THE POWER OF DIVERSITY: DATA-DRIVEN ROBUST PREDICTIVE CONTROL FOR ENERGY EFFICIENT BUILDINGS AND DISTRICTS

GEORGIOS DARIVIANAKIS, ANGELOS GEORGHIOU, ROY S. SMITH AND JOHN LYGEROS

ABSTRACT. The cooperative energy management of aggregated buildings has recently received a great deal of interest due to substantial potential energy savings. These gains are mainly obtained in two ways: (*i*) Exploiting the load shifting capabilities of the cooperative buildings; (*ii*) Utilizing the expensive but energy efficient equipment that is commonly shared by the building community (e.g., heat pumps, batteries and photovoltaics). Several deterministic and stochastic control schemes that strive to realize these savings, have been proposed in the literature. A common difficulty with all these methods is integrating knowledge about the disturbances affecting the system. In this context, the underlying disturbance distributions are often poorly characterized based on historical data. In this paper, we address this issue by exploiting the historical data to construct families of distributions which contain these underlying distributions with high confidence. We then employ tools from data-driven robust optimization to formulate a multistage stochastic optimization problem which can be approximated by a finite-dimensional linear program. The proposed method is suitable for tackling large scale systems since its complexity grows polynomially with respect to the system variables. We demonstrate its efficacy in a numerical study, in which it is shown to outperform, in terms of energy cost savings and constraint violations, established solution techniques from the literature. We conclude this study by showing the significant energy gains that are obtained by cooperatively managing a collection of buildings with heterogeneous characteristics.

1. INTRODUCTION

Approximately 20-40% of the total energy consumption in the developed countries is attributed to the building sector, an amount that often exceeds even the industrial and transportation sectors [1]. Concerns about the growing environmental impact of building energy consumption, led the EU and the US government to set the target of a net zero-energy for 50% of their commercial buildings by 2040 [2]. In this context, active building energy management has attracted considerable attention with substantial efforts to be devoted to developing sophisticated control schemes that are capable of reducing the buildings energy impact while ensuring comfortable conditions for the building users [3–6]. Nevertheless, the opportunities for large savings within individual buildings can be limited, and depend on the specific building actuation systems and construction characteristics [7–9].

Further savings can be envisaged by cooperatively managing the aggregated energy demands of a collection of buildings in a district via an energy hub. The energy hub is a conceptual entity that provides the interface between the building community and the power grid by utilizing shared energy generation, conversion and storage equipment (e.g., heat pumps, batteries and photovoltaics) [10]. In this setting, significant energy gains can be obtained by exploiting the *diversity* of the available energy sources, equipment and building characteristics. However, the main body of the literature separates the optimal control of the energy hub

from that of the buildings. A number of papers treat the building energy demands as exogenous signals, which are typically estimated using building simulation environments such as EnergyPlus [11]. These studies focus on the control of the devices within the energy hub, employing either deterministic [12,13] or stochastic [14] formulations.

A large body of literature addresses the building control problems using deterministic schemes [6–9], as they are scalable and suitable for problems with long prediction horizons. However, deterministic schemes suffer from frequent constraints violations due to their inability to handle the system disturbances [15]. Stochastic schemes can potentially address this issue but very often they are either not practically scalable (e.g., see the discussion in [16] on the application of the scenario approach [17, 18] to building control problems), or their reliability strongly depends on the exact knowledge of the distributional characteristics of the system disturbances [19]. Poor estimation of these underlying distributions can lead to significant performance deterioration [20].

To tackle this issue, distributionally robust methods have recently been introduced in the literature [21]. Instead of considering an exact distribution, these methods account for all distributions in a family which shares only a few structural parameters, such as moments and/or support information [22,23]. In this context, the historical data are only partially exploited to obtain a rough estimate of these structural parameters. On the other hand, purely data-driven approaches which systematically exploit the historical data information to determine appropriate families of distributions, have also been suggested [24,25]. These methods provide the tools to reformulate an originally infinite dimensional problem to a finite dimensional convex semi-definite optimization problem (SDP). Typically, the number of constraints in this problem depends on the historical data size, which limits the scalability of the method when dealing with a large data set. Although polynomial algorithms exist for solving SDPs [26], these algorithms are computationally demanding limiting the applicability of these methods to small problem instances.

Our goal is to develop a data-driven stochastic control scheme that is capable of cooperatively operating the energy hub and the district buildings. This paper extends the preliminary work in [27], providing a refined robust approach on handling the system disturbances by systematically exploiting the available historical realizations of these stochastic processes. In particular:

(1) We propose a data-driven approach that exploits the historical data to train linear models of the exogenous disturbances, and construct families of distributions that encompass the true disturbance distributions with high confidence.

(2) We exploit the structure of these families of distributions to formulate a robust multistage stochastic optimization problem that minimizes the wost-case expected energy costs of the system. The size of the resulting optimization problem is independent of the historical data size. We approximate this infinite dimensional problem by a finite-dimensional linear program that scales polynomially with respect to the prediction horizon length, and more importantly, unlike the SDP approaches cited above, it can effectively be solved for large scale systems.

(3) We demonstrate in an extensive numerical study the efficacy of the proposed method which is shown to outperform the optimally tuned deterministic equivalent in terms of energy consumption and constraint violations. We observe that higher cost benefits are obtained by merging buildings with dissimilar operation plans, rather than diverse construction characteristics.

The paper is organized as follows. In Section 2, we review in a more compact way, the modeling approach presented in [27]. The main contributions of this paper are summarized in Sections 3 and 4 where the developed data-driven distributionally robust methods, and the techniques associated with the derivation of
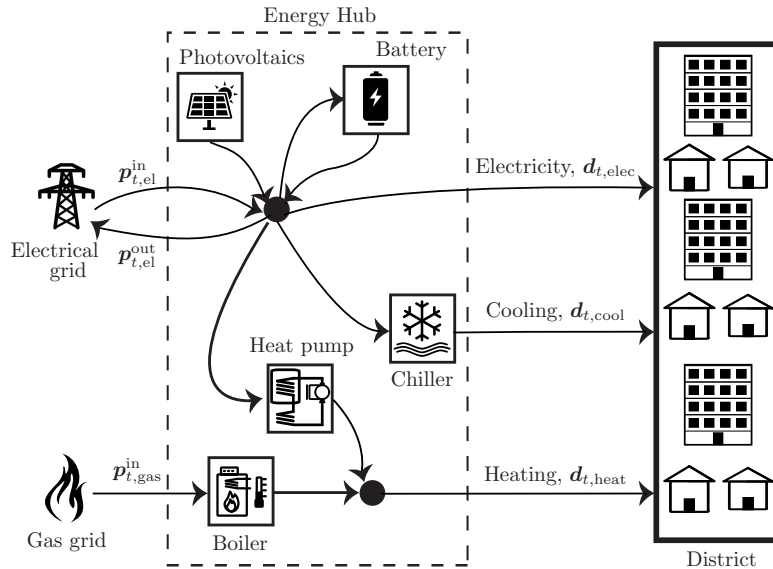
FIGURE 1. Heating, cooling and electricity network of a district.

a tractable approximation to the infinite dimensional stochastic optimization problem, are discussed. We conclude this paper with an extensive numerical study performed in Section 5. The proofs of the propositions can be found in the Appendix.

Notation: All random vectors appearing in this paper are defined on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{E}(\cdot)$ denotes the expectation operator with respect to $\mathbb{P}$. Random vectors are represented in boldface, while their realizations are denoted by the corresponding symbols in normal font. For given matrices $(A_1, \ldots, A_m)$, we define $A := \operatorname{diag}(A_1, \ldots, A_m)$ as the block-diagonal matrix with elements $(A_1, \ldots, A_m)$ on its diagonal. Given vectors $(v_1, \ldots, v_m)$, $v_i \in \mathbb{R}^{k_i}$, we define $[v_1, \ldots, v_m] := [v_1^\top, \ldots, v_m^\top]^\top \in \mathbb{R}^k$ with $k = \sum_{i=1}^m k_i$, as their vector concatenation. We denote by $\mathbf{1}$ and $\mathbf{0}$ the vectors with components all one and zero, respectively. The dimension of the corresponding vectors and the vector concatenations, will be clear from the context.

2. System modelling

In this section, we describe the energy hub and building dynamics using discrete time, bilinear models affected by stochastic exogenous disturbances. We assume that these disturbances evolve according to stochastic processes $\{\boldsymbol{\xi}_t\}_{t \in \mathcal{T}}$, where $\mathcal{T} = \{1, \ldots, T\}$, and T is the length of the horizon considered. The vector $\boldsymbol{\xi}_t$ encompasses all the stochastic processes affecting the energy hub and building dynamics, such as the ambient temperature, solar radiation and internal gains of the buildings.

2.1. Energy hub dynamics

We visualize the energy hub as a *conceptual* entity that houses and interconnects a number of conversion, storage and production devices that are shared by the building community. An energy hub essentially provides the interface between the energy grid and the building community. As depicted in the illustrative example of Fig. 1, the energy hub is capable of (*i*) purchasing electricity and gas to meet the electricity, cooling and heating demand of the building community, and (*ii*) selling the electricity produced by the photovoltaics or stored in the battery to maximize profit. In the following, we define the set $\mathcal{K}$ to include the energy hub devices,

and the sets $\mathcal{G}$ and $\mathcal{O}$ to include the energy streams from the grid and building community, respectively. In the example of Fig. 1, $\mathcal{K} = \{$Battery, Chiller, Heat pump, Boiler, Photovoltaics$\}$, $\mathcal{G} = \{$Electrical grid, Gas grid$\}$ and $\mathcal{O} = \{$Electricity, Cooling, Heating$\}$.

We model every device $i \in \mathcal{K}$ using linear dynamics and constraints, as follows:

$$\begin{aligned} \boldsymbol{x}_{t+1,i} &= A_i\,\boldsymbol{x}_{t,i} + B_i\,\boldsymbol{u}_{t,i} + C_i\,\boldsymbol{\xi}_t, \\ F_{x,i}\,\boldsymbol{x}_{t,i} &+ F_{u,i}\,\boldsymbol{u}_{t,i} + F_{\xi,i}\,\boldsymbol{\xi}_t \le h_i. \end{aligned} \tag{1}$$

The vectors $\boldsymbol{x}_{t,i}$ and $\boldsymbol{u}_{t,i}$ denote the hub device internal states and control inputs, respectively. The matrices A_i, B_i and C_i have appropriate dimensions and can be derived from the device characteristics. Finally, we assume linear operational constraints captured by the matrices $F_{x,i}$, $F_{u,i}$, $F_{\xi,i}$ and h_i. As discussed in [10,13], linear approximations for the energy hub devices are reasonable at this level of abstraction. Examples of such models for the numerical study of Section 5 are given in Appendix 6.

The energy balancing nodes are used to model the interconnection of the energy hub devices, as follows:

$$H_p\,\boldsymbol{p}_t + H_u\,\boldsymbol{u}_t + H_d\,\boldsymbol{d}_t = \boldsymbol{0}, \tag{2}$$

where the vector $\boldsymbol{p}_t = [\boldsymbol{p}_{t,1}^{\text{in}}, \boldsymbol{p}_{t,1}^{\text{out}}, \dots, \boldsymbol{p}_{t,|\mathcal{G}|}^{\text{in}}, \boldsymbol{p}_{t,|\mathcal{G}|}^{\text{out}}]$ contains the non negative decision variables of the grid energy streams, with $\{\boldsymbol{p}_{t,i}^{\text{in}}\}_{i \in \mathcal{G}}$ and $\{\boldsymbol{p}_{t,i}^{\text{out}}\}_{i \in \mathcal{G}}$ denoting the energy purchased from, and sold to the grid, respectively. In a similar way, the vector $\boldsymbol{d}_t = [\boldsymbol{d}_{t,1}, \dots, \boldsymbol{d}_{t,|\mathcal{O}|}]$ concatenates the demands of the building community, and $\boldsymbol{u}_t = [\boldsymbol{u}_{t,1}, \dots, \boldsymbol{u}_{t,|\mathcal{K}|}]$ captures the decision variables of the energy hub devices. The matrices H_p, H_u and H_d have proper dimensions and model the power flows affecting the respective balancing node.

2.2. Building dynamics

We model the building dynamics using bilinear state space models, motivated by the resistance-capacitance models of [28]. The accuracy of these bilinear models was validated against established building simulation software [11], and real buildings [9]. We denote by $\mathcal{B}$ the set of district buildings and for each building $i \in \mathcal{B}$ we assume a bilinear model that captures the temperature evolution of the rooms, walls, ceiling and floors, as follows,

$$\boldsymbol{x}_{t+1,i} = A_i\boldsymbol{x}_{t,i} + \big(B_i + \boldsymbol{x}_{t,i}^{\top}E_i\big)\boldsymbol{u}_{t,i} + \big(D_i + \boldsymbol{v}_{t,i}^{\top}C_i\big)\boldsymbol{\xi}_t, \tag{3}$$

where $\boldsymbol{u}_{t,i}$ contains the inputs to these building actuation systems which are not coupled with the disturbances (e.g., radiators, thermally activated building structures (TABS), air handling unit (AHU), floor heating). The rest of the control inputs (e.g. , position of the blinds) are captured by $\boldsymbol{v}_{t,i}$. The system matrices A_i, B_i, C_i, D_i and E_i depend on the specific building characteristics (e.g., number of rooms, construction material, window fraction area and actuation units).

The state to input bilinear terms in (3) severely complicate the design of computationally tractable controllers. To address this issue, we replace the bilinear terms $\boldsymbol{x}_{t,i}^{\top}E_i\boldsymbol{u}_{t,i}$, with the linear terms $\widehat{x}_i^{\top}E_i\boldsymbol{u}_{t,i,j}$, where $\widehat{x}_i$ is the initial condition for the states of building i. Therefore, the dynamics of the approximated system are given as:

$$\boldsymbol{x}_{t+1,i} = A_i\,\boldsymbol{x}_{t,i} + B_i(\widehat{x}_i)\,\boldsymbol{u}_{t,i} + C_i(\boldsymbol{v}_{t,i})\boldsymbol{\xi}_t + D_i\boldsymbol{\xi}_t, \tag{4a}$$

The matrix $B_i(\widehat{x}_i)$ and the matrix function $C_i(\boldsymbol{v}_{t,i})$, are readily constructed by the primitive description of the dynamics in (3). More sophisticated approaches could also be envisioned (e.g. linearizing about a state trajectory), but we do not pursue this further here for the shake of simplicity.

We consider linear operational constraints (e.g., radiator and AHU limitations), as follows:

$$F_{x,i}\,\boldsymbol{x}_{t,i} + F_{u,i}\,\boldsymbol{u}_{t,i} + F_{v,i}\,\boldsymbol{v}_{t,i} + F_{\xi,i}\,\boldsymbol{\xi}_t \le h_i\,, \tag{4b}$$

where the matrices $F_{x,i}$, $F_{u,i}$, $F_{v,i}$, $F_{\xi,i}$ and f_i are derived using the BRCM Toolbox [28], and are of appropriate dimensions. In addition to the operational constraints in (4b), we consider user specified comfort ranges given as,

$$\text{lb}_{t,i} \leq \boldsymbol{x}_{t,i} \leq \text{ub}_{t,i}, \tag{4c}$$

where $\text{lb}_{t,i}$ and $\text{ub}_{t,i}$ are lower and upper bounds. One typically considers bounds only on the room temperatures, so many of the upper/lower bounds can be assumed to be plus/minus infinity, effectively eliminating the corresponding constraints. These bounds may vary during the day to reflect occupancy patterns (e.g., office buildings temperature bounds are often relaxed during the night time since the building is empty). We will refer to constraint set (4c) as the *comfort constraints* [29].

2.3. **Coupling of buildings to the energy hub**

We model the coupling between the energy hub and the buildings with the following set of equality constraints:

$$\boldsymbol{d}_{t,j} = \sum_{i \in \mathcal{B}} \boldsymbol{\eta}_{i,j}^{\top} \boldsymbol{u}_{t,i}, \quad \forall j \in \mathcal{O}, \tag{5}$$

where $\boldsymbol{\eta}_{i,j}$ is a $(0,1)$-matrix that models whether the building actuation system $\boldsymbol{u}_{t,i}$ is connected to the j-th output energy source of the hub, $\boldsymbol{d}_{t,j}$.

To simplify notation, we compactly rewrite Eqns (1), (2), (4) and (5), as follows:

$$\begin{aligned}
\boldsymbol{x} &= \boldsymbol{B}(\widehat{x})\boldsymbol{u} + \boldsymbol{C}(\boldsymbol{v})\boldsymbol{\xi} + \boldsymbol{D}\boldsymbol{\xi}, \\
\boldsymbol{F}_{\pi}\boldsymbol{\pi} + \boldsymbol{F}_{\xi}\boldsymbol{\xi} &\leq \boldsymbol{h},
\end{aligned} \tag{6}$$

where the concatenated vector $\boldsymbol{\pi}$ is defined as $\boldsymbol{\pi} = [\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}]$. Note that all the vectors in (6) are considered over the time horizon $\mathcal{T}$, e.g., $\boldsymbol{p} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_{|\mathcal{T}|}]$, while the vectors $\boldsymbol{x}$, $\boldsymbol{u}$ are defined such they include the state and input variables, respectively, for both the energy hub devices and the buildings.

3. Modelling the uncertainty

To model the evolution of the stochastic process $\boldsymbol{\xi}_t$ the exact characterization of its distribution is needed. However, this information is typically unavailable. Nevertheless, historical data such as past realizations of $\boldsymbol{\xi}_t$ are usually available. A simple approach is to construct an empirical distribution using the historical data set, and then use it as a proxy for the true distribution. However, if the data set is small, then there might be several distributions that can describe these data points. As indicated in [20], if one arbitrarily chooses an element from this family of distributions, e.g., the empirical distribution, then the solution of the resulting stochastic optimization problem can differ significantly from the solution in which the true distribution is used. In the following, we adopt a robust perspective to this problem by training linear models of the exogenous disturbances and constructing families of distributions that describe the disturbance realizations during past years. We then formulate an optimization problem whose solution addresses all distributions in the constructed family.

3.1. **Dynamics of the stochastic process**

We denote by $\mathcal{D}$ the set that contains all the sources of disturbances appearing in our problem, e.g., solar radiation, ambient temperature, internal gains. For each $i \in \mathcal{D}$, we model the disturbance $\boldsymbol{\xi}_{t,i}$ as a

deterministic forecast $f_{t,i}$, plus the stochastic error term $\boldsymbol{e}_{t,i}$. To simplify notation, we omit index i in the subsequent discussion, i.e.,

$$\boldsymbol{\xi}_t = f_t + \boldsymbol{e}_t. \tag{7a}$$

Weather forecasts are easily accessible from national weather services (e.g., COSMO-7 of MeteoSwiss [30]) while forecasts of the anticipated internal gains can be obtained from standard weekly profiles of typical building configurations (e.g., office and residential buildings [29]). The dynamical evolution of $\boldsymbol{e}_t$, is captured by a first order autoregressive system,

$$\boldsymbol{e}_{t+1} = \alpha_t \boldsymbol{e}_t + \boldsymbol{w}_t, \tag{7b}$$

where α_t is time-varying constant, and $\boldsymbol{w}_t \in W_t$ is the stochastic process governing the stochastic evolution of the error. The autoregressive evolution of $\boldsymbol{e}_t$ is motivated by systematic errors between forecast and actual realization. Indeed, the forecast provided by MeteoSwiss can differ from the true realization of the uncertain parameters due to, among other things, the spatial difference between the local weather station and the building leading to error correlation over time. Additionally, $\boldsymbol{w}_t$ models the noise from imperfect forecasting and possible measurement noise from the sensing devices.

Equation (7) suggests that instead of constructing the probability distribution associated with $\boldsymbol{\xi}_t$, we can equivalently construct the probability distribution associated with $\boldsymbol{w}_t$ which we denote by $\mathbb{P}_t$. We make the following structural assumptions for the random variables $\boldsymbol{w}_t$: (i) they are normally distributed, (ii) they are mutually independent for all time stages $t \in \mathcal{T}$ and disturbances $i \in \mathcal{D}$, and (iii) they are stationary with respect to different days (i.e., α_t and the distribution of w_t for the same t but for different days is the same). Essentially these assumptions imply that first order models adequately capture the correlation over time, and one can ignore the correlation between different days and disturbances. We validate these hypotheses in Section 5.2 based on real disturbance data.

3.2. Data-driven uncertainty sets

For each disturbance $i \in \mathcal{D}$, the historical data set consists of realization and forecast pairs of the form $\{(\xi_t^k, f_t^k)\}$ with t denoting the time of the k-th day that the data was recorded. We consider N records of forecast and realization compatible with the sampling time used for the models of Section 2 (typically, daily records for several years sampled hourly).

We calculate the constants α_t describing the dynamics in (7) using least squares fitting. In particular, for each $t \in \{1, \ldots, 24\}$, we solve,

$$\min_{\alpha_t \in \mathbb{R}} \ \sum_{k=1}^{N} \left((\xi_{t+1}^k - f_{t+1}^k) - \alpha_t (\xi_t^k - f_t^k) \right)^2, \tag{8}$$

and using the optimal solution construct the residual data points as follows:

$$w_t^k := (\xi_{t+1}^k - f_{t+1}^k) - \alpha_t (\xi_t^k - f_t^k), \ \ k = 1, \ldots, N.$$

We denote by $\mathcal{S}_t := \{w_t^k\}_{k=1}^N$ the set of residuals derived from the historical data, and subsequently denote by $\widehat{\mu}_t$ and $\widehat{\sigma}_t^2$ their empirical mean and variance, respectively.

We construct a family of distributions, $\mathcal{P}_t$, that are compatible with the residuals of our historical data $\mathcal{S}_t$. The set $\mathcal{P}_t$ is sometimes referred as the ambiguity set in the robust optimization literature [21, 24]. We consider an ambiguity set, $\mathcal{P}_t$, of the following form:

$$\mathcal{P}_t = \{\mathcal{N}(\mu_t, \sigma_t^2) : \underline{\mu}_t \leq \mu_t \leq \overline{\mu}_t, \ \underline{\sigma}_t^2 \leq \sigma_t^2 \leq \overline{\sigma}_t^2\}. \tag{9a}$$

The constants $(\underline{\mu}_t, \overline{\mu}_t, \underline{\sigma}_t^2, \overline{\sigma}_t^2)$, are selected to ensure that given $\mathcal{S}_t$, the true distribution, $\mathbb{P}_t$ of $\boldsymbol{w}_t$, is an element of $\mathcal{P}_t$ with high probability. This is achieved by utilizing concepts from statistical hypothesis theory. In particular, the statistic associated with the chi-square hypothesis test, $h^\chi := (N-1)\widehat{\sigma}_t^2/\sigma_t^2$, follows the chi-square distribution with $(N-1)$ degrees of freedom [31]. To ensure that $\underline{\sigma}_t^2 \leq \sigma_t^2 \leq \overline{\sigma}_t^2$, with probability at least $1 - \delta_t^\chi$, we set,

$$
\begin{aligned}
\underline{\sigma}_t^2 &= (N-1)\,\widehat{\sigma}_t^2/q_{(N-1)}^\chi(\delta_t^\chi/2), \\
\overline{\sigma}_t^2 &= (N-1)\,\widehat{\sigma}_t^2/q_{(N-1)}^\chi(1-\delta_t^\chi/2).
\end{aligned}
\tag{9b}
$$

Here, $q_{(N-1)}^\chi(\cdot)$ denotes the quantile function of the chi-square distribution with $(N-1)$ degrees of freedom.

We construct the bounds of μ_t in a similar way. Given that the $\boldsymbol{w}_t$ follows a normal distribution and its variance σ_t^2 follows the chi-square distribution, $h^{st} := (\widehat{\mu}_t - \mu_t)/\sqrt{\widehat{\sigma}_t^2/N}$, follows a student t-distribution with $N-1$ degrees of freedom, [31]. To ensure that $\underline{\mu}_t \leq \mu_t \leq \overline{\mu}_t$, with probability at least $1 - \delta_t^{\mathrm{st}}$, we set,

$$
\begin{aligned}
\overline{\mu}_t &= \widehat{\mu}_t + q_N^{st}(\delta_t^{\mathrm{st}}/2)\sqrt{\widehat{\sigma}_t^2/N}, \\
\underline{\mu}_t &= \widehat{\mu}_t - q_N^{st}(\delta_t^{\mathrm{st}}/2)\sqrt{\widehat{\sigma}_t^2/N},
\end{aligned}
\tag{9c}
$$

where, $q_N^{st}(\cdot)$ denotes the quantile function of the Student's t-distribution with $(N-1)$ degrees of freedom.

The bounds given in (9) explicitly determine the family of distributions, $\mathcal{P}_t$. The following proposition provides the confidence, $1 - \delta_t$, by which the true disturbance distribution, $\mathbb{P}_t$, is a member of $\mathcal{P}_t$.

Proposition 1. *Given the sample data, $\mathcal{S}_t$, let $\mathbf{P}_{\mathcal{S}_t}$ be the $|N|$-fold product distribution of $\mathbb{P}_t$, then*

$$
\mathbf{P}_{\mathcal{S}_t}\big(\mathbb{P}_t \in \mathcal{P}_t\big) \geq 1 - \delta_t,
$$

where $\delta_t = \delta_t^{st} + \delta_t^\chi$.

Let us now define the multivariate stochastic process $\boldsymbol{w} = [\boldsymbol{w}_{0,1}, \ldots, \boldsymbol{w}_{|\mathcal{T}|,|\mathcal{D}|}] \in W$, with $W = W_{0,1} \times \ldots \times W_{|\mathcal{T}|,|\mathcal{D}|}$. We characterize the joint distribution $\mathbb{P}$ of $\boldsymbol{w}$, as follows,

$$
\begin{aligned}
\mathcal{P} = \Big\{ \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \ &\text{with } \boldsymbol{\mu} = (\mu_{1,1}, \ldots, \mu_{|\mathcal{T}|,|\mathcal{D}|}) \\
&\text{and } \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_{1,1}^2, \ldots, \sigma_{|\mathcal{T}|,|\mathcal{D}|}^2) \\
&\text{s.t.} \mathbb{P}_{1,1} \in \mathcal{P}_{1,1}, \ldots, \mathbb{P}_{|\mathcal{T}|,|\mathcal{D}|} \in \mathcal{P}_{|\mathcal{T}|,|\mathcal{D}|} \Big\}.
\end{aligned}
$$

We close this section, by compactly rewriting the disturbance modelling equations over the horizon T, as follows:

$$
\boldsymbol{\xi} = \boldsymbol{H}(\widehat{e})\boldsymbol{w},
$$

where the matrix $\boldsymbol{H}(\widehat{e})$ is readily constructed from (7) and depends on the forecast, and the vector $\widehat{e} = [e_{0,1}, \ldots, e_{0,|\mathcal{D}|}]$ containing the current error disturbance measurements.

4. Problem formulation

Our objective is to minimize the worst-case expected cost of energy purchased from the grid by the building community, over a finite horizon. This must be achieved while satisfying the dynamics and constraints of the devices and the buildings in the system. This problem can be formulated as a multistage stochastic linear

program, as follows:

$$\min_{} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\boldsymbol{c}^{\top} \boldsymbol{p} \right)$$

$$\text{s.t.} \quad \boldsymbol{p}, \boldsymbol{u} \in \mathcal{C}, \ v \in \mathcal{R}, \ \boldsymbol{\pi} = [\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{x}, \boldsymbol{u}, v],$$

$$\boldsymbol{\xi} = \boldsymbol{H}(\widehat{e})\boldsymbol{w}, \tag{10}$$

$$\boldsymbol{x} = \boldsymbol{B}(\widehat{x})\boldsymbol{u} + \boldsymbol{C}(v)\boldsymbol{\xi} + \boldsymbol{D}\boldsymbol{\xi},$$

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\boldsymbol{F}_{\pi}\boldsymbol{\pi} + \boldsymbol{F}_{\xi}\boldsymbol{\xi} \le \boldsymbol{h} \right) \ge 1 - \epsilon,$$

where $\boldsymbol{c}_t = [c_{t,1}^{\text{in}}, c_{t,1}^{\text{out}}, \ldots, c_{t,|\mathcal{G}|}^{\text{in}}, c_{t,|\mathcal{G}|}^{\text{out}}]$ contains the time varying prices of the grid energy streams. Note that the prices $\{c_{t,i}^{\text{in}}\}_{i \in \mathcal{G}}$ for purchasing energy from the grid are positive scalars, while $\{c_{t,i}^{\text{out}}\}_{i \in \mathcal{G}}$ are negative. Furthermore, we select a chance constraints formulation with worst-case violation probability ϵ. The decision variables v that are coupled with the disturbances $\boldsymbol{\xi}$, are allowed to take values in the generic finite-dimensional vector space $\mathcal{R}$. We choose the decision variables $\boldsymbol{p}, \boldsymbol{u}$ to be *strictly causal* disturbance feedback policies. For instance, $\boldsymbol{p}_t$ is the strictly causal, vector valued function of the energy purchased from the grid at time $t \in \mathcal{T}$, defined as $\boldsymbol{p}_t : W_1 \times \ldots \times W_{t-1} \to \mathcal{R}$. In the following, we denote by $\mathcal{C}$, the infinite-dimensional function space of *strictly causal* disturbance feedback policies. Finally, the expected cost of the objective has been formulated to account for the worst-case multivariate distribution $\mathbb{P}$ in the ambiguity set $\mathcal{P}$.

4.1. **Constraint relaxation**

The proposed control architecture is based on a receding horizon implementation of the system. In this setting, the mismatch between the building prediction model and real system, can lead to comfort bounds violations. To address infeasible instances of Problem (10) due to this issue, we relax the comfort constraints (4c), as follows:

$$\max\{\text{lb}_{t,i} - \boldsymbol{x}_{t,i}, 0, \boldsymbol{x}_{t,i} - \text{ub}_{t,i}\} \le \boldsymbol{s}_{t,i},$$

where we refer $\boldsymbol{s}_{t,i}$ as the slack variable. To this end, we rewrite the compactly formulated inequality (6), as follows:

$$\boldsymbol{F}_{\pi}\boldsymbol{\pi} + \boldsymbol{F}_{\xi}\boldsymbol{\xi} - \boldsymbol{s} \le \boldsymbol{h}, \tag{11}$$

where $\boldsymbol{s} \in \mathcal{C}$ denotes the concatenated slack variable vectors over the horizon T. Notice that the slack variables are only introduced to deal with building model mismatch which can be an issue on a receding horizon implementation. On the other hand, the chance constraint formulation in Problem (10), is primarily used to address in a probabilistic fashion the extreme realization of the uncertain parameters $\boldsymbol{w}$.

4.2. **Optimization over linear feedback policies**

The optimal solution of Problem (10) remains intractable due to the infinite dimensional structure of its decision variables. However, a tractable approximation of Problem (10) can be obtained by restricting the decision variables to the finite dimensional space of affine policies denoted as $\mathcal{C}_{\text{aff}}$. A strictly causal affine policy, e.g. for the variables p_t is given by

$$\boldsymbol{p}_t = p_{0,t} + \sum_{s=1}^{t-1} P_{s,t} \, \boldsymbol{w}_s, \tag{12}$$

where $p_{0,t} \in \mathcal{R}$, and matrices $P_{s,t}$ of appropriate dimensions. We refer to (12) as the strictly causal affine decision rule (ADR) [32]. We refer to the linear policy in (12) which disregards the disturbance history (e.g., $\boldsymbol{p}_t = p_{0,t}$), as open loop policy (OLP). The non-adaptive nature of the OLP policy provides an even more

conservative controller parametrization with the benefit of a considerably smaller number of optimization variables.

In this context, the approximated variant of Problem (10) is given as follows:

$$
\begin{aligned}
\min \ \sup_{\mathbb{P} \in \mathcal{P}} \ & \mathbb{E}_{\mathbb{P}} \left(\boldsymbol{c}^\top \boldsymbol{p} + \gamma \mathbf{1}^\top \boldsymbol{s} \right) \\
\text{s.t.} \ \ & \boldsymbol{p}, \boldsymbol{u}, \boldsymbol{s} \in \mathcal{C}_{\text{aff}}, \ v \in \mathcal{R}, \ \boldsymbol{\pi} = [\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{x}, \boldsymbol{u}, v], \\
& \boldsymbol{\xi} = \boldsymbol{H}(\widehat{e}) \boldsymbol{w}, \\
& \boldsymbol{x} = \boldsymbol{B}(\widehat{x}) \boldsymbol{u} + \boldsymbol{C}(v) \boldsymbol{\xi} + \boldsymbol{D} \boldsymbol{\xi}, \\
& \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \Big(\boldsymbol{F}_\pi \boldsymbol{\pi} + \boldsymbol{F}_\xi \boldsymbol{\xi} - \boldsymbol{s} \le \boldsymbol{h} \Big) \ge 1 - \epsilon,
\end{aligned}
\tag{13}
$$

where the additional term $\gamma \mathbf{1}^\top \boldsymbol{s}$ in the objective penalizes the constraint violations in (11), with parameter $\gamma \in \mathbb{R}_+$. A discussion for appropriate values of γ can be found in [33].

4.3. Chance constraint approximation

The main body of the literature exploits the structure of the ambiguity sets to propose semi-definite [22,25] and second-order cone [34] reformulations of the distributionally robust chance constraints. Although convex problems with conic constraints are generally tractable, they are computationally demanding for large systems. An alternative approach is to construct a set $\widehat{W} \subseteq W$ such that the feasible region of the corresponding robust constraint is a subset of the feasible region of the distributionally robust chance constraint, i.e.,

$$
\text{if } \boldsymbol{F}_\pi \boldsymbol{\pi} + \boldsymbol{F}_\xi \boldsymbol{\xi} - \boldsymbol{s} \le \boldsymbol{h}, \ \forall \boldsymbol{w} \in \widehat{W},
\tag{14a}
$$

$$
\text{then } \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \Big(\boldsymbol{F}_\pi \boldsymbol{\pi} + \boldsymbol{F}_\xi \boldsymbol{\xi} - \boldsymbol{s} \le \boldsymbol{h} \Big) \ge 1 - \epsilon.
\tag{14b}
$$

Such methods are discussed in [24] where the authors compute $\widehat{W}$ described by non-linear constraints. Nevertheless, these approaches typically lead to semi-definite reformulations of the robust constraint (14a). To tackle this issue, we resort to a more stringent condition, in which we require that at least, $1 - \epsilon$, of the probability mass of each distribution, $\mathbb{P}$, in the ambiguity set $\mathcal{P}$, is contained in $\widehat{W}$. To this end, we construct for every $i \in \mathcal{D}$ and $t \in \mathcal{T}$, the compact convex set $\widehat{W}_{t,i}$, as follows:

$$
\begin{aligned}
\widehat{W}_{t,i} = \Big\{ \boldsymbol{w}_{t,i} \mid \ & \boldsymbol{w}_{t,i} \ge \underline{\mu}_{t,i} - \Phi^{-1} \Big(1 - \underline{\beta}_{t,i} \Big) \overline{\sigma}_{t,i} \,, \\
& \boldsymbol{w}_{t,i} \le \overline{\mu}_{t,i} + \Phi^{-1} \Big(1 - \overline{\beta}_{t,i} \Big) \overline{\sigma}_{t,i} \ \Big\}.
\end{aligned}
\tag{15a}
$$

where the constants $\underline{\mu}_{t,i}$, $\overline{\mu}_{t,i}$, and $\overline{\sigma}_{t,i}$ are given in (9c) and (9b), respectively. Moreover, $\Phi^{-1}(\cdot)$ denotes the inverse cumulative normal distribution function, and $\underline{\beta}_{t,i}$ and $\overline{\beta}_{t,i}$ are positive constants chosen as,

$$
\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{D}} \Big(\underline{\beta}_{t,i} + \overline{\beta}_{t,i} \Big) = \epsilon.
\tag{15b}
$$

Finally, we set $\widehat{W}$ to be,

$$
\widehat{W} = \widehat{W}_{1,1} \times \ldots \times \widehat{W}_{|\mathcal{T}|,|\mathcal{D}|}.
\tag{15c}
$$

Proposition 2. *Let $\widehat{W}$ defined in (15). Then, the following probabilistic guarantee holds:*

$$
\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\boldsymbol{w} \in \widehat{W}) \ge 1 - \epsilon.
$$

Notice that satisfying (14a) together with the condition in Proposition 2 is more stringent than the actual chance constraint. In other words,

$$\text{if } \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(\boldsymbol{w} \in \widehat{W}) \geq 1 - \epsilon, \text{ and (14a) holds,}$$
$$\text{then } \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\boldsymbol{F}_{\pi}\boldsymbol{\pi} + \boldsymbol{F}_{\xi}\boldsymbol{\xi} - \boldsymbol{s} \leq \boldsymbol{h}\Big) \geq 1 - \epsilon.$$

Imposing this additional condition, we further restrict the feasible region with the benefit of gaining computational tractability due to the simple hyperrectangular structure of $\widehat{W}$. This, in turn, allows us to address the large scale problems examined in this paper.

4.4. **Objective function reformulation**

We employ the epigraph representation to equivalently rewrite the worst-case expectation in the objective function of Problem (13). In particular, the linearity of the objective function allows us to replace the expected value with:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{c}^{\top}\boldsymbol{p} + \gamma\mathbf{1}^{\top}\boldsymbol{s}\right) = \sup_{\mathbb{P}\in\mathcal{P}}\left(\boldsymbol{c}^{\top}\widetilde{\boldsymbol{p}} + \gamma\mathbf{1}^{\top}\widetilde{\boldsymbol{s}}\right) = \tau, \tag{16}$$

which can equivalently be written as,

$$\boldsymbol{c}^{\top}\widetilde{\boldsymbol{p}} + \gamma\mathbf{1}^{\top}\widetilde{\boldsymbol{s}} \leq \tau, \ \forall\boldsymbol{\mu} \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}], \tag{17}$$

where $\widetilde{\boldsymbol{p}}$, and $\widetilde{\boldsymbol{s}}$ are derived from (12) by replacing the stochastic variable $\boldsymbol{w}$ by its expected value $\boldsymbol{\mu}$ (e.g., $\widetilde{\boldsymbol{p}}_t = p_{0,t} + \sum_{s=1}^{t-1} P_{s,t}\,\boldsymbol{\mu}_t$). Notice that the epigraph representation in (17) is exact with the scalar variable τ replacing the original worst-case objective.

We conclude this section by providing the final approximation of Problem (10) as follows:

$$
\begin{aligned}
\min \ \ & \tau \\
\text{s.t.} \ \ & \boldsymbol{p}, \boldsymbol{u}, \boldsymbol{s} \in \mathcal{C}_{\text{aff}}, \ v \in \mathcal{R}, \ \boldsymbol{\pi} = [\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{x}, \boldsymbol{u}, v], \\
& \boldsymbol{c}^{\top}\widetilde{\boldsymbol{p}} + \gamma\mathbf{1}^{\top}\widetilde{\boldsymbol{s}} \leq \tau, \ \forall\boldsymbol{\mu} \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}], \\
& \left. \begin{aligned} \boldsymbol{\xi} &= \boldsymbol{H}(\widehat{e})\boldsymbol{w}, \\ \boldsymbol{x} &= \boldsymbol{B}(\widehat{x})\boldsymbol{u} + \boldsymbol{C}(v)\boldsymbol{\xi} + \boldsymbol{D}\boldsymbol{\xi}, \\ \boldsymbol{F}_{\pi}\boldsymbol{\pi} + \boldsymbol{F}_{\xi}\boldsymbol{\xi} - \boldsymbol{s} &\leq \boldsymbol{h}, \end{aligned} \right\} \ \forall\boldsymbol{w} \in \widehat{W}.
\end{aligned}
\tag{18}
$$

Problem (18) still retains its infinite structure involving a continuum space of decision variables and constraints. However, as indicated in [32, 35], Problem (18) can be reformulated into a linear optimization problem by employing traditional robust optimization techniques that use duality to translate the semi-infinite structure into a finite number of linear constraints. Unlike other robust convex programs which employ semi-definite programming, this problem is computationally tractable and can be solved in very high dimensions.

5. Numerical results

In this section, we perform numerical studies to assess the performance of the ADR and OLP control methodologies. We compare these methods to the classical and commonly used in practice *certainty equivalence problem* (CEP) in which all random variables of Problem (10) are replaced with their expected value. To make the comparison of these control strategies more transparent, we do not consider the possibility of selling energy to the grid. Moreover, we analyze the sensitivity of the considered control methods with respect to parameters deviations, and investigate the potential gains that can be obtained by cooperatively managing buildings with heterogeneous construction and operation characteristics.
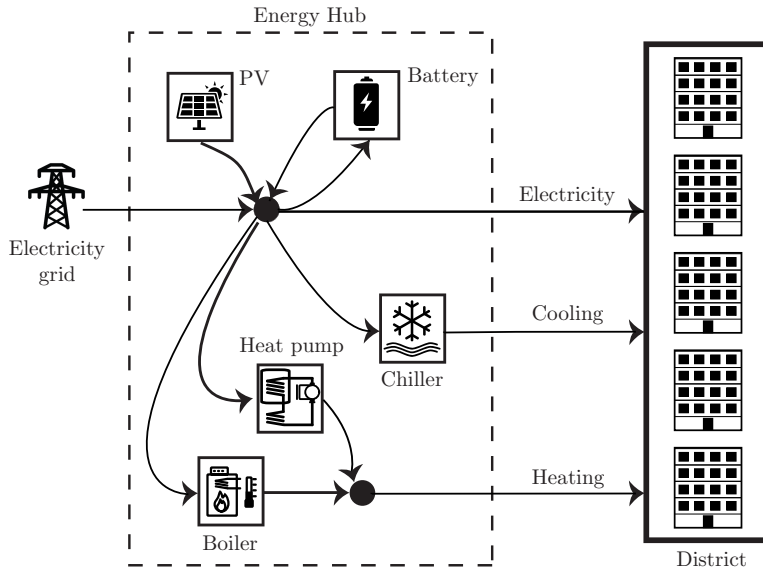
FIGURE 2. Energy hub configuration for numerical studies.

5.1. **Problem configuration**

We consider districts composed of buildings with roughly the same floor area which are connected through an energy hub that comprises five devices: chiller, boiler, heat pump, photovoltaics (PV) and battery, as depicted in Fig. 2. To keep the ratio between demand and supply in the energy hub, relatively constant, we linearly scale the capacities of the hub devices based on the number of buildings in our system. A detailed description of the hourly discretized dynamics and constraints governing the energy hub components can be found in the Appendix 6.

The buildings have heterogeneous construction characteristics, summarized in Table 1. In particular, we use hourly discretized building models described in [9] characterized by the building type BT ∈ {Swiss Passive (SP), Swiss Average(SA), Swiss Target (ST)}, and construction type CT = {heavy, light}. Each building consists of 5 rooms which are characterized by the window fraction area WFA = {30%,50%,80%} and their corresponding facade orientation. The dynamics and constraints of each building are generated using the BRCM toolbox [28] to which the individual building construction details along with the specifications of the control devices (radiators, AHU, TABS, blinds) are provided.

Each of the building models in Table 1, consists of 113 states. Model reduction techniques are used to derive a simplified, but sufficiently accurate, building model. For instance, using the *bilinear balanced truncation*

Building specifications					
No.	Area(m^2)	WFA	BT	CT	Input Devices
1	420	30%	SP	heavy	AHU, blinds, radiator
2	420	50%	SP	light	AHU, blinds, TABS
3	441	80%	SA	light	AHU, blinds, TABS
4	441	50%	SA	heavy	AHU, blinds, radiator
5	374	50%	ST	heavy	AHU, blinds, radiator

TABLE 1. Summary of the 5 buildings used in the simulations.

Time	Cost	Winter Bounds	Summer Bounds
05:00 - 23:00	0.145 CHF/kWh	$[21, 25]°C$	$[20, 23]°C$
23:00 - 05:00	0.097 CHF/kWh	$[15, 30]°C$	$[15, 30]°C$

TABLE 2. Electricity day/night tariff variations and comfort constraints bounds.

method described in [36], we generate a 59 state model, with a maximum absolute error of less than $0.1°C$, during step response simulations for building 1 of Table 1. By evaluating the Hankel matrix of this building system, we identified the time constants associated with the three most controllable and observable modes as 11.74 days, 4.47 hours and 4.24 minutes. Approximately the same time constants were identified for the other building systems in Table 1. The reduced order linear models are used only for prediction purposes, while the original bilinear models, given in (3), are employed to simulate the buildings dynamics in the closed-loop implementation.

We compare the performance of the ADR, OLP and CEP control designs using the metrics of purchased grid energy, room constraint violations and solution time. The cost of the energy purchased from the grid is measured in Swiss Franc (CHF), and the room constraint violations are measured in Kelvin hours (Kh). We assume time-varying electricity tariffs and comfort constraints bounds, as given in Table 2. The same comfort bounds are used in every room and building considered in the system. The disturbance forecasts and realizations are the same as those used by the OptiControl project [37], for the city of Zürich during the years 2006 and 2007. The data set from year 2006 is used to train the linear disturbance models and construct the families of distributions given in equations (8) and (9). The 2007 dataset is then used to test the performance of the controllers developed based on the resulting distribution families. Finally, we choose as soft constraint penalization, $\gamma = 10^3$, confidence levels $\delta_{t,i}^{\chi} = \delta_{t,i}^{\mathrm{st}} = 0.01$, and constraint violation level $\epsilon = 0.01$. We select the violation levels for the upper and lower bounds of the i-th disturbance at time t, as $\overline{\beta}_{t,i} = \underline{\beta}_{t,i} = \epsilon/(2|\mathcal{T}||\mathcal{D}|)$.

5.2. Disturbance model verification

We consider a disturbance set $\mathcal{D}$ which comprises seven sources of uncertainty; ambient and ground temperatures, four sources of solar radiation (North, South, West and East), and building internal gains. The analysis in Section 3 requires that the residual uncertain parameters $\boldsymbol{w}_t$ are normally distributed and independent over time. We verify these assumptions using the Shapiro-Wilk [38] and Pearson [31] hypothesis tests on historical, weather and occupancy, data for the year 2006.

Time	AT	SRS	SRE	IG
07:00	0.989	0.991	0.972	0.940
08:00	0.996	0.997	0.996	0.945
09:00	0.994	0.995	0.993	0.923
10:00	0.993	0.965	0.992	0.967
11:00	0.994	0.990	0.987	0.920
12:00	0.992	0.941	0.981	0.928
13:00	0.997	0.996	0.988	0.945
14:00	0.995	0.995	0.984	0.965

TABLE 3. Statistic of the Shapiro-Wilk normality test.

In Table 3, we report the values of the Shapiro-Wilk test statistic from 7:00 to 14:00 for the ambient temperature (AT), solar radiation south (SRS) and east (SRE), and internal gains (IG). We choose to present the results from 7:00 to 14:00 since these are the hours with the greatest variation in mismatch between forecasts and realizations. Values which are close to 1 indicate that the sample data are compatible with a normal distribution. In particular, the normality hypothesis is accepted at the significance level of 0.1 for the atmospheric processes, and 0.01 for the internal gains. These results are in accordance with the study in [39] where it is shown that normal distributions sufficiently capture the evolution of weather processes. By contrast, in [40] it is argued that Poisson distributions should be used to generate occupancy profile trajectories. Although, Poisson distributions can be approximated by Gaussians [41], we stress that the bounds in (9) are useful even if the underlying data are mildly non-Gaussian, as suggested in [31, §11.3].

Time	07:00	08:00	09:00	10:00	11:00	12:00
07:00	1.00	0.04	0.02	0.01	-0.01	0.01
08:00	0.04	1.00	0.05	0.05	-0.01	0.00
09:00	0.02	0.05	1.00	0.06	0.03	-0.09
10:00	0.01	0.05	0.06	1.00	0.00	-0.09
11:00	-0.01	-0.01	0.03	0.00	1.00	-0.03
12:00	0.01	0.00	-0.09	-0.09	-0.03	1.00

TABLE 4. Statistic of the Pearson correlation test.

In Table 4, we provide the values of the Pearson test statistic from 7:00 to 12:00 for the SRS and the results are similar for the other sources of uncertainty. This correlation statistic can range between plus and minus one with values close to zero denoting probably uncorrelated data. We emphasize that for every disturbance, the hypothesis of uncorrelated data is accepted at the significance level of 0.05 which provides a strong statistical evidence for the validity of our uncertain disturbance assumptions.

5.3. Prediction horizon selection

We investigate the effect of the prediction horizon on the performance of the ADR, OLP and CEP control methodologies. In Fig. 3, we show the respective curves generated by conducting a receding horizon simulation during the first week of January 2007 for building 1 in Table 1. We observe that the cost of purchased grid energy associated with the CEP solution method is the least, at the expense of frequent comfort constraint violations. On the contrary, the stochastic approximations (OLP and ADR), are less cost efficient but lead to many fewer constraint violations.

We can identify several factors associated with the selection of a suitable prediction horizon:

(*i*) The dominant time constants of the buildings were roughly identified as 11 days, 4 hours and 4 minutes. Hence, a horizon of $T \geq 7$ hours facilitates the controller to excite the modes of the system that determine its short term evolution. We emphasize that contrary to the ADR and CEP methods, the long horizons deteriorate the performance of the OLP method due to its inability to adapt on the growing size of the disturbance uncertainty.

(*ii*) The six hour gap between day and night comfort bounds, given in Table 2. A prediction horizon of $T \geq 6$, is required for the system to anticipate the comfort bounds of the next day. In this way, it can utilize the heating capacity of the buildings and the battery storage to exploit the day-night tariff structure of the electricity prices.
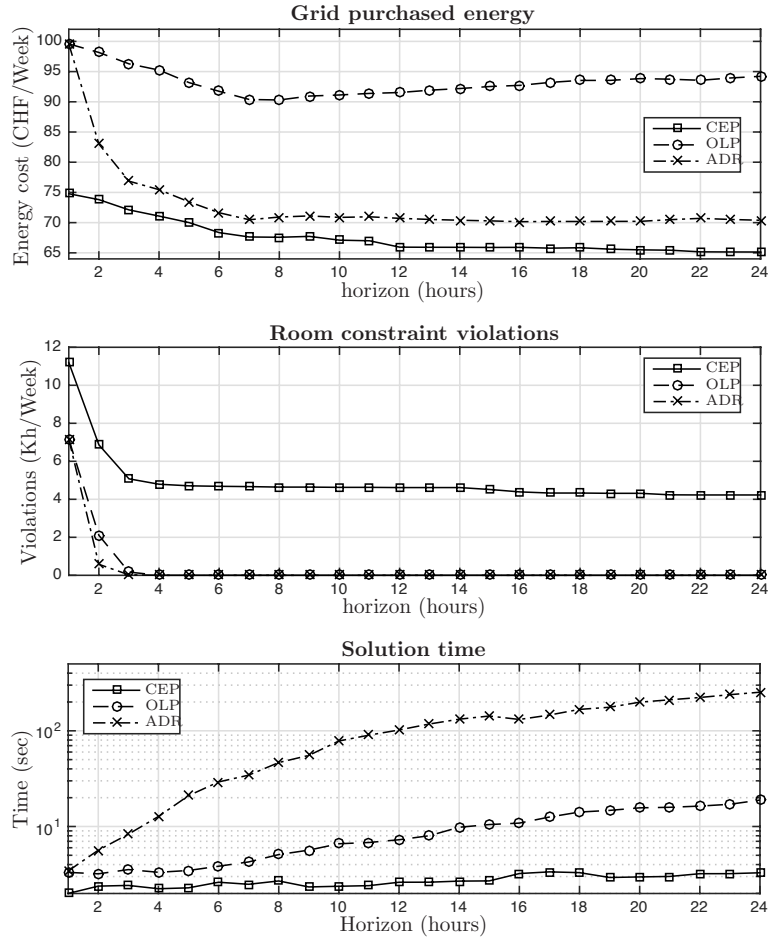
FIGURE 3. Purchased grid energy, room constraint violations and solution time with respect to the prediction horizon length T during a typical winter week.

(*iii*) The renewable energy production peaks around midday. A horizon of $T \geq 6$, is sufficient for the system to anticipate the cost-free energy that will be available from the photovoltaic units. In this way, it can efficiently utilize the battery to fully exploit the potential excess of energy.

(*iv*) The computation time associated with the ADR and OLP control methods which increases with the prediction horizon length.

The above analysis suggests that a prediction horizon of $T = 8$ hours provides a reasonable trade-off between foresight of the controller and computational tractability. This value is used for the rest of this section.

5.4. **Solution method selection**

To generalize our observations regarding the comparison of the ADR, OLP and CEP control methodologies. The simulation experiment is extended to a district composed of the 5 buildings summarized in Table 1. The system is simulated in a receding-horizon fashion using data realizations of 12 consecutive weeks (restarting at the beginning of each week) for the winter and summer periods of 2007, starting January 1st and June 29th, respectively.

We calculated the purchased energy cost and room constraint violations for each one of these 12 weeks, and we report the results in Table 5. The table entries correspond to the *(empirical mean, empirical standard*

		Winter	
Method	Cost (p.u.)	Violations (Kh/Week)	Basis (CHF/Week)
CEP	(1.00, 0.21)	(4.31, 0.68)	327.56
OLP	(1.34, 0.19)	(0.00, 0.00)	
ADR	(1.07, 0.21)	(0.03, 0.01)	

		Summer	
Method	Cost (p.u.)	Violations (Kh/Week)	Basis (CHF/Week)
CEP	(1.00, 0.17)	(2.23, 0.14)	64.19
OLP	(1.07, 0.18)	(0.03, 0.01)	
ADR	(1.03, 0.16)	(0.06, 0.01)	

TABLE 5. Receding horizon performance results.

deviation) over these 12 weeks scenarios. Notice that we present the energy costs in the per unit (p.u.) system using as base values the mean costs occurring for the CEP method during the winter and summer period, respectively. Once again, we observe that the CEP solution method is the most cost effective at the expense of significant comfort constraint violations. The ADR and OLP control methodologies result in higher cost of purchased grid energy but results in fewer constraint violations. Notice that the ADR approximation is considerably more cost efficient than the OLP, while it achieves the same level of constraint violation. This is attributed to the nature of the ADR method which takes into account the potential adaptation of future decisions based on the realizations of the disturbance variables.
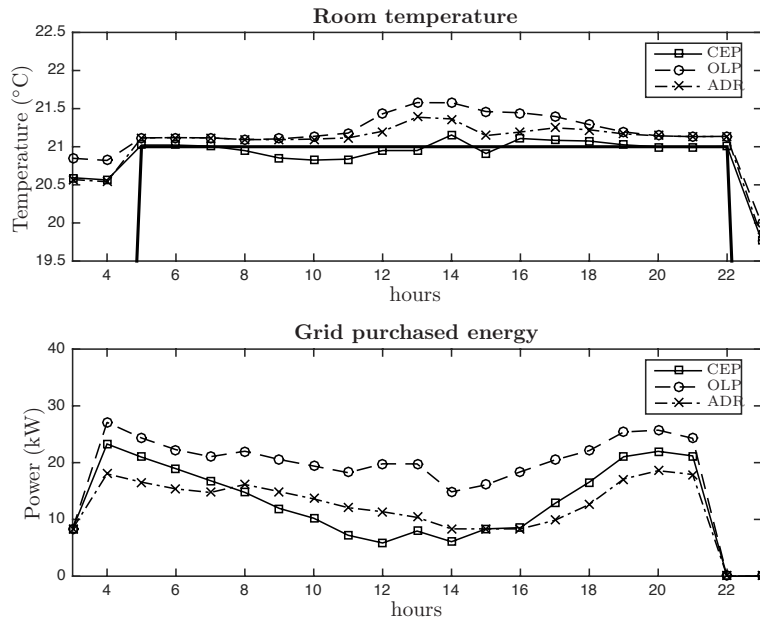

FIGURE 4. Profile trajectories for the mean room temperature of Building 1, and the total purchased grid energy by the community during a typical winter day.

To better visualize the behavior of the three approximations, the trajectories for the mean room temperature of Building 1 (c.f. Table 1), and the amount of purchased grid energy, are depicted in Figure 4 for Monday, 1st January 2007. The CEP method operates very near comfort constraints and leads to frequent

violations, while the OLP produces more conservative results by keeping the room temperature well inside the comfort range at the expense of consumption cost benefits. The poor performance of the CEP method with respect to the constraint violations can be explained when considering that the approximation assumes a deterministic evolution of the uncertain parameters. On the other hand, the adaptive nature of the ADR produces a good compromise between the optimistic decisions made by the CEP and the conservative decisions of the OLP. We remark that all three approximations utilize the load shifting capabilities of the battery by storing energy during the evening hours when electricity is cheaper, and deploying that energy in the early morning hours when the building needs to be brought back within the comfort range. As expected the energy produced from the photovoltaic unit is fully exploited to either reduce the grid energy consumption or charge the battery.

5.5. Comparison of ADR and tuned-CEP methods

We compare the stochastic ADR method with the deterministic *tuned-CEP* one. We denote by tuned-CEP, the CEP method in which the comfort bounds of the building, $i \in \mathcal{B}$, at time $t \in \mathcal{T}$, have been tightened by a fixed constant $c_b \geq 0$ Celcius degrees, as follows,

$$\min\{\boldsymbol{x}_{t,i} - \mathrm{lb}_{t,i}, \ \mathrm{ub}_{t,i} - \boldsymbol{x}_{t,i}\} \geq c_b.$$

This constant c_b is computed as the minimum constraint tightening value for which the ADR and CEP methods achieve the same amount of comfort bounds violations over the simulation horizon. The tuned-CEP can be seen as a practical ad-hoc solution to deal with the increased constraint violations of the CEP method.

	Winter		
Method	Cost (p.u.)	Tightening c_b (°C)	Basis (CHF/Week)
CEP	(1.00, 0.21)	[0.38, 0.56]	327.56
ADR	(0.94, 0.19)	0	

TABLE 6. Comparison of ADR and tuned-CEP methods.

We conducted receding horizon simulations for 12 consecutive weeks of the winter 2007, and we report in Table 6 the empirical mean and standard deviation of the recorded weekly costs. For each of these weeks, we computed by repetitive trials the respective constraint tightening constant, c_b, which ranges from 0.38°C and 0.56°C. The fact that c_b varies considerably from week to week, also suggests that tuning the CEP method is not a straightforward procedure. Additionally, we observe that the cost of purchased energy associated with the tuned-CEP solution method is considerably higher than the ADR one. Therefore, the inherent ability of the ADR method to deal with the uncertainty allows it to achieve an efficient trade-off between constraint violations and energy cost without the extra effort of tuning.

5.6. District energy benefit

Finally, we investigate the potential energy cost savings that can be obtained by considering diverse, heterogeneous buildings in a district. We define the heterogeneity on buildings by means of operational and construction characteristics. Specifically, we distinguish two operation types for the buildings, commercial (COM) and residential (RES). The day-night room temperature bounds defining these types are summarized in Table 7. We select building (Bd) 1 and 3 of Table 1 to be classified as commercial, while building 2 and 4 as residential. Notice that these buildings also differ in their construction characteristics.

Time	COM Bounds		RES Bounds	
	Lower	Upper	Lower	Upper
06:00 - 09:00	15°C	30°C	21°C	25°C
09:00 - 19:00	21°C	25°C	15°C	30°C
19:00 - 23:00	15°C	30°C	21°C	25°C
23:00 - 06:00	15°C	30°C	15°C	30°C

TABLE 7. Comfort constraints bounds for commercial and residential buildings.

We conduct five receding horizon simulations during the first week of January 2007. In each of these experiments, we consider a different combination of residential and commercial buildings from Table 1, and we report the cost of purchased grid energy for the CEP, ADR and OLP methods in Table 8. Notice that in each simulation experiment, we appropriately scale the hub devices with respect to the number of buildings that are connected to the energy hub.

	ADR	OLP	CEP	Basis (CHF/Week)
RES (Bd 2)	1.13	1.39	1.00	56.37
RES (Bd 4)	1.15	1.41	1.02	
COM (Bd 1)	1.28	1.47	1.23	
COM (Bd 3)	1.31	1.49	1.26	
RES+RES (Bd 2+4)	2.27	2.80	2.02	
COM+COM (Bd 1+3)	2.58	2.96	2.49	
COM+RES (Bd 1+2)	**1.51**	**1.90**	**1.37**	

TABLE 8. Cost of purchased energy for various district configurations during a typical winter week.

We observe that the cost benefits obtained by merging buildings with different construction characteristics but similar operation plans can be limited. On the contrary, significant gains are obtained by aggregating buildings with dissimilar operation plans. There are two reasons for this: (*i*) The energy shifting mainly occurs among buildings with complementary demand profiles. In this occasion, the efficient, but of limited capacity, devices of the hub (e.g., heat pump) are fully utilized during the course of day; (*ii*) The energy produced by the photovoltaics units is better exploited through the storage when the peak demands of buildings do not coincide over time. This flexibility on the demand profiles results to a considerably less purchased energy from the grid.

6. CONCLUSION

This paper presents a unified data-driven control framework for the problem of cooperatively managing the aggregated energy demands of buildings in a district. It exploits the available historical data to train linear models with additive uncertainties that effectively capture the evolution of the stochastic processes in the system. The underlying distributions of these additive uncertainties are shown to belong to Gaussian families of distributions which are constructed off-line using the historical data information. We exploit the simple structure of these sets to approximate the resulting robust stochastic optimization problem by a finite dimensional linear program. This program is tractable and its complexity scales polynomially. This is particularly important when addressing large scale problems such as cooperative building energy management. An extensive simulation study based on realistic data demonstrates the efficacy of the proposed method.

Among other things, our study demonstrates that additional cost benefits can be obtained by aggregating buildings with heterogeneity in the demand profiles.

As future work, we note that our problem has a decoupled structure. Indeed, the linear structure of the objective and and the weakly coupled structure of the constraints can be exploited in a number of distributed and decentralized optimization algorithms. Additionally, the problem's decoupled structure can be exploited by recently developed optimization algorithms such as the alternating direction method of multipliers (ADMM), for a fast numerical solution of the linear optimization problem.

Acknowledgments

Appendix

Proof of Proposition 1

We construct the family of distributions, $\mathcal{P}_t$ as the intersection of (i) the family, $\mathcal{P}_t^\chi$, of Gaussian distributions with unknown mean and bounds on variance, and (ii) the family, $\mathcal{P}_t^{st}$, of Gaussian distributions with unknown variance and bounds on the mean, as follows:

$$\mathcal{P}_t = \{\mathbb{P}_t \text{ such that } (\mathbb{P}_t \in \mathcal{P}_t^\chi) \wedge (\mathbb{P}_t \in \mathcal{P}_t^{st})\},$$

where,

$$\mathcal{P}_t^\chi = \{\mathbb{P}_t \text{ is Gaussian and condition (9b) is satisfied}\},$$

$$\mathcal{P}_t^{st} = \{\mathbb{P}_t \text{ is Gaussian and condition (9c) is satisfied}\}.$$

Given the data, $\mathcal{S}_t$, equations (9b) and (9c) also provide the confidence levels associated with the respective ambiguity sets,

$$\mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t^\chi) \geq 1 - \delta_t^\chi,$$

$$\mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t^{st} \,|\, \mathbb{P}_t \in \mathcal{P}_t^\chi) \geq 1 - \delta_t^{st}.$$

Notice that the bounds in (9c) are derived under the assumption that the sample variance is following the chi-square distribution. Therefore, the confidence level associated with the construction of the distribution family, $\mathcal{P}_t^{st}$, is given by the conditional probability.

We use Bayes rule to evaluate the significance level, δ_t, associated with the ambiguity set $\mathcal{P}_t$, as follows:

$$\begin{aligned}
\mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t) &= \mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t^\chi \wedge \mathbb{P}_t \in \mathcal{P}_t^{st}) \\
&= \mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t^{st} \,|\, \mathbb{P}_t \in \mathcal{P}_t^\chi) \cdot \mathbf{P}_{\mathcal{S}_t}(\mathbb{P}_t \in \mathcal{P}_t^\chi) \\
&\geq (1 - \delta_t^{st})(1 - \delta_t^\chi) \\
&= 1 - \delta_t^{st} - \delta_t^\chi + \delta_t^{st}\delta_t^\chi \\
&\geq 1 - \delta_t^{st} - \delta_t^\chi \\
&= 1 - \delta_t.
\end{aligned}$$

with the significance level, δ_t, given as, $\delta_t = \delta_t^{st} + \delta_t^\chi$.   ■

Proof of Proposition 2

The uncertain vector $\boldsymbol{w}$ comprises $|\mathcal{T}||\mathcal{D}|$ uncorrelated components. Therefore, we seek a bounded set $\widehat{W}$ that is composed by $2|\mathcal{T}||\mathcal{D}|$ inequalities that upper and lower bound every component $i = 1, \ldots, |\mathcal{T}||\mathcal{D}|$, of $\boldsymbol{w}$. Denoting by $\overline{w}_{t,i}$ and $\underline{w}_{t,i}$ the upper and lower bound, respectively, the set $\widehat{W}$ can be written as,

$$\widehat{W} = \left\{ \boldsymbol{w} \in \mathbb{R}^{|\mathcal{T}||\mathcal{D}|} : \underline{w}_{t,i} \leq w_{t,i} \leq \overline{w}_{t,i}, \, \forall t \in \mathcal{T}, \, i \in \mathcal{D} \right\}.$$

We require that,

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\boldsymbol{w} \in \widehat{W}) \geq 1 - \epsilon \,.$$

We exploit the conservative Bonferonni approximation [20] to decouple the joint chance constraints into a set of individual ones, as follows:

$$\mathbb{P}(\boldsymbol{w} \notin \widehat{W}) = \sum_{i \in \mathcal{D}} \sum_{t \in \mathcal{T}} \left(\mathbb{P}(w_{t,i} \geq \overline{w}_{t,i}) + \mathbb{P}(w_{t,i} \leq \underline{w}_{t,i}) \right)$$
$$= \sum_{i \in \mathcal{D}} \sum_{t \in \mathcal{T}} \left(\overline{\beta}_{t,i} + \underline{\beta}_{t,i} \right) = \epsilon,$$

with $\overline{\beta}_{t,i}, \underline{\beta}_{t,i} \geq 0$. Notice that in this case the Bonferonni approximation is exact since we are considering mutually exclusive events (i.e. $\mathbb{P}(w_{t,i} \geq \overline{w}_{t,i} \wedge w_{t,i} \leq \underline{w}_{t,i}) = 0$).

Assume that $w_{t,i} \sim \mathcal{N}(\mu_{t,i}, \sigma_{t,i}^2)$, then the chance constraints,

$$\mathbb{P}(w_{t,i} \leq \overline{w}_{t,i}) \geq 1 - \overline{\beta}_{t,i},$$
$$\mathbb{P}(w_{t,i} \geq \underline{w}_{t,i}) \geq 1 - \underline{\beta}_{t,i},$$

are equivalently reformulated (see [23]), as follows,

$$w_i \geq \mu_i - \Phi^{-1}\left(1 - \underline{\beta}_{t,i}\right) \cdot \sigma_i,$$
$$w_i \leq \mu_i + \Phi^{-1}\left(1 - \overline{\beta}_{t,i}\right) \cdot \sigma_i.$$

In this setting, the robustified individual chance constraints,

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(w_{t,i} \leq \overline{w}_{t,i}) \geq 1 - \overline{\beta}_{t,i},$$
$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(w_{t,i} \geq \underline{w}_{t,i}) \geq 1 - \underline{\beta}_{t,i},$$

are equivalently be reformulated as follows,

$$\left. \begin{array}{l} w_{t,i} \geq \mu_{t,i} - \Phi^{-1}\left(1 - \underline{\beta}_{t,i}\right) \cdot \sigma_{t,i} \\ w_{t,i} \leq \mu_{t,i} + \Phi^{-1}\left(1 - \overline{\beta}_{t,i}\right) \cdot \sigma_{t,i} \end{array} \right\} \quad \begin{array}{l} \forall \underline{\mu}_{t,i} \leq \mu_{t,i} \leq \overline{\mu}_{t,i}, \\ \forall \underline{\sigma}_{t,i} \leq \sigma_{t,i} \leq \overline{\sigma}_{t,i}. \end{array}$$

Exploiting the linearity of these constraints, we can explicitly compute the upper and lower bounds, as follows,

$$\underline{w}_{t,i} = \underline{\mu}_{t,i} - \Phi^{-1}\left(1 - \underline{\beta}_{t,i}\right) \cdot \overline{\sigma}_{t,i},$$
$$\overline{w}_{t,i} = \overline{\mu}_{t,i} + \Phi^{-1}\left(1 - \overline{\beta}_{t,i}\right) \cdot \overline{\sigma}_{t,i}.$$

This analysis provides the constructive proof for the structure of $\widehat{W}$ in (15) ∎

Dynamics and constraints of the energy hub devices

We choose to present the devices characteristics sized per building (Bd). In particular, we model the chiller, boiler and heat pump using a coefficient of performance [12], which gives rise to the following constraints:

$$\boldsymbol{u}_{t,\text{chiller}}^{\text{out}} = 0.7 \boldsymbol{u}_{t,\text{chiller}}^{\text{in}},$$
$$\boldsymbol{u}_{t,\text{boiler}}^{\text{out}} = 0.9 \boldsymbol{u}_{t,\text{boiler}}^{\text{in}},$$
$$\boldsymbol{u}_{t,\text{HP}}^{\text{out}} = 3 \boldsymbol{u}_{t,\text{HP}}^{\text{in}},$$

where $\boldsymbol{u}_{t,i} = [\boldsymbol{u}_{t,i}^{\text{in}}, \boldsymbol{u}_{t,i}^{\text{out}}]$ with $\boldsymbol{u}_{t,i}^{\text{in}}, \boldsymbol{u}_{t,i}^{\text{out}}$, denoting the input and output power flows on the i-th device, respectively. The capacities of the conversion units are given as follows,

$$0 \le \boldsymbol{u}_{t,\text{chiller}}^{\text{out}} \le 20 \, \text{kW/Bd},$$
$$0 \le \boldsymbol{u}_{t,\text{boiler}}^{\text{out}} \le 25 \, \text{kW/Bd}$$
$$0 \le \boldsymbol{u}_{t,\text{HP}}^{\text{out}} \le 5 \, \text{kW/Bd},$$

We consider a South oriented photovoltaic array with maximum output of $4.10 \, \text{kW/Bd}$. The photovoltaic dynamics are generated by linearizing the non-linear model of [42] with respect to the ambient temperature and the solar radiation,

$$0 \le \boldsymbol{u}_{t,\text{PV}}^{\text{out}} \le 0.1280 - 0.0019 \boldsymbol{\xi}_{t,\text{AT}} + 3.7 \boldsymbol{\xi}_{t,\text{SRS}},$$

for all $t \in \mathcal{T}$. Note that the units of $\boldsymbol{\xi}_{t,\text{SRS}}$ are measured in kW/m^2 and can typically take values $\boldsymbol{\xi}_{t,\text{SRS}} \in [0, 1]$.

We consider a lead-acid battery [43], with a 5kW/Bd capacity, giving rise to the following linear dynamical system:

$$\boldsymbol{x}_{t+1} = \begin{pmatrix} 0.51 & 0.22 \\ 0.47 & 0.78 \end{pmatrix} \boldsymbol{x}_t + \begin{pmatrix} 0.61 \\ 0.25 \end{pmatrix} \boldsymbol{u}_t^{\text{in}} + \begin{pmatrix} -0.83 \\ -0.39 \end{pmatrix} \boldsymbol{u}_t^{\text{out}},$$

where the states and control are constrained by:

$$0 \le \boldsymbol{u}_t^{\text{in}}, \boldsymbol{u}_t^{\text{out}} \le 8,$$
$$1 \le \boldsymbol{x}_{t,1} + \boldsymbol{x}_{t,2} \le 5, \; \boldsymbol{x}_t \ge 0,$$
$$0.62 \boldsymbol{x}_{t,1} + 0.27 \boldsymbol{x}_{t,2} - \boldsymbol{u}_t^{\text{out}} \ge 0,$$
$$0.84 \boldsymbol{x}_{t,1} + 0.37 \boldsymbol{x}_{t,2} + \boldsymbol{u}_t^{\text{in}} \le 2.58,$$
$$0.73 \boldsymbol{x}_{t,1} + 0.73 \boldsymbol{x}_{t,2} + \boldsymbol{u}_t^{\text{in}} \le 3.66.$$

To this end, we provide the equations describing the interconnection of the energy hub with the building community. In particular, the electricity balancing constraint is given by,

$$\boldsymbol{p}_{t,\text{elect}} + \sum_{i \in \text{E}_+} \boldsymbol{u}_{t,i}^{\text{out}} = \sum_{i \in \text{E}_-} \boldsymbol{u}_{t,i}^{\text{in}} + \boldsymbol{d}_{t,\text{elect}},$$

where $\text{E}_+ = \{\text{Photovoltaics, Battery}\}$ and $\text{E}_- = \{\text{Heat Pump, Chiller, Boiler, Battery}\}$. Similarly, the heating and cooling energy balancing constraints are given by,

$$\sum_{i \in \text{H}_+} \boldsymbol{u}_{t,i}^{\text{out}} = \boldsymbol{d}_{t,\text{heat}}, \; \text{and} \; \boldsymbol{u}_{t,\text{chiller}}^{\text{out}} = \boldsymbol{d}_{t,\text{cool}},$$

where $\text{H}_+ = \{\text{Heat Pump, Boiler}\}$. Finally, the demand for electricity, heating and cooling of the building community is given by

$$\boldsymbol{d}_{t,\text{elect}} = \sum_{i \in \mathcal{B}} \boldsymbol{u}_{t,i,\text{AHU}},$$
$$\boldsymbol{d}_{t,\text{cool}} = \sum_{i \in \mathcal{B}} \boldsymbol{u}_{t,i,\text{TABS}},$$
$$\boldsymbol{d}_{t,\text{heat}} = \sum_{i \in \mathcal{B}} \left(\boldsymbol{u}_{t,i,\text{radiator}} + \boldsymbol{u}_{t,i,\text{TABS}} \right),$$

where TABS are typically used both for heating and cooling of the buildings.

References

[1] L. Perez-Lombard, J. Ortiz, and C. Pout. A review on buildings energy consumption information. *Energy Buildings*, 40(3):394–398, 2008.

[2] I. Sartori, A. Napolitano, and K. Voss. Net zero energy buildings: A consistent definition framework. *Energy Buildings*, 48(1):220–232, 2012.

[3] L. Deori, L. Giulioni, and M. Prandini. Optimal building climate control: A solution based on nested dynamic programming and randomized optimization. In *Proc. IEEE Conf. Decision Control*, pages 4905–4910, Los Angeles, California, USA, 2014.

[4] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin. Reducing transient and steady state electricity consumption in HVAC using learning-based model-predictive control. *Proc. IEEE*, 100(1):240–253, 2012.

[5] M. Maasoumy, M. Razmara, M. Shahbakhti, and A. Vincentelli. Handling model uncertainty in model predictive control for energy efficient buildings. *Energy Buildings*, 77(1):377–392, 2014.

[6] F. Oldewurtel, D. Sturzenegger, and M. Morari. Importance of occupancy information for building climate control. *Appl. Energy*, 101(1):521–532, 2013.

[7] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy Buildings*, 45(1):15–27, 2012.

[8] J. Siroky, F. Oldewurtel, J. Cigler, and S. Privara. Experimental analysis of model predictive control for an energy efficient building heating system. *Appl. Energy*, 88(9):3079–3087, 2011.

[9] D. Sturzenegger, D. Gyalistras, M. Morari, and R. S. Smith. Model predictive climate control of a Swiss office building: Implementation, results and cost-benefit analysis. *IEEE Trans. Control Syst. Technol.*, 24(1):1–12, 2016.

[10] M. Geidl, G. Koeppel, P. Favre-Perod, B. Klockl, G. Andersson, and K. Frohlich. Energy hubs for the future. *IEEE Power Energy Mag.*, 5(1):24–30, 2007.

[11] D. Crawley, L. Lawrie, C. Pedersen, and F. Winkelmann. Energy plus: Energy simulation program. *ASHRAE J.*, 42(4):49–56, 2000.

[12] R. Evins, K. Orehounig, V. Dorer, and J. Carmeliet. New formulations of the energy hub model to address operational constraints. *Energy*, 73(1):387–398, 2014.

[13] E. Fabrizio, M. Filippi, and J. Virgone. An hourly modelling framework for the assessment of energy sources exploitation and energy converters selection and sizing in buildings. *Energy Buildings*, 41(10):1037–1050, 2009.

[14] A. Parisio, C. Del Vecchio, and A. Vaccaro. A robust optimization approach to energy hub management. *Int. J. Elec. Power Energy Sys.*, 42(1):98–104, 2012.

[15] D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Rev.*, 53(3):464–501, 2011.

[16] X. Zhang, S. Grammatico, G. Schildbach, P. Goulart, and J. Lygeros. On the sample size of randomized mpc for chance-constrained systems with application to building climate control. In *Proc. Eur. Control Conf.*, pages 478–483, Strasbourg, France, 2014.

[17] Jitka Dupačová, Nicole Gröwe-Kuska, and Werner Römisch. Scenario reduction in stochastic programming. *Math. Program.*, 95(3):493–511, 2003.

[18] G. Calafiore and M. Campi. The scenario approach to robust control design. *IEEE Trans. Autom. Control*, 51(5):742–753, 2006.

[19] A. Shapiro. Analysis of stochastic dual dynamic programming method. *Eur. J. Oper. Res.*, 209(1):63–72, 2011.

[20] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM J. Optim.*, 17(4):969–996, 2006.

[21] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Oper. Res.*, 62(6):1358–1376, 2014.

[22] B. Van Parys, D. Kuhn, P. Goulart, and M. Morari. Distributionally robust control of constrained stochastic systems. *IEEE Trans. Autom. Control*, 61(2):430–442, 2016.

[23] G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. Jordan. A robust minimax approach to classification. *J. Mach. Learn. Res.*, 3(1):555–582, 2003.

[24] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. arXiv:1401.0212, 2013.

[25] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, 58(3):595–612, 2010.

[26] H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of semidefinite programming: Theory, Algorithms, and Applications.* Springer Science & Business Media, 2012.

[27] G. Darivianakis, A. Georghiou, R. S. Smith, and J. Lygeros. A stochastic optimization approach to cooperative building energy management via an energy hub. In *Proc. IEEE Conf. Decision Control*, pages 7814–7819, Osaka, Japan, 2015.

[28] D. Sturzenegger, D. Gyalistras, V. Semeraro, M. Morari, and R. S. Smith. BRCM Matlab Toolbox: Model generation for model predictive building control. In *Proc. Amer. Control Conf.*, pages 1063–1069, Portland, USA, 2014.

[29] Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics. European committee for standardization, 2007.

[30] J. Steppeler, G. Doms, U. Schättler, H.W. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric. Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, 82(1-4):75–96, 2003.

[31] E. Lehmann and J. Romano. *Testing Statistical Hypotheses*. Springer, 2006.

[32] A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Math. Program.*, 152(1):301–338, 2014.

[33] E. Kerrigan and J. Maciejowski. Soft constraints and exact penalty functions in model predictive control. In *Proc. UKACC Int. Conf. Control*, Cambridge, UK, 2000.

[34] G. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *J. Optim. Theory App.*, 130(1):1–22, 2006.

[35] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Math. Program.*, 99(2):351–376, 2004.

[36] P. Benner and T. Damm. Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems. *SIAM J. Control Optim.*, 49(2):686–711, 2011.

[37] D. Gyalistras, A. Fischlin, M. Morari, C. Jones, F. Oldewurtel, A. Parisio, F. Ullmann, C. Sagerschnig, and A. Gruner. Final report: Use of weather and occupancy forecasts for optimal building climate control (OptiControl). Technical report, ETH Zurich, 2010.

[38] N. Razali and Y. Wah. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Analyt.*, 2(1):21–33, 2011.

[39] St. Jewson and R. Caballero. Seasonality in the statistics of surface air temperature and the pricing of weather derivatives. *Meteorol. App.*, 10(04):367–376, 2003.

[40] J. Page, D. Robinson, N. Morel, and J. L. Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy Buildings*, 40(2):83–98, 2008.

[41] David L Wallace. Asymptotic approximations to distributions. *Ann. Math. Stat.*, 29(3):635–654, 1958.

[42] H. Fakham, P. Degobert, and B. François. Control system and power management for a PV based generation unit including batteries. In *Int. Aegean Conf. Elect. Mach. Power Electron.*, pages 141–146, Bodrum, Turkey, 2007.

[43] E. I. Vrettos and S. A. Papathanassiou. Operating policy and optimal sizing of a high penetration RES-BESS system for small isolated grids. *IEEE Trans. Energy Convers.*, 26(3):744–756, 2011.