 24 hr. An update period Δ of 1 hr is used. The 24-hr forecast problem contains 253 constraints and 70 degrees of freedom and can be solved, in warm-start mode, in 10 iterations and 0.1 CPU-seconds. The relative costs (excluding demand costs) are presented in Figure 6(a). As can be seen, for a purely reactive strategy, the relative costs can go as high as 24% as a result of lack of proactiveness. In addition, we observe that a horizon of 24 hr is sufficient to achieve the minimum potential costs. The reason is that the thermal mass of the building cannot be used for a very long time because energy is lost through the wall. In fact, we found that as the building insulation is enhanced, the costs can be further reduced. To illustrate this

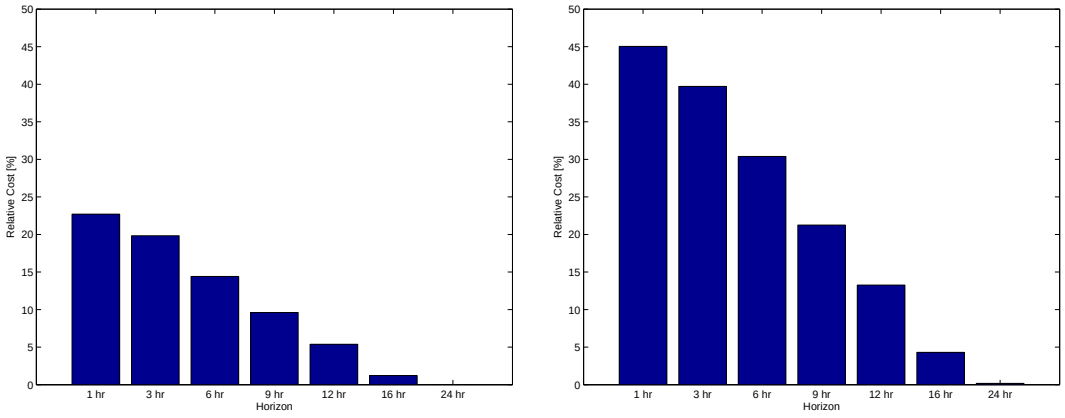
situation, in Figure 6(b) we present the relative costs with an increased wall thickness of 0.3 m. As can be seen, using a forecast of 24 hr can reduce costs by 45%. On the other hand, when the building is poorly insulated, increasing the forecast horizon does not reduce the costs. In other words, *the economic potential of adding forecast information is tightly related to the ability to store energy in the system*, which is in turn influenced by the building characteristics. The predicted cost savings agree with the results of a previous economic study on a photovoltaic-hydrogen hybrid system [26]. In that study, we found that the operating costs can be reduced by 75% by incorporating forecast information of the solar radiation. In that system, increasing the forecast horizon smooths out the control actions, resulting in higher component efficiencies and reduced power losses.

Table 1: Building model parameters.

Parameter	**Value**	**Units**	**Meaning**
β	0.001	$\frac{m^2}{hr}$	thermal diffusivity of wall
C_I	34,800	$\frac{kcal}{^oC}$	internal heat capacity
$\mathbf{k}$	0.1	$\frac{kcal}{m \cdot hr \cdot ^oC}$	conductivity of wall
S	3,500	m^2	wall total surface area
A	1,000	m^2	usable total surface area
V	10,000	m^3	building total volume
α'	4	$\frac{kcal}{m^2 \cdot hr \cdot ^oC}$	convective heat transfer coefficient (wall inner side)
α''	10	$\frac{kcal}{m^2 \cdot hr \cdot ^oC}$	convective heat transfer coefficient (wall outer side)
L	0.20	m	wall thickness
C_{elec}	0.12	$\frac{\$}{kWh}$	on-peak electricity cost
C_{elec}	0.04	$\frac{\$}{kWh}$	off-peak electricity cost
C_{gas}	0.10	$\frac{\$}{kWh}$	natural gas cost

In Figure 7 we present the temperature set-points for the 24-hr and 1-hr forecast cases during 10 days in the winter season. As can be seen, the 24-hr forecast strategy determines the optimal timing at which electric heating needs to be turned on at night. Note that the optimum timing and the peak temperature depend on the expected ambient temperature. On the other hand, the reactive strategy is not able to foresee the structure of the electricity prices. This strategy suggests that the optimal policy is to keep the temperature set-point always at the lowest possible value in order to reduce the overall heating costs. Although this strategy seems intuitive, it is clearly not optimal if the structure of the electricity rates and the thermal mass of the building can be exploited. From Figure 8, we observe that the optimal cooling policy during the summer follows a peak-shifting strategy. The resulting policy recommends letting the building cool down at night until the temperature gets close to the lower limit of the comfort zone. During the day, the building is allowed to heat up progressively until it reaches the highest limit of the comfort zone. Similar results have been obtained by Braun and coworkers [3]. The proposed closed-loop D-RTO framework can account for time variations and

correct the policy automatically on-line. In this simplified study the cooling requirements are negligible because we account only for heat gains and losses through the wall. In addition, the day-night temperature difference at this location is large during summer, as seen in Figure 8. A more detailed study should also account for internal heat gains, radiation heating, air recycling, and humidity factors. Nevertheless, these preliminary results indicate that the performance of operating strategies can benefit from anticipating the weather conditions.



a) Base Insulation b) Enhanced Insulation

Figure 6: Impact of forecast horizon on economic performance of building system.

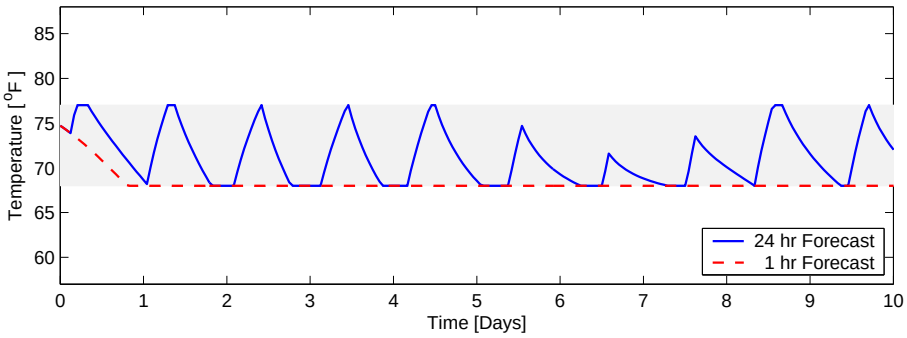


Figure 7: Optimal temperature set-points of closed-loop D-RTO with 1-hr and 24-hr forecasts. Comfort zone is highlighted in gray.

2.3 Gaussian Process Modeling

In the previous sections, we have demonstrated that important economic benefits can be realized by using weather forecast information. But several questions arise: Can we get accurate forecast information? What techniques can be used? In which form is

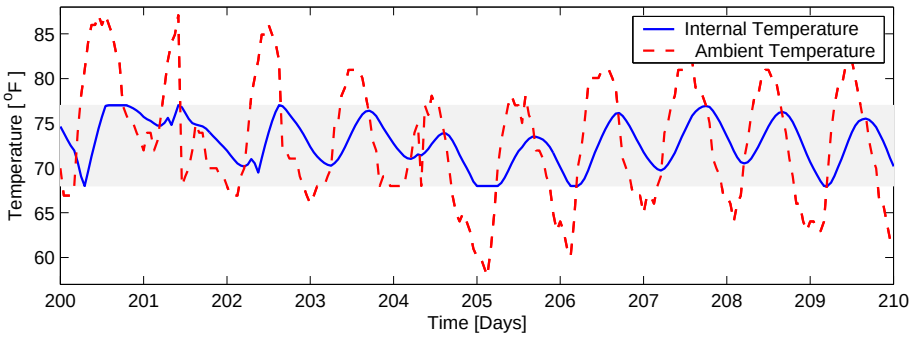


Figure 8: Internal temperature set-point and ambient temperature during 10 days in summer. Closed-loop D-RTO with a forecast of 24 hr was used. Comfort zone is highlighted in gray.

the information needed? What is the effect of neglecting uncertainty? In this section, we present a technique to derive empirical forecast and uncertainty information based solely on data. This will be used to explain how to connect the closed-loop optimization framework with the forecast capabilities. We show that, while the empirical strategy is practical and useful, it has important limitations. This fact motivates our interest in obtaining more consistent forecast information through a detailed weather model.

The most straightforward forecasting alternative is to use historical data to construct a time-series regression model. An approach that has recently received attention is Gaussian process (GP) modeling [18]. The idea is to construct the regressive model by specifying the structure of the covariance matrix rather than the structure of the dynamic model itself as in the Box-Jenkins approach [1]. We have found that this feature makes the GP approach more flexible. To illustrate the use of this technique, we construct a forecast model for the ambient temperature by regressing the future temperature (output) χ_{k+1} to the current and previous temperature values (inputs) $\chi_k, ..., \chi_{k-N}$ that can be obtained from weather information data bases. In this case, N is selected long enough to capture the periodic trends of the ambient temperature. We define the model inputs as $\mathbf{X}_{[j]} = [\chi_{k-N-j}, ..., \chi_{k-j}]$ and the outputs as $\mathbf{Y}_{[j]} = \chi_{k+1-j}$, and we collect a number of training sets $j = 0, ..., N_{train}$. We assume that the inputs are correlated through an exponential covariance function of the form

$$\mathbf{V}(\mathbf{X}_{[j]}, \mathbf{X}_{[i]}, \eta) := \eta_0 + \eta_1 \cdot \exp\left(-\frac{1}{\eta_2}\|\mathbf{X}_{[j]} - \mathbf{X}_{[i]}\|^2\right), \tag{9}$$

where η_1, η_2, and η_3 are hyperparameters estimated by maximizing the log likelihood function

$$\log p(\mathbf{Y}|\eta) = -\frac{1}{2}\mathbf{Y}\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta)\mathbf{Y} - \frac{1}{2}\log\det(\mathbf{V}(\mathbf{X}, \mathbf{X}, \eta)). \tag{10}$$

Once the optimal hyperparameters η^* are obtained, we can compute mean predictions $\mathbf{Y}^P$ with associated covariance $\mathbf{V}^P$ at a set of test points $\mathbf{X}^P$. In our context, these are the

evolving temperature trends. The resulting GP posterior distribution is

$$\mathbf{Y}^P = \mathbf{V}(\mathbf{X}^P, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta^*)\mathbf{Y} \tag{11a}$$

$$\mathbf{V}^P = \mathbf{V}(\mathbf{X}^P, \mathbf{X}^P, \eta^*) - \mathbf{V}(\mathbf{X}^P, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta^*)\mathbf{V}(\mathbf{X}, \mathbf{X}^P, \eta^*). \tag{11b}$$

The inverse of the input covariance $\mathbf{V}_\mathbf{X} := \mathbf{V}^{-1}(\mathbf{X}, \mathbf{X}, \eta^*)$ (e.g., its factorization) needs to be computed only during the training phase. With this, we can define a conceptual GP model of the form

$$\mathbf{Y}^P = \mathbf{GP}(\mathbf{X}^P, \eta^*, \mathbf{V}_\mathbf{X}). \tag{12}$$

Note that at current time t_k, we have measurements to compute only the single-step forecast $\bar{\chi}_{k+1}$. To obtain multi-step forecasts, we must propagate the GP predictions recursively. If we define the total number of forecast steps as $N_F := T/\Delta$, we can use the following algorithm,

1. **Forecast mean computation:** For $j = 1, ..., N_F$ do,
   (a) Set $\mathbf{X}^P_{[j]} \leftarrow [\chi_{k-N}, \chi_{k-N+1} ..., \chi_k]$
   (b) Compute $\mathbf{Y}^P_{[j]} = \mathbf{GP}(\mathbf{X}^P_{[j]}, \eta^*, \mathbf{V}_\mathbf{X})$
   (c) Drop last measurement, set $\chi_{k+1} \leftarrow \mathbf{Y}^P_{[j]}$, and update $k \leftarrow k + 1$
2. **Forecast covariance computation:** Compute self-covariance $\mathbf{V}(\mathbf{X}^P, \mathbf{X}^P, \eta^*)$ and cross-covariance $\mathbf{V}(\mathbf{X}^P, \mathbf{X}, \eta^*)$. Compute forecast covariance $\mathbf{V}^P$ from (11b).

This recursion generates the forecast mean $\mathbf{Y}^P = [\bar{\chi}_{k+1}, ..., \bar{\chi}_{k+T}]$ and associated covariance matrix $\mathbf{V}^P$. To illustrate the predictive capabilities of the GP modeling technique, we construct a temperature model based on the Pittsburgh, PA data presented in Figure 5. We used a total of 120 training data sets. We consider a single-step strategy $N_F = 1$ and a multi-step strategy with a horizon of $N_F = 20$. In Figure 9, we present the forecast mean and 100 samples drawn from the corresponding normal distributions $\mathcal{N}(\mathbf{Y}^P, \mathbf{V}^P)$. In the top graph, we can see that the single-step strategy provides reasonable forecasts and the uncertainty bounds encapsulate the true temperature realizations. This strategy can be used to perform operational tasks such as high-frequency regulatory control. From the economic analysis of the previous section, however, short forecasts are shown to be of limited use for long-term set-point optimization. In the bottom graph, we can see that the multi-step GP model is able to capture the periodicity of the trends. However, the forecast mean drifts away from the true temperature realizations. Moreover, the uncertainty bounds are not able to encapsulate the actual realizations. Hence, the applicability of long-term GP forecasts for set-point optimization is limited. The ambient temperature follows strong variations as a result of spatial interactions and slow metereological phenomena that cannot be taken into account through empirical modeling techniques. In the following section, we discuss how to obtain more accurate and consistent forecast information from a detailed weather model, and we present techniques to quantify the uncertainty of the predictions.

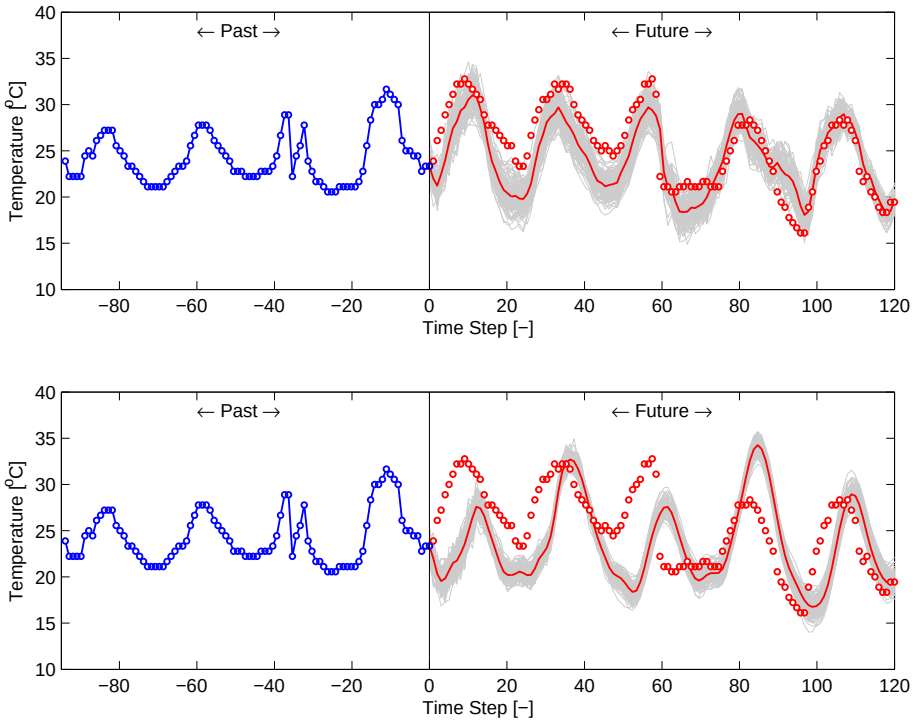


Figure 9: Temperature forecasts with single-step (top) and multi-step GP model (bottom). Forecast mean is solid line, samples are in light gray, and markers are actual realizations.

3 Numerical Weather Prediction Model

In this section, we derive an ensemble data assimilation approach based on a detailed numerical weather prediction (NWP) model to assess the forecast uncertainty. In particular, we are interested in capturing the uncertainty of the temperature field. First we develop a model for the *prior* covariance of the spatial temperature field obtained from the weather model at the current time. Next, the prior distribution is sampled and evolved through the NWP model dynamics. The resulting trajectories are then assembled to obtain an approximate forecast covariance matrix.

The Weather Research and Forecasting (WRF) model is a state-of-the-art mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs [20]. We use the current version of the model, WRF 3.1, with the default settings for the forecast and uncertainty estimation on temperature fields.

The data used in the WRF model corresponds to North American Regional Reanalysis data set that covers 160W-20W, 10N-80N, with a resolution of 10 minutes of a degree. There are 29 pressure levels (1000-100 hPa, excluding the surface) and a three-hour output frequency. The time period under consideration ranges from August 1 to August 30, 2006. This data set includes meteorological fields such as temperature, wind, and humidity, as well as geophysical forcings such as soil albedo and vegetation type.

3.1 Ensemble Approach to Uncertainty Quantification

At current time t_k, the current model states are described through random variables. We consider random variables with a Gaussian probability density function. By using the Gaussian distribution, the forecast uncertainty is described completely by the random variable mean and covariance matrix. The mean is obtained from the WRF model, which has been reconciled to past measurement data. Traditional state estimation techniques such as Kalman filtering and 4-D Var (moving horizon estimation) are used internally. The dimensions of the state vector with a coarse spatial discretization is $\mathcal{O}(10^6)$. Therefore, the covariance matrix is extremely large (grows with the square of the number of states). Hence, in practice, the covariance matrix needs to be approximated with a reduced model [5] or with an ensemble of realizations [17]. In this work, we prefer to use the ensemble approach because it can be implemented by using the WRF model as a black box.

If the dimension of a random variable x is defined as n, a given covariance matrix $\mathbf{V} \in \mathbb{R}^{n\times n}$ can be approximated by an ensemble of m realizations x_i, $1 \leq i \leq m$,

$$\begin{aligned} \mathbf{V} &:= \frac{1}{m-1}\sum_{i=1}^{m}\left(x_i-\overline{x}\right)\left(x_i-\overline{x}\right)^T \approx \mathbf{E}\left[\left(x-\overline{x}\right)\left(x_i-\overline{x}\right)^T\right], \\ \overline{x} &:= \frac{1}{m}\sum_{i=1}^{m} x_i \approx \mathbf{E}[x]. \end{aligned} \tag{13a}$$

One also has that

$$\mathbf{V} = D^{\frac{1}{2}} C D^{\frac{1}{2}}, \quad C_{i,j} = \frac{\mathbf{V}_{i,j}}{\sqrt{D_{i,i}}\sqrt{D_{j,j}}} = \frac{\mathbf{V}_{i,j}}{\sigma_i\sigma_j} = \frac{\sigma_{i,j}^2}{\sigma_i\sigma_j},\ 1 \leq i,j \leq n,$$

where C is the correlation matrix and D is a diagonal matrix holding the local variances ($D_{i,i} = \sigma_i^2$). In the context of the closed-loop stochastic optimization framework of Section 2, x represents the *current* exogenous state or disturbance $\chi(t_k)$ with mean $\overline{x} := \overline{\chi}(t_k)$ and covariance $\mathbf{V} := \mathbf{V}(t_k)$. Note that, in this case, the disturbance at t_k represents a three-dimensional spatial field $\chi(t_k, x, y, z)$. In the following, we simplify the notation by eliminating the explicit dependence on the space dimensions. Using this representation, we now discuss how to approximate the forecast covariance.

Consider the true state of the weather at time t_k, $\chi_{true}(t_k)$. Since the numerical model is not perfect, the true state at t_{k+1} is given by

$$\chi_{true}(t_{k+1}) = \mathcal{M}\left(\chi_{true}(t_k)\right) + \eta(t_k), \tag{14}$$

where $\mathcal{M}$ is the WRF model and η represents the model errors that are assumed to be unbiased with covariance Q, $\eta \in \mathcal{N}(0, Q)$. Since the current state is not known exactly, the numerical prediction at time t_{k+1}, $\chi(t_{k+1})$, is obtained from the model evolution of the *believed* state (true solution perturbed with errors) that is represented by a set of unbiased random variables $\varepsilon(t_k)$, $\varepsilon \in \mathcal{N}(0, \mathbf{V}(t_k))$. With this, we can express the future believed state as

$$\chi(t_{k+1}) = \mathcal{M}\left(\chi_{true}(t_k) + \varepsilon(t_k)\right). \tag{15}$$

Since $\chi(t_{k+1})$ becomes a random variable, we define its covariance matrix as $\mathbf{V}(t_{k+1})$. The matrix is given by

$$\begin{aligned}\mathbf{V}(t_{k+1}) &= \mathbf{E}\left[\left(\chi(t_{k+1}) - \chi_{true}(t_{k+1})\right)\left(\chi(t_{k+1}) - \chi_{true}(t_{k+1})\right)^T\right] &\text{(16a)}\\ &= \mathbf{E}\left[\left(\mathcal{M}\left(\chi_{true}(t_k) + \varepsilon(t_k)\right) - \left(\mathcal{M}\left(\chi_{true}(t_k)\right) + \eta(t_k)\right)\right)\right. \\ &\qquad \left.\cdot\left(\mathcal{M}\left(\chi_{true}(t_k) + \varepsilon(t_k)\right) - \left(\mathcal{M}\left(\chi_{true}(t_k)\right) + \eta(t_k)\right)\right)^T\right]. &\text{(16b)}\end{aligned}$$

This formula reduces to the Kalman filter covariance update under certain special conditions. To illustrate, we first assume that the initial condition errors $\varepsilon(t_k)$ and model errors $\eta(t_k)$ are uncorrelated. Consequently,

$$\mathbf{E}\left[\left(\mathcal{M}\left(\chi_{true}(t_k) + \varepsilon(t_k)\right)\right)\eta(t_k)^T\right] = 0.$$

Moreover, if we assume that the error growth is well approximated by a linearized model, then

$$\mathcal{M}\left(\chi_{true}(t_k) + \varepsilon(t_k)\right) - \mathcal{M}\left(\chi_{true}(t_k)\right) = M \cdot \varepsilon(t_k)\,.$$

where $M = \frac{d\mathcal{M}}{dy}$. It follows that (16a) becomes

$$\begin{aligned}\mathbf{V}(t_{k+1}) &\approx \mathbf{E}\left[\left(M \cdot \varepsilon(t_k) + \eta(t_k)\right)\left(M \cdot \varepsilon(t_k) + \eta(t_k)\right)^T\right]\,, \\ &= \mathbf{E}\left[M\varepsilon(t_k)\varepsilon^T(t_k)M^T\right] + \mathbf{E}\left[M \cdot \varepsilon(t_k)\eta(t_k)^T\right] + \mathbf{E}\left[\eta(t_k)\varepsilon^T(t_k)M^T\right] + \mathbf{E}\left[\eta(t_k)\eta(t_k)^T\right]\,, \\ &= M\mathbf{V}(t_k)M^T + Q\,. &\text{(17)}\end{aligned}$$

Equation (17) represents a linear approximation of the exact error covariance. The ensemble approach, on other other hand, propagates the uncertainties in the current state field through the nonlinear WRF model according to (15). The covariance matrix is approximated by using an ensemble of realizations generated by sampling the prior distribution $\varepsilon(t_k) \in \mathcal{N}(0, \mathbf{V}(t_k))$. We obtain multi-step trajectories by recursive model propagation of each realization

$$\mathbf{Y}_{[j,i]} := \chi(t_{k+j}, i) = \underbrace{\mathcal{M}(\mathcal{M}(...\mathcal{M}}_{j \text{ times}}(\chi_{true}(t_k) + \varepsilon_i(t_k)))),\ j = 1, ..., N_F, i = 1, ..., m, \quad (18)$$

From these trajectories, we compute the multi-step forecast mean $\mathbf{Y}^P$ and covariance matrix $\mathbf{V}^P$ using the ensemble approximations (13). Note that the ensemble approach is able to capture the strong model nonlinearities more accurately. To estimate the forecast covariance, however, we need to specify the prior covariance $\mathbf{V}(t_k)$ which in many cases cannot be computed because of the large state dimensionality. We next introduce a method to approximate this matrix.

3.2 The NCEP Method for Prior Covariance Estimation

The National Centers for Environmental Prediction (NCEP) method [17; 7; 10] has been used to estimate the *spatial* uncertainty information. This method was known formerly as the National Meteorological Center (NMC) method. The idea is to estimate characteristic correlation distances to construct an artificial covariance matrix. The inferred characteristic horizontal correlation distance for this case is approximated by $L_H = 2$ degrees and by $L_V = 500$ meters in the vertical direction. This can be obtained by physical intuition. The spatial correlation function between two spatial points $\chi(t_k, x_i, y_i, z_i)$ and $\chi(t_k, x_j, y_j, z_j)$ is defined as

$$C_{i,j} = \exp\left(-\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{L_H^2} - \frac{(z_j - z_i)^2}{L_V^2}\right). \tag{19}$$

The correlation function (19) is used to construct the covariance matrix from which the ensemble for the initial state field is drawn. Here, we focus on the temperature field $T(t_k, x, y, z)$. The true initial temperature field is not known exactly, but we assume that it is correctly represented by an unbiased random vector $\varepsilon_T(t_k)$. With this, the temperature field $T(t_k)$ is characterized by a random vector $T^B(t_k)$ with the following properties:

$$T^B(t_k) = T(t_k) + \varepsilon_T(t_k)\,,\ \mathbf{E}\left[\varepsilon_T(t_k)\varepsilon_T(t_k)^T\right] = \mathbf{V}_{TT}(t_k)\,,\ \mathbf{E}\left[\varepsilon_T(t_k)\right] = 0$$
$$\Rightarrow T^B(t_k) \in \mathcal{N}\left(T(t_k), \mathbf{V}_{TT}(t_k)\right)\,.$$

The initial temperature field is approximated by an m-member ensemble drawn from

$$T_j^B(t_k) = T(t_k) + GCG^T\xi_j\,,\ 1 \le j \le m\,,\ \xi \in \mathcal{N}(0,1)\,,$$

where $GCG^T \approx \mathbf{V}_{TT}(t_k)$. Here, matrix G transforms the unbalanced variables into full quantities for temperature and is defined as $G_{[j]} := \sigma_G(z_j)\mathbb{I}$, $j = 1, \ldots, nz$, and $G = \text{diag}\left(\mathrm{G}_{[1]}, \ldots, \mathrm{G}_{[\mathrm{nz}]}\right)$, where

$$\sigma_G(z_j) = \mathbf{E}\left[T(t_k, x, y, z_j)\right] / \max_i\left(\mathbf{E}\left[T(t_k, x, y, z_i)\right]\right).$$

This covariance can then be used to compute the forecast mean and covariance through the ensemble approach.

3.3 Validation of Weather and Uncertainty Model

We next validate the forecast and uncertainty temperature model by using the independent measurements of Section 2.2 for the Pittsburgh area [16]. In the left graph of Figure 10, we show a multi-step ensemble of 30 members for temperature realizations, the expected temperature value, and measurements for five days (Aug. 1–6, 2006). This corresponds to a total of $N_F = 120$ forecast steps with $\Delta = 1$hr. In the right graph we present the reconstructed forecast distribution (mean $\pm$ 3σ) obtained from the ensembles. Note that the forecast errors are small ($\pm 5^oC$) and the uncertainty envelope encloses the true (measured)

solution. In addition, the model can capture long-term temperature trends except at the third day, where an unusual temperature drop is observed. In Figure 11(a), we present the hourly evolution of the standard scores of the actual realizations for the multi-step GP model and for the weather model. In Figure 11(b), we present the cumulative standard error of the actual realizations for the multi-step GP model, for the weather model and for the standard normal (in this figure, larger is better). For example, a standard error less than 2 is obtained 65% of the time by the weather model but only 25% by the GP model. We conclude that the weather uncertainty model is far more accurate.

The forecast errors for the GP model are within $\pm 15\,\sigma$. The forecast errors of the weather model are always within $\pm 5\,\sigma$ except at the third day, where the temperature drop of the third day can be clearly identified. The scores analysis suggests that this sudden temperature drop might be due to a sensor malfunction. If we drop this malfunction from the data, the cumulative error would be far closer to the normal for the weather model in Figure 11(b). Nonetheless, while this seems reasonable, the lack of information about that event will introduce our bias into the assessment. We therefore report the best information available and note that, in any event, the weather model is far more accurate than the GP model.

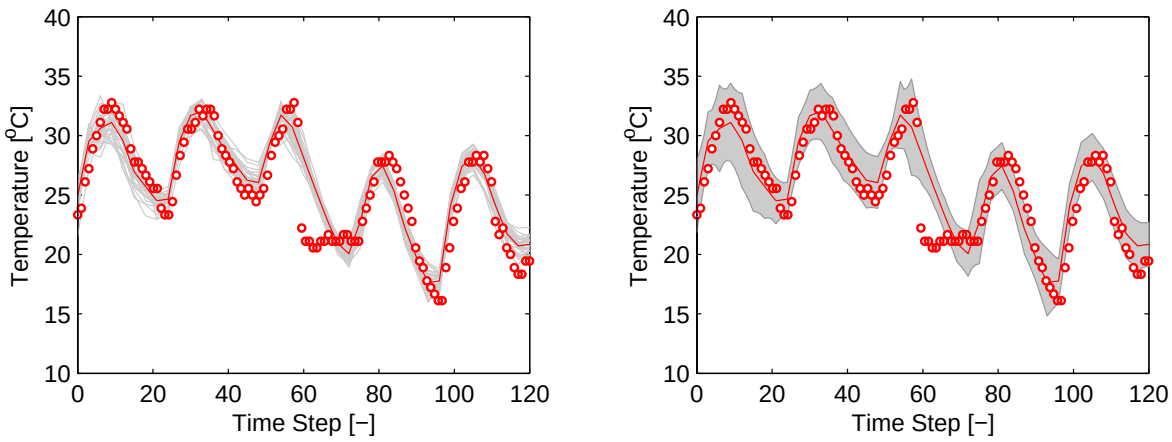
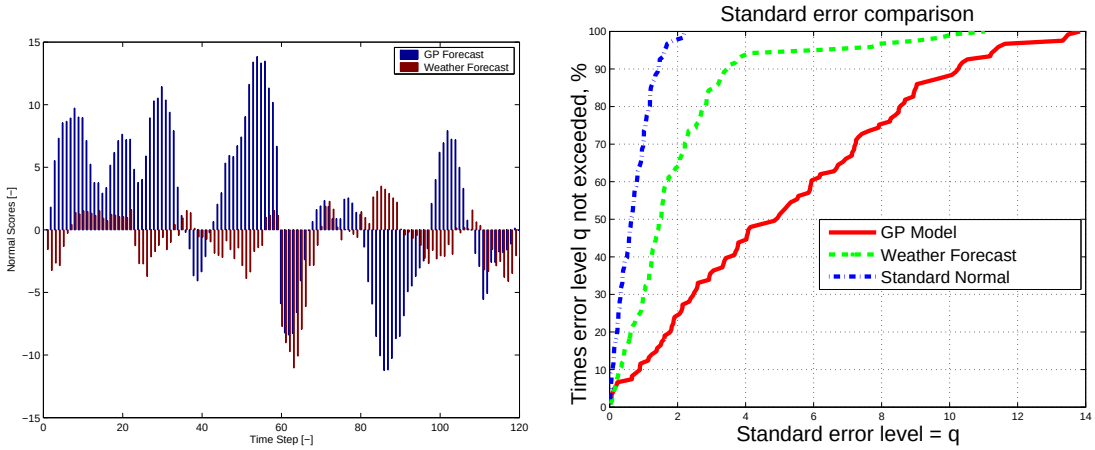


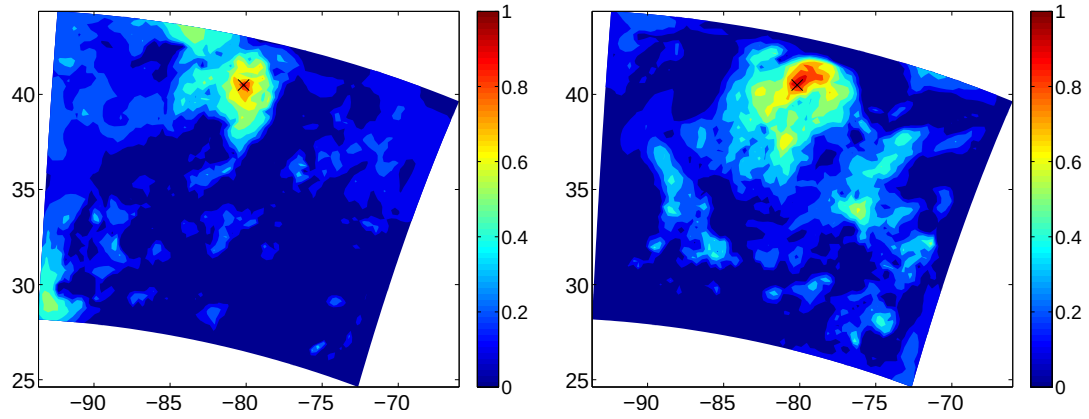
Figure 10: Temperature forecast for the Pittsburgh airport area during August 1–6, 2006. Forecast mean and 30 ensembles obtained from perturbations of initial temperature field (left). Reconstruction of forecast distribution from ensembles - mean $\pm\ 3\,\sigma$ (right).

The ensemble forecast provides sufficient information to describe the spatial error distribution. In Figure 12, we illustrate the horizontal correlation field for the temperature error in the Pittsburgh area corresponding to 10 a.m. August 1 and August 2, 2006 [16]. Note that the error field widens in time as the uncertainty of the forecast increases. Note also that strong temperature variations can arise in relatively narrow regions. Therefore, we emphasize that accounting for spatial effects is critical for accurate and consistent forecasts.



a)Hourly plot b) Cumulative error plot

Figure 11: Standard scores for GP and weather model forecasts.



a) August 1, 2006 b) August 2, 2006

Figure 12: Correlation field for the temperature errors in the Pittsburgh area at 10 a.m. on consecutive days.

3.4 Integrative Study for Building System

We revisit the building system of Section 2.2 in order to illustrate the impact of using forecast uncertainty information in the operating strategy. Here, we solve the SSA approximation of the stochastic counterpart of problem (8) over a horizon of 5 days. The SSA problem is given in equation (20). We use 100 samples drawn from the forecast distributions of the multi-step GP model and of the weather model shown in Figures 9 and 10, respectively. After discretization, the resulting NLP contains 130,900 constraints and 357 degrees of freedom. The problem can be solved, in warm-start mode, in 20 iterations and 68 CPU-seconds. The resulting open-loop profiles for the building internal temperature are presented in Figure 13. In the top graph, we present the temperature

profile for the ideal strategy where perfect forecast information is assumed. Since there is no uncertainty, the predicted temperature profile matches the actual realization. Note that the optimal set-point policy hits continuously the bounds of the comfort zone, as it tries to take advantage of the on-peak and off-peak electricity rates to minimize costs. In the middle graph, we present the optimal temperature profiles obtained using forecast information from the GP model. Since the uncertainty structure of the GP forecast is not able to capture the ambient temperature, the actual realization of the internal building temperature goes outside the comfort zone. In the bottom graph, we see that the use of weather model forecast results in an temperature trajectory that stays within the comfort zone at all times.

The cost penalty sustained by the NCEP weather uncertainty approach from subsections 3.1 and 3.2 when compared to the GP model uncertainty approach is about 10-20% of the ideal cost. One should bear in mind, however, that the GP model cost turns out to be infeasible for the actual realizations (the building temperature significantly exits the comfort level), so using only cost as a performance is misleading in this case. We could easily imagine some financial measure of the violation and report it to balance the perceived cost drop. Nevertheless, given the complex regulatory nature of the comfort level limits, their violation cost may easily be understated. For example, the 10CFR434 federal regulations for new federal, commercial and multi-family high-rise residential buildings require compliance with the comfort zone at least 98 % of the time the building is occupied. Therefore using the GP model would result in the operator being in violation of the federal law (the constraint violation in Figure 13 would be out of compliance more than 30 % of the period stated), the cost of which is difficult to fully assess. Given the difficulty of pricing the violation, it is more beneficial to regard the situation from the constrained optimization perspective and state that feasibility takes precendence over low cost. We conclude that the weather uncertainty model is the only one that has a sufficiently accurate description of the uncertainty to result in a feasible policy at a cost that is still substantially lower than the reactive policy cost.

We also note that the variance of the predicted temperature realizations increases with time. Since the comfort zone is very narrow ($\approx 5^o$C), high-precision forecast information is needed to realize economic benefits.

$$\min_{\varphi_c^{elec}(\tau),\varphi_h^{gas}(\tau),\varphi_h^{elec}(\tau)} \quad \frac{1}{N_S}\sum_{j=1}^{N_S}\int_{t_k}^{t_k+T}\left[C_{elec}(\tau)\varphi_c^{elec}(\tau)+C_{elec}(\tau)\varphi_h^{elec}(\tau)+C_{gas}\varphi_h^{gas}(\tau)\right]d\tau$$

$$C_I\cdot\frac{\partial T_I^j}{\partial\tau} = \varphi_h^{gas}(\tau)+\varphi_h^{elec}(\tau)-\varphi_c^{elec}(\tau)-S\cdot\alpha'\cdot\left(T_I^j(\tau)-T_W^j(\tau,0)\right) \tag{20a}$$

$$\frac{\partial T_W^j}{\partial\tau} = \beta\cdot\frac{\partial^2 T_W^j}{\partial x^2} \tag{20b}$$

$$0 = \alpha'\left(T_I^j(\tau)-T_W^j(\tau,0)\right)+\mathbf{k}\cdot\left.\frac{\partial T_W^j}{\partial x}\right|_{(\tau,0)} \tag{20c}$$

$$0 = \alpha''\left(T_W^j(\tau,L)-T_A^j(\tau)\right)+\mathbf{k}\cdot\left.\frac{\partial T_W^j}{\partial x}\right|_{(\tau,L)} \tag{20d}$$

$$T_I^{min} \le T_I^j(\tau) \le T_I^{max} \tag{20e}$$

$$T_I(0) = T_I^k \tag{20f}$$

$$T_W(0,x) = T_W^k(x),\ \tau\in[t_k,t_k+T], j=1,...,N_S. \tag{20g}$$

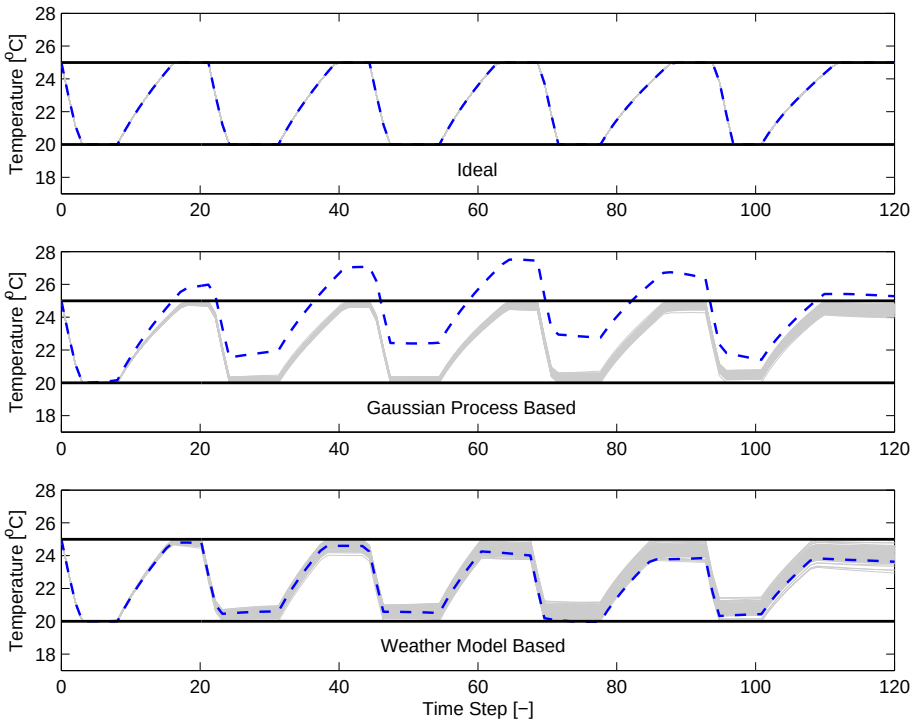


Figure 13: Performance of weather forecast-based operating strategies. Thermal comfort zone is highlighted by thick solid lines, predicted temperatures are gray lines, and actual realizations are dashed lines.

4 Conclusions and Future Work

In this work, we demonstrate that significant costs reductions can be achieved by using operation technology that can anticipate the weather conditions in a systematic manner. We show that adding forecast information provides a mechanism to compute proactive policies that can lead to enhanced performance in systems with slow responses.

We discuss strategies to obtain weather forecast information. We emphasize that a Gaussian-process empirical strategy provides quick estimates but is limited to short horizons and can lead to inconsistent uncertainty bounds. Motivated by these facts, we discuss the potential of using detailed weather models to obtain forecasts. We demonstrate that these models are capable of providing more accurate forecasts and are able to capture temporal and spatial correlations of the state fields. We extend a weather model to provide forecast covariance information through the ensemble-based approach.

As future work, we are interested in establishing a full connection between the weather model forecasts and the stochastic programming framework. To do so, we first must implement the ensemble-based approach in a closed-loop manner. Since the weather model is extremely computationally expensive, a dedicated, centralized parallel computing architecture is needed that can send forecast information to individual control agents. In addition, since the amount of data to be handled is huge, strategies must be established to communicate only the essential forecast statistical information. Another important issue is the fact that the weather model provides information over relatively coarse fields that need to be mapped to the specific location of the energy system under consideration. To this end, we are interested in using a Gaussian process framework to interpolate the spatiotemporal fields. We are also interested in addressing the complexity of large-scale stochastic programming problems through adaptive sampling and variance reduction techniques. Moreover, we will consider more comprehensive models for energy systems in order to refine the current estimates of the economic gains.

Acknowledgments

This work was supported by the Department of Energy, through Contract No. DE-AC02-06CH11357.

References

[1] G. E. P. Box, G. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice-Hall, New Jersey, 1994.

[2] J. E. Braun. Reducing energy costs and peak electricty demand through optimal control of building thermal storage. *ASHRAE Transactions*, 96(2):876–888, 1990.

[3] J. E. Braun, K. W. Montgomery, and N. Chaturvedi. Evaluating the performance of building thermal mass control strategies. *HVAC&Research*, 7(4):403–428, 2001.

[4] C. Carrara, L. F. Fantini, and A. Lorenzi. Influence of the outdoor temperature variation on the performance of a building heating system. In C. J. Hoogendoorn, editor, *Energy Conservation in Heating, Cooling and Ventilating Buildings*. 1978.

[5] E.M. Constantinescu, T. Chai, A. Sandu, and G.R. Carmichael. Autoregressive models of background errors for chemical data assimilation. *Journal of Geophysical Research*, 112:D12309, 2007.

[6] D. B. Crawley, L. K. Lawrie, C. O. Pedersen, R. J. Liesen, D. E. Fisher, R. K. Strand, R. D. Taylor, R. C. Winkelmann, W. F. Buhl, Y. J. Huang, and A. E. Erdem. EneryPlus: A new-generation building energy simulation program. In *Proceedings of Building Simulation '99*, volume 1, pages 81–88, 1999.

[7] R. Franke and E. Barker. Vertical correlation functions for temperature and relative humidity errors. *Monthly Weather Review*, 128(12):3962–3981, 2000.

[8] H. K Geyer and R. K Ahluwalia. GCtool for fuel cell systems design and analysis: User documentation. Argonne National Laboratory, 1998.

[9] J. Kadam, W. Marquardt, M. Schlegel, T. Backx, O. Bosgra, and P.J. Brouwer. Towards integrated dynamic real-time optimization and control of industrial processes. *Proceedings Foundations of Computer-Aided Process Operations (FOCAPO2003)*, pages 593–596, 2003.

[10] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471, 1996.

[11] S. A. Klein, J. A. Duffie, and W. A. Beckman. TRNSYS: A transient simulation program. *ASHRAE Transactions*, 82:623–633, 1976.

[12] Argonne National Laboratory. PSAT (Powertrain Systems Analysis Toolkit), 2009. `http://www.transportation.anl.gov/`.

[13] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1):215–241, 2006.

[14] T. Markel, A. Brooker, T. Hendricks, V. Johnson, K. Kelly, B. Kramer, M. O'Keefe, S. Sprik, and K. Wipke. ADVISOR: A systems analysis tool for advanced vehicle modeling. *Journal of Power Sources*, 110(2):255–266, 2002.

[15] T.E. Marlin and A.N. Hrymak. Real-time operations optimization of continuous processes. In *Fifth International Conference on Chemical Process Control (CPC-5)*, 1996.

[16] National Oceanic and Atmospheric Association. National Weather Service Forecast Office, 2009. `http://www.erh.noaa.gov/pbz/hourlyclimate.htm`.

[17] D.F. Parrish and J.C. Derber. The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763, 1992.

[18] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006.

[19] A. Ruszczynski and A. Shapiro. *Stochastic Programming (Handbooks in Operations Research and Management Series)*. Elsevier Science BV, Amsterdam, 2003.

[20] W.C. Skamarock, J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang, and J.G. Powers. A description of the Advanced Research WRF version 2. Technical Report Tech Notes-468+ STR, NCAR, 2005.

[21] O. Ulleberg. The importance of control strategies in PV-hydrogen systems. *Solar Energy*, 76:323–329, 2004.

[22] Kansas State University. Comparing fuel costs of heating and cooling systems. *Technical Report*, 2003.

[23] S. R. Vosen and J. O. Keller. Hybrid energy storage systems for stand-alone electric power systems: optimization of system performance and cost through control strategies. *International Journal of Hydrogen Storage*, 24(12):1139–1156, 1999.

[24] A. Wächter and L. T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Program.*, 106:25–57, 2006.

[25] D. C. White. Online optimization: What have we learned? *Hydrocarbon Processing*, 77:55–69, 1998.

[26] V. M. Zavala, M. Anitescu, and T. Krause. On the Optimal On-Line Management of Photovoltaic-Hydrogen Hybrid Energy Systems. In R. M. de Brito Alves, C. A. Oller do Nascimento, and E. C. Biscaia Jr., editors, *Proceedings of 10th International Symposium on Process Systems Engineering*, 2009.