Model Predictive Control Based on Stochastic Grey-Box Models

Christian Ankerstjerne Thilker, Rune Grønborg Junker, Peder Bacher, John Bagterp Jørgensen, and Henrik Madsen

Acronyms

ACF	Autocorrelation function
AIC	Akaike's information criterion
AR	Autoregressive
CCF	Cross-correlation function
CDEKF	Continuous–discrete extended Kalman filter
CTSM-R	Continuous-time stochastic modelling in R
EKF	Extended Kalman filter
EV	Electrical vehicle
FF	Flexibility function
FI	Flexibility index
HMM	Hidden Markov model
KF	Kalman filter
LRT	Likelihood-ratio test
LS	Least squares
MLE	Maximum likelihood method
MPC	Model predictive control
PI	Proportional integral
PID	Proportional integral derivative

C. A. Thilker (✉) · R. G. Junker · P. Bacher · J. B. Jørgensen
Technical University of Denmark, Lyngby, Denmark
e-mail: chant@dtu.dk; rung@dtu.dk; pbac@dtu.dk; jbjo@dtu.dk

H. Madsen
Technical University of Denmark, Lyngby, Denmark

Norwegian University of Science and Technology, Trondheim, Norway
e-mail: hmad@dtu.dk; henrik.madsen@ntnu.no

S. Ploix et al. (eds.), *Towards Energy Smart Homes*,
https://doi.org/10.1007/978-3-030-76477-7_11

PV	Photo-voltaic cells
SDE	Stochastic differential equation
SOC	State of charge
RMS	Root mean square

1 Introduction

If the proliferation of renewable energy sources is to be continued, solutions for the related problems have to be implemented. The problems include, but are not limited to, mismatch in generation and load, voltage deviation, congestion, and demand ramps. While historically these problems were manageable through control of the generation, this will not be an option in the future, as a majority of power generation will be coming from intermittent renewable energy sources. On the other hand, recent advances and adoption of digital solutions and smart devices present new opportunities for smart energy *demand* [33, 44, 49, 50], by utilising the inherent *energy flexibility* [25]. With buildings accounting for around 40% of energy demand [31], they have been identified as key assets in this context [34]. However, to actually make buildings smart and unlock the inherent energy flexibility, suitable methods for controlling them have to be employed. While smart buildings seek to resolve high-level problems, there remains the dilemma that the buildings themselves are subject to decentralised and independent control and given over to the controllers commissioned by the building owners. To deal with this, it has been proposed to use a two-level control hierarchy [12, 17, 58] in which the upper level consists of controllers that formulate price signals. The price signals are then sent to the lower-level controllers that are controlling energy flexible systems such as smart buildings. The objective of the lower-level controllers is to minimise costs that, if the price signals are formulated correctly, also solve the grid problems [19, 29]. A generalisation of this hierarchical setup of nested controllers is described as the smart-energy operating system (SE-OS) in [36, 44, 45].

For smart control of buildings, both one-way and two-way communication setups are used—often referred to as indirect and direct control, respectively. The simplest and most resilient setup is achieved by one-way communication where a price signal is sent to a group of buildings in a certain part of the grid. In the paper [12], it is shown how consumers, which are sensitive to varying prices, can be used to control the electricity load using a one-way price signal. Estimation of the price response is based on data measurable at grid level, removing the need to install sensors and communication devices between each individual consumer and the price-generating entity.

A sizeable list of examples of smart control of buildings from EBC's Annex 67 project can be found in [24] with details on the central control strategies presented in [56]. The potential of the energy flexibility of buildings was thoroughly investigated in this project, and it was found that by applying suitable control methods, it was found that the suitable application of control methods exposed enormous potential for energy flexibility in buildings. Another important project in regards to this is

the SmartNet project, which, together with the CITIES project, demonstrated the potential of automatic energy flexible control for a number of buildings with an indoor swimming pool [4, 36].

In this chapter, special focus will be put on how to formulate extended and advanced disturbance models in such a way that short-term forecasts are well described by them—the most important disturbances to consider for smart control of buildings are weather-related.

The models will be formulated as stochastic grey-box models. This modelling framework bridges the gap between physical and statistical modelling that makes it possible to combine knowledge from physics and statistics in an optimal way. The grey-box models will be formulated using *discretely observed stochastic differential equations* written down as *continuous–discrete-time stochastic state-space models*. In statistics, such models are also called *continuous state-space Hidden Markov models*.

Grey-box models are typically rather simple models in terms of physics, but they are formulated with emphasis on the stochastic part of the models. This implies that we will be able to use rigorous statistical model techniques and that the models enable for an efficient use of online sensors for control and forecasting. This modelling framework has been used to describe the thermal dynamics of buildings [5, 9, 26, 27, 40] and energy systems in many control-oriented projects [16].

First, we will describe the grey-box modelling framework in Sect. 2. Next, we shall describe some examples of grey-box models for buildings and smart building-related components in Sect. 4. This includes models for heat pumps, stationary, and mobile batteries (EVs).

The states of a building, e.g., the indoor air temperature, are heavily influenced by the weather conditions, and a special focus of this chapter is to establish rather simple stochastic models for the most important weather variables. In relation to control, the weather acts as a disturbance, and in order to obtain the best possible controllers, it is important that the controllers are able to take advantage of short-term forecasts of the disturbances. Models for the most relevant disturbance variables for control of buildings are described in Sect. 5.

The theory for model predictive control is outlined in Sect. 6. A special emphasis is put on how models for predicting the weather variables can be integrated into the concepts of model predictive control, and this is the topic of Sect. 7.

2 Grey-Box Models

The models in this chapter are based on the *grey-box* modelling framework. This framework is typically based on a non-linear model with a partial theoretical structure and some unknown parts derived from data. Consequently, the grey-box framework bridges the gap between models based on first principle (*white-box* models) and models based solely on data (*black-box* models) (Fig. 1).

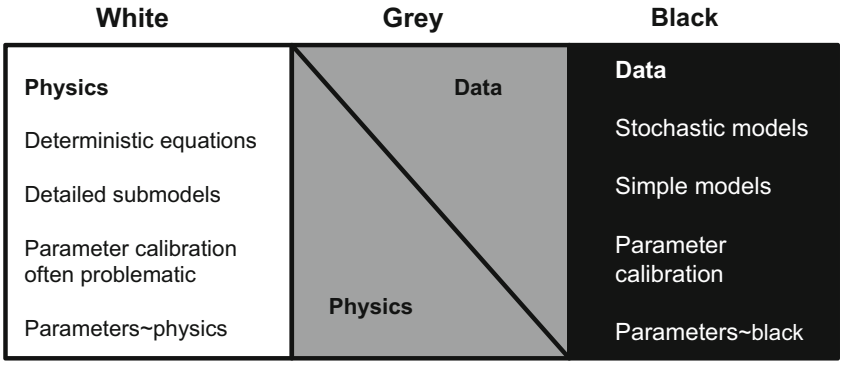

Fig. 1 Grey-box modelling bridges the gap between white- and black-box modelling

Grey-box models are formulated as a state-space model where the dynamics of the states is described in continuous time by a set of stochastic differential equations (SDEs) (*system equations*). The discrete-time observations are related to the states by a set of static equations (*observation equations*). Hence, a grey-box is formulated as continuous–discrete-time stochastic state-space model in the form

$$\mathrm{d}\boldsymbol{x}(t) = \underbrace{f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t), t)\mathrm{d}t}_{\text{Drift}} + \underbrace{g(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t), t)\mathrm{d}\boldsymbol{\omega}(t)}_{\text{Diffusion}} , \tag{1}$$

$$\boldsymbol{y}_k = h(\boldsymbol{x}(t_k)) + \boldsymbol{v}_k , \qquad \boldsymbol{v}_k \sim N(\boldsymbol{0}, \mathsf{R}_\mathsf{v}) , \tag{2}$$

where $\boldsymbol{x}$ is the system vector, $\boldsymbol{\omega}$ is a standard Wiener process (also often called a Brownian motion), and f and g are the drift and diffusion functions, respectively. h is the observation function and $\boldsymbol{v}_k$ is the observation noise. The drift function is the deterministic part of the SDE, whereas the diffusion function describes all the uncertainties not properly described in the drift.

If the system in (1)–(2) is linear, the model is written as

$$\mathrm{d}\boldsymbol{x}(t) = (\mathsf{A}\boldsymbol{x}(t) + \mathsf{B}\boldsymbol{u}(t) + \mathsf{E}\boldsymbol{d}(t))\,\mathrm{d}t + \Sigma\mathrm{d}\boldsymbol{\omega}(t) , \tag{3}$$

$$\boldsymbol{y}_k = \mathsf{C}\boldsymbol{x}(t_k) + \boldsymbol{v}_k , \qquad \boldsymbol{v}_k \sim N(\boldsymbol{0}, \mathsf{R}_\mathsf{v}) , \tag{4}$$

where A, B, E, C, and Σ are matrices governing the state evolution, input, disturbance, observation, and noise, respectively.

Modelling physical systems using SDEs provides a natural method to represent the phenomenon as it evolves in continuous time. In contrast to discrete-time models, prior physical knowledge about the system can rather easily be included, and the estimated parameters do not depend on the sampling time.

There are many reasons for introducing the system noise (the diffusion term):

- *Modelling approximations*. For example, the dynamics, as described by the drift term, might be an approximation to the true system.
- *Unrecognised and unmodeled inputs*. Some variables that are not considered, such as wind speed, may affect the system.
- *Noise in measurements of input variables*. In such cases the measured input signals are regarded as the actual input to the system, and the deviation from the true input is described by the noise term.

In the observation equation, a noise term is also introduced. The reason for this noise term is:

- *Noise in measurements of output variables*. The sensors that measure the output signals are affected by noise and drift.

It seems reasonable to assume that the system noise and the measurement noise are independent.

This chapter focuses on simple grey-box models, describing the heat dynamics of a building and related components such as a heat pump and batteries (stationary and mobile). The main purpose is to describe the dynamics of the building and relevant components. In particular we shall focus on how the heat dynamics are affected by outdoor climate.

2.1 A Simple Linear Grey-Box Model

Let us consider a simple second-order grey-box model for the thermal dynamics of a building ([40]). Here, the so-called RC formulation is used and the thermal capacity is lumped into two states, and each of these states has an associated thermal mass.

$$\begin{bmatrix} dT_m \\ dT_i \end{bmatrix} = \begin{bmatrix} \frac{-1}{r_i c_m} & \frac{1}{r_i c_m} \\ \frac{1}{r_i c_i} & -\left(\frac{1}{r_a c_i} + \frac{1}{r_i c_i}\right) \end{bmatrix} \begin{bmatrix} T_m \\ T_i \end{bmatrix} dt + \begin{bmatrix} 0 & 0 & A_w p/c_m \\ 1/(r_a c_i) & 1/c_i & A_w(1-p)/c_i \end{bmatrix} \begin{bmatrix} T_a \\ \phi_h \\ \phi_s \end{bmatrix} dt + \begin{bmatrix} d\omega_m(t) \\ d\omega_i(t) \end{bmatrix} . \quad (5)$$

$$\boldsymbol{T}_r(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \boldsymbol{T}(t) + \boldsymbol{e}(t) . \quad (6)$$

The states of the model are given by the temperature T_m of a large heat accumulating medium with the heat capacity c_m and by the temperature T_i of the room air and possibly the inner part of the walls with the capacity c_i. The term r_i is the resistance against heat transfer between the room air and the large heat accumulating medium, while r_a is the resistance against heat transfer from the room air to the ambient air with the temperature T_a.

The input energy is supplied by the electrical heaters ϕ_h and the solar radiation that penetrates through the windows facing south $A_w\phi_s$, where A_w is the effective window area. The effective window area is the window area corrected for shade effects and absorption and reflection by the triple glazed windows. Note that only the indoor air temperature is measured.

This model has been identified in [40]. It is concluded that for the considered building, this second-order model provides a good description of all the variations in the data since the residuals are white noise. In that paper, it is described how the parameters are estimated using a maximum likelihood method, and furthermore, it was concluded that all the solar radiation is influencing the indoor air since the assumption $p = 0$ seems reasonable (the parameter p was not significant).

3 Identification of Grey-Box Models

Formulating suitable grey-box models is an iterative process in which physical considerations are combined with information obtained by statistical observations. The typical starting point is to formulate the mathematical equations governing the most important physical dynamics. These equations are then used as the initial model. Next, the parameters of the model are estimated, and finally, the model is used to generate residuals. These residuals are key to the model validation step, and if it is concluded that the residuals still show systematic behaviour, then the residuals are analysed in order to identify how the model can be improved and extended.

3.1 Initial Model Structure Identification

Typically, the initial model order, i.e., the number of state equations, and the dominating structure of the model are determined by physics. However, also statistical methods are useful. For instance, it is well known that the autocorrelation and partial autocorrelation function contain important information about the order of (linear) models. The following step-by-step guide summarises the procedure of formulating grey-box models:

1. Make a drawing of the physical system that includes the various methods for heat transfer (conductive, convective, and radiation).
2. Write down the mass and energy balance equations for the system.
3. Determine the *causality* of the system. Which time series data can be considered as input and which as output? For instance, for a building with feedback or controlled internal air temperature, the output could be the heat consumption, whereas for a building with no feedback, e.g., when the heating signal is determined by a PRBS signal, the internal air temperature could be the output.

4. Evaluate if any *non-linear* phenomena must be taken into account explicitly in the initial phase (later on statistical methods can be used for identifying non-linear phenomena). Such phenomena could be significantly influenced by wind speed, complicated glass construction, humidity, influence from rainfall, etc. Some non-linear effects can be described by a transformation of the input variables. In [54] the non-linear effect of solar radiation is described in a grey-box model using spline basis functions.
5. Evaluate if any *non-stationary* phenomena must be taken into account explicitly in the initial phase (later on statistical methods can be used for identifying non-stationary phenomena). Examples of such phenomena could be the fermentation of a new concrete building, moisture in the construction, opening of windows and doors, etc. For control applications, slowly varying non-stationary phenomena can be handled by considering adaptive and recursive methods [1, 38].

3.2 Estimation of Model Parameters

Typically, the model parameters are estimated either using the least squares method (LS) or the maximum likelihood estimation (MLE) method. The advantage of the MLE method is that this method also allows for estimating the parameters related to the noise term. Here, we briefly introduce the MLE method for estimating parameters in grey-box models. The method is described in detail in [30].

Given a sequence of measurements $\mathcal{Y}_N = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_N\}$, the likelihood function is the joint probability density of all the observations but considered as a function of the unknown parameters. Thus, the likelihood function can be written as the product of the one-step ahead conditional densities:

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}_N, \mathcal{U}_N) = \prod_{k=1}^{N} p(\boldsymbol{Y}_k|\boldsymbol{\theta}, \mathcal{Y}_{k-1}, \mathcal{U}_k) p(\boldsymbol{X}_0|\boldsymbol{\theta}) \,, \tag{7}$$

where $p(\boldsymbol{Y}_k|\boldsymbol{\theta}, \mathcal{Y}_{k-1}, \mathcal{U}_k)$ is the probability of observing $\boldsymbol{Y}_k$ given the previous observations, inputs, and set of parameters $\boldsymbol{\theta}$. This is the so-called exact likelihood function that contains a parameterisation of the density associated with the initial state $\boldsymbol{X}_0$.

Since the systems are assumed to be driven by Wiener processes for which the increments are Gaussian, the one-step ahead density for linear systems is also Gaussian. For most non-linear systems, this is still a reasonable assumption, and this assumption can be checked—see, e.g., [6].

In the Gaussian case, the conditional density is completely characterised by the conditional mean (the prediction) and the conditional covariance. By introducing the one-step prediction error (also called the *innovation error* or *residuals*)

$$\boldsymbol{\epsilon}_{k|k-1} = \boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1} \,, \tag{8}$$

and the associated covariance, $\mathsf{R}_{k|k-1} = \mathrm{Var}(\boldsymbol{Y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta})$, the likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_N, \mathcal{U}_N) = p(\mathcal{Y}_N|\mathcal{U}_N, \boldsymbol{\theta}) \tag{9}$$

$$= \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_{k|k-1}^{\top}\mathsf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_{k|k-1} \right)}{\sqrt{\det\left(\mathsf{R}_{k|k-1}\right)\left(\sqrt{2\pi}\right)^{\mathrm{L}}}} \right) p(\boldsymbol{X}_0|\boldsymbol{\theta}) \,, \tag{10}$$

where L is the dimension of the observation space. Using the logarithm, we obtain the log-likelihood function

$$l(\boldsymbol{\theta}; \mathcal{Y}_N, \mathcal{U}_N) = -\frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{\epsilon}_{k|k-1}^{\top}\mathsf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_{k|k-1} + \log\left(\det\left(\mathsf{R}_{k|k-1}\right)(2\pi)^{\frac{\mathrm{L}}{2}}\right)\right)$$
$$+ \log(p(\boldsymbol{X}_0|\boldsymbol{\theta})).$$

The parameter estimates are found by maximising the log-likelihood function

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{l(\boldsymbol{\theta}; \mathcal{Y}_N, \mathcal{U}_N)\right\}. \tag{11}$$

The corresponding value of the log-likelihood is the observed maximum likelihood value given the available data set.

For linear models, the conditional mean and covariance are calculated using an ordinary Kalman filter, while for non-linear models, an extended Kalman filter is used. See [30] for further details.

3.3 Uncertainty of Parameter Estimates

Uncertainty of parameter estimates is an essential output of any statistical parameter estimation scheme. This uncertainty lies in the facilitation of subsequent statistical tests. For the software implementation used here [28, 39], an estimate of the uncertainty of the parameter estimates is obtained by using the fact that by the central limit theorem the ML estimator is asymptotically Gaussian with mean θ and covariance:

$$\Sigma_{\hat{\theta}} = \mathsf{H}^{-1}, \tag{12}$$

where the matrix H is given by

$$h_{ij} = -E\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(l(\theta|\mathcal{Y}_N)\right)\right\}, i, j = 1, \ldots, p \,.$$

An approximation to H can be obtained from

$$h_{ij} \approx -\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(l(\theta|\mathcal{Y}_N)\right) \right)\Big|_{\theta=\hat{\theta}}, i, j = 1, \ldots, p \ ,$$

which is simply the Hessian evaluated at the maximum of the log-likelihood function. To obtain a measure of the uncertainty of the individual parameter estimates, the covariance matrix is decomposed as

$$\Sigma_{\hat{\theta}} = \sigma_{\hat{\theta}} \mathsf{R} \sigma_{\hat{\theta}} \ , \tag{13}$$

into $\sigma_{\hat{\theta}}$, which is a diagonal matrix of the standard deviations of the parameter estimates, and R, which is the corresponding correlation matrix.

3.4 Selection of Model Structure

Basically, the two main categories of problems related to the order of the model are:

1. **Model too simple:** A common problem is that *the residuals* for a given model are *autocorrelated*. In this case the model needs to be extended (for grey-box models, more states are needed). Another common problem is that the residuals are *cross-correlated* with some explanatory variables (e.g., large residuals for large wind speeds). In this case, this (or these) explanatory variable needs to be included in the model.
2. **Model too large:** A common problem is that some of the *parameters* are *insignificant*. In order to ensure a reliable estimation of the performance, the amount of parameters must be reduced by removing insignificant parameters.

3.5 Model Validation

If the residuals from a given modelling step show systematic variation, then the model is too simple and it can be improved. Consequently, model validation is a very important step in model building.

The following methodologies can be used in relation for model validation:

1. **Test for white noise residuals.**
   Typically, the autocorrelation function (ACF) of the residuals is used here. If a test for white noise residuals fails, then the model must be extended by extending the model order, which for grey-box models is the number of states.
2. **Test for dependency with inputs.**

Plot residuals against the inputs to see if any dependency exists. The cross-correlation function (CCF) (see [38, p. 230]) can be used to identify linear dependencies that have to be added to the model.

3. **Test for parameter significance.**
   See the next section on model validation. Here, it is mentioned that if a parameter is found to be insignificant, then in general this parameter should be removed from the model and the parameters of the reduced model are estimated.
4. **Check for correlation between parameters.**
   Most software for parameter estimation provides a correlation matrix of the estimated parameters. A numerically very high (say larger than .98) correlation between two parameter estimates indicates that one of these two parameters should be either excluded from the model or fixed to some physically assumed values.

3.6 Comparison of Models

1. **Test between (nested) models.**
   If two models are nested, i.e., the smaller model (B) can be found just by removing parts of a larger model (A), then the *likelihood-ratio test (LRT)* is very useful. The LRT value is given as $D = 2(\log L(A) - \log L(B))$, where $\log L(A)$ is the logarithm of the likelihood function for model A. For grey-box modelling, the asymptotic test principles based on Wilks' Theorem are used. Given that the model can be reduced to model B, the quantity D is according to Wilks' Theorem asymptotically $\chi^2(k-m)$ distributed, where k and m are the number of parameters in models A and B, respectively. For large values of D, it is concluded that the best model is the larger model. See, e.g., [42] for further details.
   In CTSM-R, the value $\log L$ is found using summary().
2. **Comparison between (non-nested) models**. If two models are non-nested, then methods based on *information criteria* can be used—see page 174 in [38]. This consists of computing an information criterion, such as the AIC or BIC:

$$AIC = 2k - 2\log L(A), \quad BIC = 2\log(N) - 2\log L(A)\ .$$

The preferred model is then simply found as the model with the lowest information criteria. Alternatively, and preferably, when a lot of data are available, cross-validation can be used [8]. In its simplest form, this procedure can be summarised as:

a. Split data into two parts, $\mathcal{Y}_{\text{Train}}$ and $\mathcal{Y}_{\text{Validate}}$. A typical split is 80% for $\mathcal{Y}_{\text{Train}}$ and 20% for $\mathcal{Y}_{\text{Validate}}$.
b. Estimate model parameters using only the data contained in $\mathcal{Y}_{\text{Train}}$. With these model parameters, compute the one-step residuals for $\mathcal{Y}_{\text{Validate}}$.

c. Using these residuals, evaluate either the likelihood function (9) or the sum of squared residuals (RMS).
d. The model with the highest likelihood or lowest RMS is preferred.

4 Smart Building-Related Models

This section presents multiple models for a building where the model of the building itself is the same, but the heating system models and control strategies are different. The first model uses conventional electrical heaters (radiators) to supply heat to the room air. The second model uses a heat pump: a compressor heats water, which then flows into pipes based under the floor. This ground-sourced heating is very efficient electricity-wise due to the compressor and is thus an attractive solution for heating.

The models are formulated using stochastic differential equations (SDEs), and the stochastic model for the building is closely related to the simple model introduced in Sect. 2.1.

4.1 The Heat Pump Model

Halvgaard et al. [20] describe a model for a building with a heat pump that is reused in this chapter. The model includes the same two important states as the model in Sect. 2.1: the room air temperature and the floor medium temperature. Additionally, it includes the temperature of the water connected to the heat pump. That makes the system states $\boldsymbol{x}(t) = \left[T_r(t), T_f(t), T_w(t)\right]^T$. Table 1 lists and describes all variables in the model, and Fig. 2 shows an illustration of the smart building model and the directions of the heat dynamics.

Regarding the disturbances of the model, two elements are of high importance: the solar radiation and ambient air temperature. The solar radiation generally plays a double role regarding the heating of buildings: it directly enters a building through, e.g., windows, but it also highly influences the ambient air temperature, which in

Table 1 Description of the variables in the heat pump model

Variable	Unit	Description
T_r	°C	The room air temperature
T_f	°C	The temperature of the floor medium
T_w	°C	The temperature of the water in the compressor and pipes
T_a	°C	The ambient air temperature
W_c	W	The energy delivered to the compressor of the heat pump
ϕ_s	W/m^2	The solar radiation entering the building

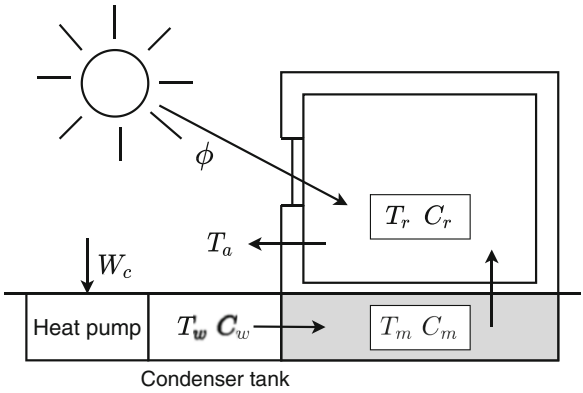

Fig. 2 An illustration of the smart house model. Each state has an associated heat capacity and temperature. The arrows indicate the positive direction of energy flow

turn affects the building. Later sections will describe and model these dynamics and cross-correlations.

Based on Fig. 2, the equations below describe the overall building dynamics

$$\begin{aligned} C_r dT_r(t) &= \left(Q_{fr}(t) - Q_{ra}(t) + A_w p \phi_s(t)\right) dt + \sigma_r d\omega_r(t) \,, \\ C_f dT_f(t) &= \left(Q_{wf}(t) - Q_{fr}(t) + A_w (1-p) \phi_s(t)\right) dt + \sigma_f d\omega_f(t) \,, \\ C_w dT_w(t) &= \left(\eta W_c(t) - Q_{wf}(t)\right) dt + \sigma_w d\omega_w(t) \,, \end{aligned} \tag{14}$$

where C_r, C_f, and C_w are heat capacities for the room air, floor, and water, respectively. ω_r, ω_r, and ω_w are Wiener processes for each state, and σ_r, σ_f, and σ_w are noise constants. The heat flows are given by

$$\begin{aligned} Q_{ra}(t) &= r_{ra}^{-1} \left(T_r(t) - T_a(t)\right) \,, \\ Q_{fr}(t) &= r_{fr}^{-1} \left(T_f(t) - T_r(t)\right) \,, \\ Q_{wf}(t) &= r_{wf}^{-1} \left(T_w(t) - T_f(t)\right) \,. \end{aligned} \tag{15}$$

We can write the set of SDEs in linear form as in (3)

$$\begin{aligned} \mathsf{A} &= \begin{bmatrix} -\frac{1}{r_{fr}C_r} - \frac{1}{r_{ra}C_r} & \frac{1}{r_{fr}C_r} & 0 \\ \frac{1}{r_{fr}C_f} & -\frac{1}{r_{wf}C_f} - \frac{1}{r_{fr}C_f} & \frac{1}{r_{wf}C_f} \\ 0 & \frac{1}{r_{wf}C_w} & -\frac{1}{r_{wf}C_w} \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 0 \\ 0 \\ \frac{\eta}{C_w} \end{bmatrix}, \\ \mathsf{E} &= \begin{bmatrix} \frac{1}{r_{ra}C_r} & A_w \frac{(1-p)}{C_r} \\ 0 & A_w \frac{p}{C_f} \\ 0 & 0 \end{bmatrix}, \quad \mathsf{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \end{aligned} \tag{16}$$

with the variables $\boldsymbol{x}(t) = \left[T_r(t), T_f(t), T_w(t)\right]^T$, $u(t) = W_c(t)$, $\boldsymbol{d}(t) = [T_a(t), \phi_s(t)]^T$. Table 2 lists and briefly describes all the parameters in the model.

4.2 The Electrical Heater Model

The model using electrical heaters is almost identical to the one introduced in Sect. 2.1, with the exception of p, which is significant. Otherwise the parameters for this model are the same as used for the heat pump model. By separating the input and disturbance variables in the previous introduced model, we obtain the following linear second-order system:

$$\begin{aligned} \mathsf{A} &= \begin{bmatrix} -\frac{1}{r_{fr}C_r} - \frac{1}{r_{ra}C_r} & \frac{1}{r_{fr}C_r} \\ \frac{1}{r_{fr}C_f} & -\frac{1}{r_{fr}C_f} \end{bmatrix}, \ \mathsf{B} = \begin{bmatrix} \frac{1}{C_w} \\ 0 \end{bmatrix}, \\ \mathsf{E} &= \begin{bmatrix} \frac{1}{r_{ra}C_r} & A_w \frac{(1-p)}{C_r} \\ 0 & A_w \frac{p}{C_f} \end{bmatrix}, \ \mathsf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \end{aligned} \tag{17}$$

with the variable $\boldsymbol{x}(t) = \left[T_r(t), T_f(t)\right]^T$, $u(t) = W_c(t)$, $\boldsymbol{d}(t) = [T_a(t), \phi_s(t)]^T$.

4.3 Buildings with Stationary Batteries and Electrical Vehicles

In the very near future, electrical vehicles (EVs) will be in almost every household, and it is believed by many that future smart buildings will include stationary

Table 2 The values used in the model for a single smart home in (14) and (17)

Parameter	Value	Unit	Description
C_r	810	kJ/°C	Heat capacity constant for the room air
C_f	3315	kJ/°C	Heat capacity constant for the floor
C_w	836	kJ/°C	Heat capacity constant of the water in the pipes
r_{ra}	0.036	kJ/(°C h)	Resistance against heat transfer between the room air and the ambient air
r_{fr}	0.0016	kJ/(°C h)	Resistance against heat transfer between the floor and the room air
r_{wf}	0.036	kJ/(°C h)	Resistance against heat transfer between the water and the floor
p	0.1		The fraction of energy from the solar energy into the room air
η	3		The heat pump coefficient of performance
A_w	2.9		The effective window area

batteries [21, 57, 63]. The latter has the purpose of storing electricity harvested from photo-voltaic cells (PVs) and to buy and sell electricity from the market when the price is low and high, respectively. Adding a stationary battery and potentially an EV also greatly increases the *flexibility* of a building. We follow the modelling approach as in [63]. For more extended state-space models for batteries, see [7, 62]. We shall not, however, use or demonstrate these models in this chapter.

The fundamental differential equation governing the state of charge (SOC) of an (very simplified) integrating battery has the form

$$\dot{\gamma}(t) = \frac{V(t)}{Q} i(t) \,, \tag{18}$$

where $\gamma \in [0, 1]$ is the SOC (0 is discharged, while 1 is fully charged), V is the voltage, Q is the total battery capacity, and i is the current. We can rewrite this as the *power* flowing in and out of the battery

$$\dot{\gamma}(t) = \frac{1}{Q} \left(\eta^+ P^+(t) - \eta^- P^-(t) \right) \,, \tag{19}$$

where P^+ and P^- are the power flow in and out of the battery and η^+ and η^- are the respective efficiency constants. The corresponding SDE formulation is

$$\mathrm{d}\gamma(t) = \left(\eta^+ P^+(t) - \eta^- P^-(t) \right) \mathrm{d}t + \sigma_\gamma \mathrm{d}\omega(t) \,. \tag{20}$$

If the smart building is equipped with both a stationary battery and an EV, then we need a description of both

$$\mathrm{d}\gamma_{ev}(t) = \left(\eta^+_{ev} P^+_{ev}(t) - \eta^-_{ev} P^-_{ev}(t) \right) \mathrm{d}t + \sigma_{ev} \mathrm{d}\omega_{ev}(t) \,, \tag{21a}$$

$$\mathrm{d}\gamma_{bat}(t) = \left(\eta^+_{bat} P^+_{bat}(t) - \eta^-_{bat} P^-_{bat}(t) \right) \mathrm{d}t + \sigma_{bat} \mathrm{d}\omega_{bat}(t) \,, \tag{21b}$$

where γ_{ev} and γ_{bat} are the EV and stationary battery SOC, respectively. P^+_{bat} and P^-_{bat} are the bought and sold electricity from the market, and P^+_{ev} and P^-_{ev} are charging and discharging the EV battery, respectively.

The two batteries thus add two additional states to the smart building state-space model. To write the model in state-space form requires some assumptions. First, we assign the EV usage as a disturbance variable, $d_{ev}(t) = P^-_{d,ev}(t)$. Second, we assume that all the electricity generated by PVs and bought from the market go directly to the stationary battery. The solar radiation is thus also a disturbance, $d_{bat}(t) = \phi_s(t)$. The input variable for the battery is $u_{bat}(t) = \left[P^+_{bat}(t), P^-_{bat}(t), W_c(t), P^+_{ev}(t) \right]^T$. Writing the state-space formulation for the *entire* smart building including batteries, the state-space variables become $\boldsymbol{x}(t) = \left[T_r(t), T_f(t), T_w(t), \gamma_{ev}(t), \gamma_{bat}(t) \right]^T$, $\boldsymbol{u}(t) = \left[P^+_{bat}(t), P^-_{bat}(t), W_c(t), P^+_{ev}(t) \right]^T$, and $\boldsymbol{d}(t) = \left[T_a(t), \phi_s(t), P^-_{ev}(t), \phi_s(t) \right]^T$. Similarly, the continuous-time linear system likewise is

$$\mathsf{A} = \begin{bmatrix} -\frac{1}{r_{fr}C_r} - \frac{1}{r_{ra}C_r} & \frac{1}{r_{fr}C_r} & 0 & 0\,0 \\ \frac{1}{r_{fr}C_f} & -\frac{1}{r_{wf}C_f} - \frac{1}{r_{fr}C_f} & \frac{1}{r_{wf}C_f} & 0\,0 \\ 0 & \frac{1}{r_{wf}C_w} & -\frac{1}{r_{wf}C_w} & 0\,0 \\ 0 & 0 & 0 & 0\,0 \\ 0 & 0 & 0 & 0\,0 \end{bmatrix}, \tag{22a}$$

$$\mathsf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\eta}{C_w} & 0 & 0 & 0 \\ 0 & \frac{\eta_{ev}}{Q_{ev}} & 0 & 0 \\ -\frac{\eta_{bat}}{Q_{bat}} & -\frac{\eta_{bat}}{Q_{bat}} & \frac{\eta_{bat}}{Q_{bat}} & -\frac{\eta_{bat}}{Q_{bat}} \end{bmatrix}, \quad \mathsf{C} = \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix}, \tag{22b}$$

$$\mathsf{E} = \begin{bmatrix} \frac{1}{r_{ra}C_r} & \frac{(1-p)}{C_r} & 0 & 0 \\ 0 & \frac{p}{C_f} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\eta_{ev}}{Q_{ev}} & 0 \\ 0 & 0 & 0 & \frac{\eta_{bat}}{Q_{bat}}\eta_{pv}n_{pv} \end{bmatrix}. \tag{22c}$$

5 Disturbance Modelling

In this section, we shall use the well-documented meteorological data presented in [3] as the foundation for the grey-box models for the disturbances. Table 3 lists and describes each attribute of the data, which is collected from two weather stations in Værløse and Taastrup in Denmark. Samples are taken hourly from January 1, 1967 to December 31, 1973. The cloud cover is measured on the so-called okta scale. An okta is an integer in the range from 0 to 9, where 0 is completely clear skies, gradually gets more cloudy up till 8 that is fully overcast. Okta 9 is the class of non-observable cloud cover conditions, e.g., in foggy weather or heavy snow fall. We thus denote the okta state space by

$$C = \{0, 1, \ldots, 8, 9\}\,. \tag{23}$$

Table 3 Facts about the data and they are measured

Attribute	Notation	Unit	Measurement method
Cloud cover	$\{c', c, \kappa, Z_\kappa\}$	okta	Measured once every hour
Diffuse radiation	I_D	W/m^2	The average of 6 observations within an hour
Direct radiation	I_N	W/m^2	The average of 6 observations within an hour
Net radiation	R_n	W/m^2	The average of 6 observations within an hour
Ambient air temperature	T_a	°C	The average of 6 observations within an hour

5.1 Cloud Cover

The type, height, and amount of cloud cover have enormous influence on energy levels and balances of the lower atmosphere, i.e., the local weather close to Earth's surface. The variations of the solar radiation are mainly due to the absorption of energy by the molecules of the clouds. For example, in case of a heavy cloud cover, much less solar radiation gets through the atmosphere down to the surface. In a control context of a smart building that has PVs and is able to harvest energy from the sun, it is crucial to know the amount of solar radiation available. The cloud cover is undoubtedly the single most important factor in this case. The cloud cover also has a big impact on the air temperature in the lower atmospheric layers. When the rays from the sun hit the Earth's surface, a certain fraction gets absorbed and heats up the soil that, in turn, heats up the air. A good model for the cloud cover is thus a crucial element of a disturbance model for describing the local weather.

5.1.1 Discrete State-Space Cloud Cover Model

Figure 3 shows a plot of the cloud cover data from March. The overall dynamics seems fluctuating and is to some extend random. However, the cloud cover seems to spend more time and be more stable at both the ends of the scale (around okta 0 and 8).

Due to the discrete measure of the cloud cover, it is tempting to opt for a discrete state-space model to describe the cloud cover. An example of this is a continuous-time Markov model, see, e.g., [53]. The literature does describe successful models of this kind using both homogeneous and in-homogeneous models to describe the diurnal behavioural variation of the cloud cover in Denmark [41, 43]. The results confirm that the cloud cover is more stable in the clear sky and overcast states, with greater fluctuation in the middle states.

This model very well describes the probabilistic dynamics of the cloud cover. It also supplies estimates of the future expected value through the *Kolmogorov forward equations*. But the rest of the disturbance models will be formulated as continuous state-space models and rely on SDEs. Therefore, in a combined disturbance model for the smart building and in an MPC framework, a SDE describing the cloud cover

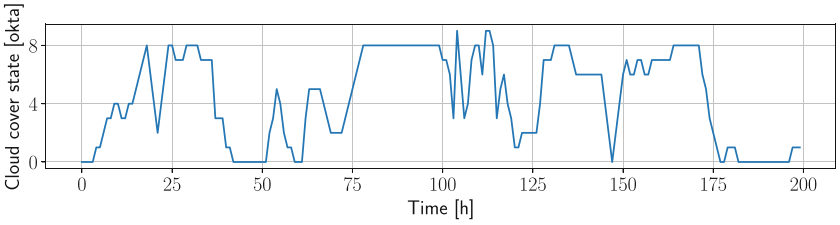

Fig. 3 A sample visualisation of the cloud cover data in March

becomes more convenient. For this reason, we shall now formulate, estimate, and validate a SDE-based model for the cloud cover.

5.1.2 Continuous State-Space Model Based on Stochastic Differential Equations

Recall the typical form of SDEs, explained in Sect. 4

$$\boldsymbol{x}(t) = f(\boldsymbol{x}(t), t)\mathrm{d}t + g(\boldsymbol{x}(t), t)\mathrm{d}\boldsymbol{\omega}(t) \ . \tag{24}$$

The analysis from the results of the Markov models in, e.g., [43] show that the cloud cover is governed by very special dynamics. It turns out that the process is less likely to move when it is in the end points (okta 0 and 8) and more likely for middle oktas. Formulating a SDE with these dynamics is not a trivial process. A first observation is that the cloud cover state space has boundaries. It is thus important to ensure that the SDE does not allow the process to go outside the boundaries. Let $\mathcal{K} = [0, 1]$ be the set of real numbers from 0 to 1, and let $\kappa \in \mathcal{K}$ denote the cloud cover state on *some* normalised scale. Starting with the diffusion, g, a very appropriate function can be

$$g(\kappa(t), t) = \sigma_\kappa \kappa(t) \left(1 - \kappa(t)\right) \ , \tag{25}$$

where σ_κ is a constant (Fig. 4). This choice ensures that the diffusion goes to zero in both ends of the okta scale and is also largest in the middle—which is desirable in order to make the process stay at either end longer time while ensuring that the middle states are more transient. In the grey-box modelling framework, we should assume some structure on the drift function, but it can be useful to also use some flexible functions such as the Legendre polynomials to allow the data to freely form them to easily maximise the likelihood function. The work in this chapter uses the following *mean reverting process*:

$$\mathrm{d}\kappa(t) = \theta\left(\kappa(t)\right)\left(\mu(\kappa(t)) - \kappa(t)\right)\mathrm{d}t + \sigma_\kappa \kappa(t)(1 - \kappa(t))\mathrm{d}\omega(t) \ , \tag{26}$$

where μ is the mean value and θ is the reversion speed. This SDE has a state-dependent diffusion term that has some nice modelling features, but it also has some pretty significant disadvantages from an estimation and simulation standpoint. The next part will detail these disadvantages.

5.1.3 Transformation into a State-Independent Diffusion Process

State-dependent diffusion terms in SDEs give rise to problems in estimation and simulation [46]. Two of the more influential problems are:

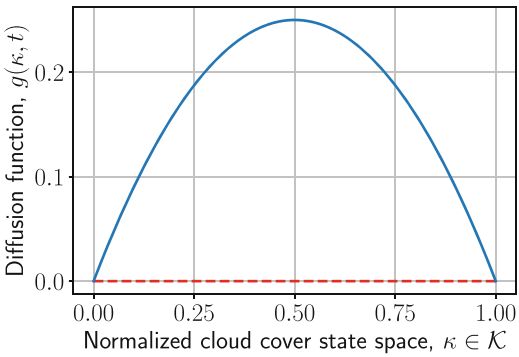

Fig. 4 The diffusion function in (25) with $\sigma_\kappa = 1$

- *Simulation* from a SDE with state-dependent diffusion can have slower convergence rate and then require more computational power.
- *Predictions* using methods like the extended Kalman filter (EKF) can be wrong and even illegal if they go outside of the bounds of the domain due to the linearisation.

A popular solution to this problem in the literature is to use the so-called *Lamperti transformation* [48].

The Lamperti transformation heavily relies on the result from stochastic calculus called Ito's lemma. Informally speaking, Ito's lemma corresponds to the chain rule for stochastic calculus: given a process, X, and a function, $\psi(X)$, Ito's lemma states the derivative of the function as a stochastic process $Z = \psi(X)$. That way, we can alternatively view the lemma as the equivalence of two processes, X and Z, by a closed formula using a transformation ψ. Consider a strictly positive process: by taking the natural logarithm of the process, we create a new process that lives on the entire real line. To obtain the differential equation governing this process, usually we would use the chain rule. But for a stochastic process, we need Ito's lemma. The special case where we choose a function that results in a constant diffusion term, we call the Lamperti transformation. Let us start by stating Ito's very famous lemma [22] (using some simplifying notation).

Lemma 1 (Ito's Lemma) *Let X be an Ito's process in the form*

$$\mathrm{d}X = f(X,t)\mathrm{d}t + g(X,t)\mathrm{d}\omega.$$

Let the function $\psi(X,t) \in C^2(\mathbb{R} \times [0,\infty))$, and then the process

$$Z = \psi(X,t)$$

is an Ito's process. Furthermore, Z is governed by the process

$$\mathrm{d}Z = \left(\frac{\partial\psi}{\partial t}(X,t) + f(X,t)\frac{\partial\psi}{\partial X}(X,t) + \frac{1}{2}\frac{\partial^2\psi}{\partial X^2}(X,t)g(X,t)^2\right)\mathrm{d}t + \frac{\partial\psi}{\partial X}(X,t)g(X,t)\mathrm{d}\omega.$$

By choosing $\partial\psi/\partial X(X,t)$ to be equal to $1/g(X,t)$, the diffusion term becomes exactly a unit for the Ito-transformed process, Z. The following theorem states this result [22, 46].

Theorem 1 (Lamperti Transformation) *Let X be an Ito's process as in Lemma 1. Define the function*

$$\psi(X,t) = \int \frac{1}{g(x,t)}\mathrm{d}x\bigg|_{x=X}$$

. If $\psi(X,t)$ is bijective onto $\mathbb{R}$, then Z has a unit diffusion term and has the following process:

$$\mathrm{d}Z = \left(\frac{\partial\psi}{\partial t}(\psi^{-1}(Z,t),t) + \frac{f(\psi^{-1}(Z,t),t)}{g(\psi^{-1}(Z,t),t)} - \frac{1}{2}\frac{\partial g}{\partial X}(\psi^{-1}(Z,t),t)\right)\mathrm{d}t + \mathrm{d}\omega.$$

Applying the Lamperti transformation on the specific SDE in (26) (except for leaving a constant on the diffusion term), the Lamperti-transformed process, $Z_\kappa = \psi(\kappa,t)$, is

$$\begin{aligned} Z_\kappa = \psi(\kappa,t) &= \int \frac{1}{x(1-x)}\mathrm{d}x\bigg|_{x=\kappa} \\ &= \log(\kappa) - \log(1-\kappa) = \log\left(\frac{\kappa}{1-\kappa}\right). \end{aligned} \tag{27}$$

For $\kappa \in \mathcal{K}$, the process Z_κ is in all of the real numbers, $\mathbb{R}$. The inverse of ψ is

$$\kappa = \psi^{-1}(Z_\kappa,t) = \frac{\exp(Z_\kappa)}{1+\exp(Z_\kappa)}. \tag{28}$$

Using Theorem 1, the *state-independent* Lamperti-transformed process becomes

$$\begin{aligned} \mathrm{d}Z_\kappa &= \left(0 + \frac{f(\kappa,t)}{\kappa(1-\kappa)} + \kappa - \frac{1}{2}\right)\mathrm{d}t + \sigma_\psi \mathrm{d}\omega \\ &= \left(\frac{f\left(\frac{\exp(Z_\kappa)}{\exp(Z_\kappa)+1},t\right)}{\frac{\exp(Z_\kappa)}{\exp(Z_\kappa)+1}\left(1-\frac{\exp(Z_\kappa)}{\exp(Z_\kappa)+1}\right)} + \frac{\exp(Z_\kappa)}{\exp(Z_\kappa)+1} - \frac{1}{2}\right)\mathrm{d}t + \sigma_\psi \mathrm{d}\omega \\ \mathrm{d}Z_\kappa &= f_\psi(Z_\kappa)\mathrm{d}t + \sigma_\psi \mathrm{d}\omega\,. \end{aligned} \tag{29}$$

To get an intuition of how the Lamperti transformation works, Fig. 5 shows an example of a SDE with simple linear drift while having the diffuse function as in (25), together with the Lamperti-transformed drift. For the Lamperti-transformed process, the drift will make sure that it always stays around 0 and does not go towards $\pm\infty$. But the most interesting feature of the Lamperti drift is perhaps that

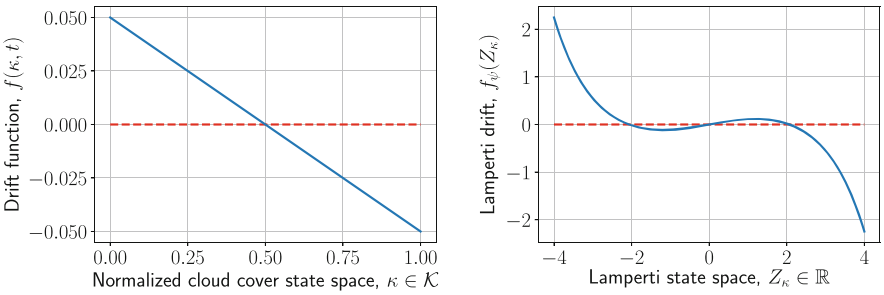

Fig. 5 A simple linear drift function (left) and its Lamperti-transformed equivalent (right) from the process $\mathrm{d}X = 0.1 \cdot (0.5 - X)\mathrm{d}t + X(1 - X)\mathrm{d}\omega$

there is a stable stationary point at $Z_k = 0$ driving the process towards zero. The explanation should be found in the shape of the diffusion function that is largest in the middle and goes to zero in the ends. Around $Z_k = 0$, the diffusion is the dominating force and has zero mean—i.e., is expected to keep the process around zero due to the zero mean (at least expectation-wise). The variance, however, will make sure to drive the process away from zero. But when the process gets too far out, the drift will dominate again and force it towards zero.

Due to the advantages of dealing with a state-independent SDE, all estimation, simulation, and prediction happen in the Lamperti domain in (29) and are subsequently transformed back to the original cloud cover domain by ψ^{-1}. We now turn to estimate parameters in (26).

5.1.4 Estimation of Parameters Embedded in the SDE

As previously mentioned, we need to base the choice on the drift function of the SDE on physical properties of the process we attempt to model. We choose to use the following SDE:

$$\mathrm{d}\kappa = \theta\sqrt{\kappa(1-\kappa)}\left(\frac{\exp(P_7(\kappa))}{1+\exp(P_7(\kappa))} - \kappa\right)\mathrm{d}t + \sigma\kappa(1-\kappa)\mathrm{d}\omega\,, \tag{30}$$

where $P_7(\kappa)$ is a linear combination of the first seven Legendre polynomials. The model in (30) has a very complex mean value. The intuition is that it allows the mean value to move rather freely in the range from 0 to 1, depending on the cloud cover state. The reverting-speed term $\sqrt{\kappa(1-\kappa)}$ may seem like an over-complication in the model. But previous cloud cover modelling attempts suggest that the process spends more time in the ends of the okta scale. The term $\sqrt{\kappa(1-\kappa)}$ makes the drift smaller at the ends of the scale and therefore intuitively makes the process stay there for longer before reverting back to the middle.

Before we are able to estimate parameters in the SDE in (26), we require a transformation of the data, $\zeta : C \to \mathcal{K}$, due to the diffusion term (since it requires the process to be in the interval $\mathcal{K} = [0, 1]$). First, we let okta 9 be missing observations in the data, such that the observable okta state space (and the state space we should model) is $\{0, 1, \ldots, 8\}$. Choosing a good transformation is not straightforward. We cannot simply divide the okta state space by 8. Doing so implies that okta 0 in C corresponds to 0 in $\mathcal{K}$ and likewise okta 8 corresponds to 1 in $\mathcal{K}$. But the drift and diffusion of the SDE equal zero for $\kappa = 0$ and $\kappa = 1$ and the SDE gets stuck. Furthermore, the Lamperti transformation is not well defined for these values.

In the discrete okta state space, C, the distance between each pair of neighbouring states is the same. However, in the continuous state space $\mathcal{K}$, this is definitely not given. In fact, the definition of the oktas [3] indicates that the end points of okta pairs $\{0, 1\}, \{7, 8\} \in C$ are more alike compared to the rest of oktas. Thus by moving the end points of the okta scale closer together as in Fig. 6 before dividing by 8, we might obtain a good transformation that behaves well. We thus choose the following transformation for the cloud cover to get it into $\mathcal{K}$:

$$\kappa = \zeta(c) = c/8\,, \quad c \in C\,,\ \kappa \in \mathcal{K}\,, \tag{31}$$

where the oktas 0 and 8 have been perturbed according to Fig. 6. But how much should we move the end points as to get the best model? To answer this question, we use *Akaike's information criterion* (AIC) as a measure to compare estimated models. We perform a small grid search for the positions of oktas 0 and 8 around the points 0.5 and 7.5 and choose the model that performs best in terms of the AIC value.

To estimate the parameters in the SDE, we apply ML estimation using the continuous–discrete extended Kalman filter (CDEKF), see, e.g., [10, 23]. We use the CDEKF to predict from the SDE. Let $\boldsymbol{\theta}$ denote the set of parameters in the model. Given $\boldsymbol{\theta}$, we use the CDEKF to calculate the 1-step prediction and variance of the state, $\hat{\kappa}_{k|k-1}(\boldsymbol{\theta})$ and $R_{k|k-1}(\boldsymbol{\theta})$. Let $\epsilon_k(\boldsymbol{\theta}) = \hat{\kappa}_{k|k-1}(\boldsymbol{\theta}) - \kappa_k$ be the prediction error of the state using $\boldsymbol{\theta}$, and then the ML estimate is given by (9).

Table 4 shows the result of the grid search and suggests that the perturbation $(\text{okta0}, \text{okta8}) = (0.6, 7.4)$ is by far the best choice in terms of AIC—not surprisingly, since we expected the end points to behave more like their neighbours.

Figure 7 shows histograms of the long-term distributions of the data and model 3 using the state transformation in Fig. 6 with the values $(\text{okta0}, \text{okta8}) = (0.6, 7.4)$.

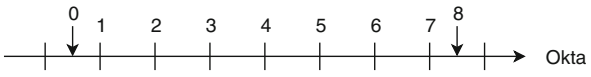

Fig. 6 An improved cloud cover transformation, ζ. Due to oktas 0 and 1, and okta 7 and 8, supposedly being more alike compared to the other oktas, we propose the following transformation

Table 4 The AIC value for each combination of end points in the grid search using the model in (30)

Okta 0 \ Okta 8	7.40	7.45	7.50	7.55
0.50	21079	21432	22436	22557
0.55	21330	21591	21680	26515
0.60	20442	22817	23294	24994

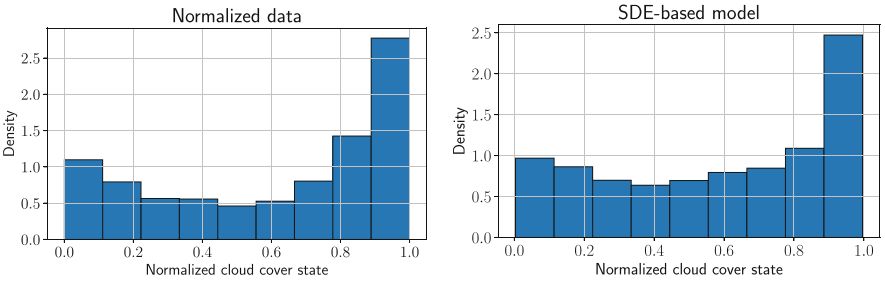

Fig. 7 The long-term distribution of the data (left) and model 3 (right) using the improved state transformation in (31) with oktas 0 and 8 moved to 0.6 and 7.4, respectively

Fig. 8 The autocorrelation function of the 1-step prediction residuals of model using the okta transformation as in (31)

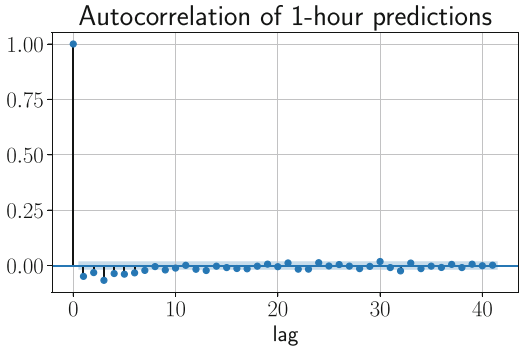

Table 5 Parameter estimates for the model in (30) with the locations of oktas 0 and 8 moved to 0.60 and 7.40. $\hat{p}_i$, $i = 1, \ldots 7$, is the parameter for the ith Legendre polynomial

Parameter	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{\theta}$	$\hat{\sigma}$
Estimate	–53.1	14.6	–42.3	8.8	–58.1	–30.3	–45.7	0.187	0.835
Std. Err.	2.13	0.831	1.715	0.434	2.273	1.329	1.873	0.003	0.012

Even though the distributions are not identical, the model mimics the overall pattern very well. The autocorrelation function in Fig. 8 is also close to zero as desired.

Table 5 shows the parameter estimates for the model. It can be hard to interpret the model by simply looking at the parameters since the Legendre polynomial parameters do not make much sense by themselves. Instead, we show the drift and Lamperti drift functions of the cloud cover state in Fig. 9 to give an intuition of the model. It truly has a very non-trivial and complex shape. It is partly the mean value function, $\exp(P_7(\kappa))/(1+\exp(P_7(\kappa)))$, and the reverting-speed function, $\theta\sqrt{\kappa(1-\kappa)}$, that makes this possible. The former allows the drift function to make

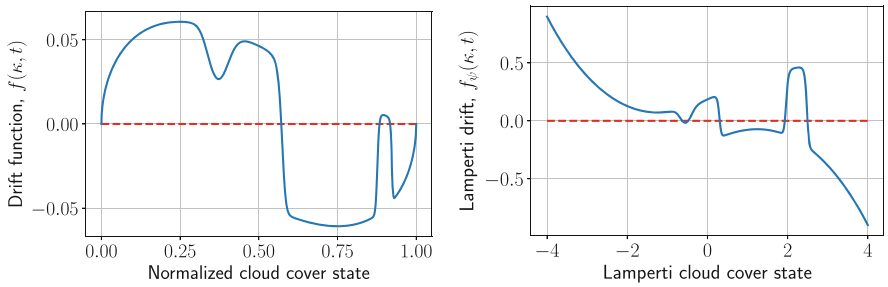

Fig. 9 Left: The drift function of model 3 in (30) using the state transformation in (31). Right: The same drift function, but in the *Lamperti domain*. The dashed line simply indicates the zero line

sudden changes from positive to negative and vice versa (which happens around $\kappa = 0.6$ and $\kappa = 0.9$). The ladder makes sure that the drift goes to zero in both ends and gives it the overall bending shape. The very sharp bend seen in the drift around $\kappa = 0.9$ surely seems odd and out of place. But it has a crucial role in the long-term distribution. It creates a stable stationary point for the process and therefore makes the process stay in the overcast states for more time. This is especially visible in the long-term distribution by the larger density in the overcast state compared to the clear skies states.

The Lamperti drift is harder to interpret, as it lives in the logistic domain. But it illustrates how the Lamperti process behaves and how the state dependence affects it.

5.2 *Solar Radiation*

Now that we have established a model describing the cloud cover dynamics based on SDEs, we move on to describe the next component of an advanced disturbance model: the solar radiation. It is responsible for some of the fast heating dynamics influencing the indoor air temperature and is thus an important disturbance.

5.2.1 Modification of the Cloud Cover Data

The data introduced in Sect. 5.1 are averages of multiple observations within an hour, except for the cloud cover. The cloud cover value taken at time t_k is thus not representative for the cloud cover in the interval $[t_k, t_{k+1}]$. To obtain better estimates for such cloud cover values, we use the average of the cloud cover at time t_k and t_{k+1}

$$c_{k+1} = (c'_{k+1} + c'_k)/2\,, \tag{32}$$

where c'_k is the raw cloud cover data. The corrected cloud cover values live in the state space $c_k \in \{0, 0.5, 1, \ldots, 8.5, 9\}$. We use these values for the cloud cover throughout this section.

5.2.2 Solar Radiation Components and Modelling Approach

The term *global solar radiation* covers all the short-wave radiation at the surface of the Earth. These are high-energy rays that transfer large amounts of energy that turns into heat when absorbed by objects or electricity by the PVs. Previous modelling attempts range from simple polynomial fits to complex black-box neural networks, see, e.g., [15, 59, 60, 65]. The global solar radiation consists of two components: *diffuse* and *direct* radiation. The direct radiation is all short-wave radiation travelling undisturbed to the Earth's surface. The diffuse radiation is all short-wave radiation that is reflected from molecules in the atmosphere. The fundamental relationship is

$$\phi_s(t) = I_N(t)\sin(\alpha(t)) + I_D(t)\,, \tag{33}$$

where ϕ_s, I_N, and I_D are the global, direct, and diffuse radiation, respectively, and α is the solar elevation angle. That is, obtaining models for each component gives a model for the global radiation. Lambert–Beer's law gives an analytical expression for the intensity of the radiation when it arrives at the Earth's surface

$$I_\lambda = I_{0\lambda} e^{-\int \mu_\lambda(s)\mathrm{d}s}\,, \tag{34}$$

where $I_{0\lambda}$ is the initial intensity and μ_λ is the attenuation of the medium the ray travels in. The integral in (34) is hard to evaluate since $\kappa_{a\lambda}$ is difficult to estimate for the atmosphere due to its very non-uniform density and dependence on the solar elevation.

We shall employ a more data-driven approach, namely kernel regression [64]. It estimates the conditional expectation, $E(Y|X = x)$, of a variable. In our case, we estimate the expected solar radiation given the cloud cover okta, $c \in C$, and the solar elevation angle, $\alpha(t)$. Local linear regression has an advantage over *constant* regression in that it generally induces less bias in the ends of the support. Figure 10 shows the direct radiation for some selected oktas (the diffuse radiation is omitted). It shows that for larger oktas, the data seem more scattered—and behave very poorly for okta 8.

Figure 11 shows the result of applying local linear kernel regression on the direct and diffuse radiation data for each okta. The conditional expectations are in line with the physical properties of the radiation types: The direct radiation is highest when there is close to zero clouds, while the diffuse radiation tops for a certain presence of clouds.

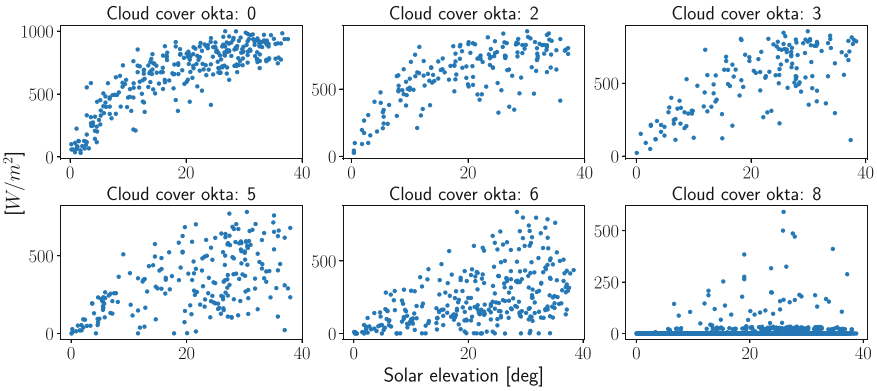

Fig. 10 The direct radiation for some example oktas: 0, 2, 3, 5, 6, and 8

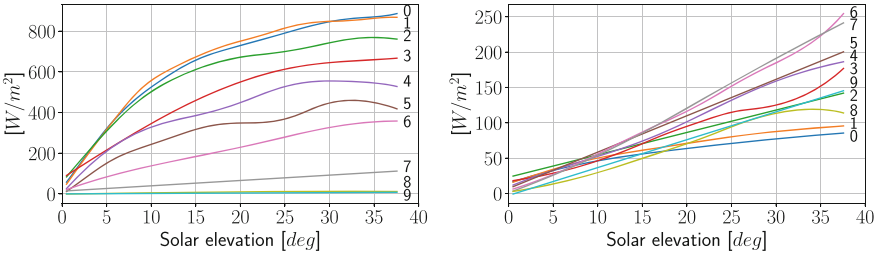

Fig. 11 The results from the kernel regression applied on the direct and diffuse radiation. For simplicity, only half of the okta values are shown

5.2.3 Describing the Deviation and Autocorrelation

Now that we have described the conditional expectation, we move on to describe the deviation and the potential autocorrelation left in the residuals. Figure 12 shows the residuals of the kernel regression applied to the direct radiation (we omit the diffuse radiation due to space limitation, but the behaviour is the same). It also suggests a rough linear increase in the standard deviation is the case (the same is true for the diffuse radiation). That is, we employ the following model for the standard deviation for each okta:

$$\sigma^{(c)}(t) = \beta_0^{(c)} + \beta_1^{(c)}\alpha(t) \,, \tag{35}$$

for both the direct and diffuse radiations. $\beta_0^{(c)}$ and $\beta_1^{(c)}$ are constant parameters for each $c \in C$. Let $\epsilon_k^{(c)} = \hat{y}_k^{(c)} - y_k^{(c)}$ be the residual, and let $\sqrt{R_k} = \sigma_k^{(c)}$ be the standard deviation at time t_k, $k = 1, 2 \ldots N$, for okta c. The above sets of parameters, $\{\beta_0^{(c)}, \beta_1^{(c)}\}$, can be estimated using ML estimation as in (9) using a numerical solver. The parameter estimates can be found in [61].

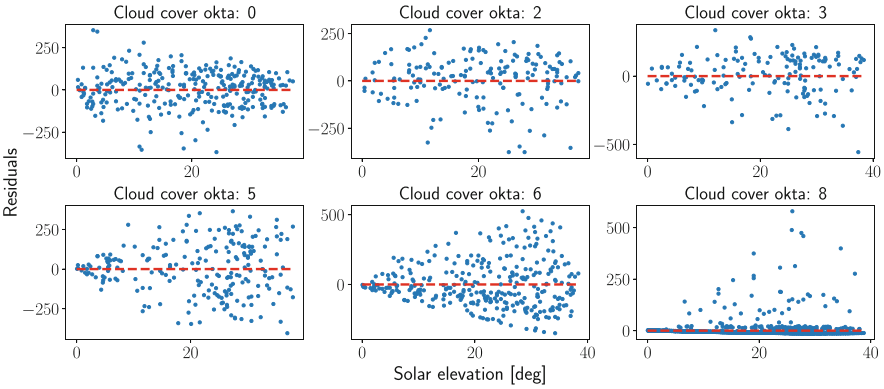

Fig. 12 The residuals for the direct radiation

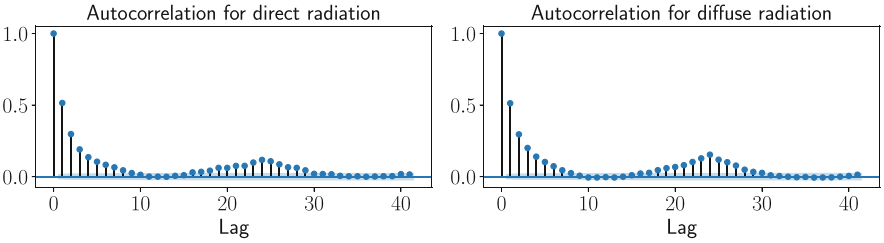

Fig. 13 The ACF for the direct and diffuse radiations

The cloud cover is not the single cause of variation in the solar radiation. For example, vapour, dust, ozone, and other particles give rise to *autocorrelation*. Let

$$e(t) = \frac{\hat{\epsilon}^{(c)}(t)}{\hat{\beta}_0^{(c)} + \hat{\beta}_1^{(c)}\alpha(t)} \tag{36}$$

be the standardised residuals of the direct and diffuse data (i.e., two processes). Note that each of the processes is standard normal distributed and independent of the cloud cover. We consider the two processes as a *multivariate* time series with the variable, $\boldsymbol{e}_k = \left[e_{N,k}, e_{D,k}\right]^T$, with missing observations (during night-time). See [38] for how to deal with missing observations in a time series. Figure 13 now shows the autocorrelation in $\boldsymbol{e}_k$. The fast exponential decay in the first few lags suggests that a first-order *autoregressive* (AR) model is necessary

$$\begin{aligned} \boldsymbol{e}_k &= \Phi \boldsymbol{e}_{k-1} + \boldsymbol{\epsilon}_k \,, \\ \boldsymbol{a}_k &= \boldsymbol{e}_k + \boldsymbol{\epsilon}_{a,k} \,, \end{aligned} \tag{37}$$

where Φ is the AR coefficients and $\boldsymbol{\epsilon}_k \sim N(\mathbf{0}, \mathsf{Q_e})$ and $\boldsymbol{\epsilon}_{a,k} \sim N(\mathbf{0}, \mathsf{R_a})$ are the process and observation noise, respectively. $\boldsymbol{e_k}$ is thus the noise process driven by its previous values and a noise term, $\boldsymbol{\epsilon}_k$. The observation equation (ref equation) is a white noise process that encumbers the observations with noise, $\boldsymbol{\epsilon}_{a,k}$, which we also need to estimate. We estimate the parameters by applying ML estimation and use the Kalman filter to estimate the covariance matrices for the noise terms. The results become

$$\Phi = \begin{bmatrix} 0.609 & 0.109 \\ (0.013) & (0.009) \\ 0 & 0.675 \\ & (0.010) \end{bmatrix}, \quad \mathsf{R_a} = \begin{bmatrix} 0.160 & 0 \\ (0.011) & \\ 0 & 0.162 \\ & (0.019) \end{bmatrix}, \quad \mathsf{Q_e} = \begin{bmatrix} 0.466 & 0.160 \\ (0.019) & (0.005) \\ 0.160 & 0.456 \\ (0.005) & (0.016) \end{bmatrix} \tag{38}$$

with the standard errors in parentheses beneath the estimate.

5.3 Net Radiation

The net radiation itself is not directly important for describing the heat dynamics of a building. But it is an important meteorological variable that heavily influences the ambient air temperature. The model for the ambient air temperature thus requires a model for the net radiation. The net radiation, also known as the net flux, is the balance of the total energy at the boundary of the atmosphere. It is simply the sum of the total outgoing and incoming energy of the atmosphere. A negative net radiation corresponds to more energy leaving the atmosphere and vice versa. In general, the net radiation, R_n, is given by the analytical formula [2]

$$R_n = (1 - \alpha_g)\phi_s + L_u + L_d \ , \tag{39}$$

where α_g is the albedo fraction and L_u and L_d are the upward and downward components of the long-wave radiation. The albedo is the fraction of global solar radiation that is reflected on the Earth's surface into space, and $(1 - \alpha_g)$ is therefore the fraction of global radiation that is partly absorbed and that becomes long-wave radiation.

Figure 14 shows the net radiation for a single March, which makes the dependence on the time of the day clear. A simple and convenient model to use for our purpose is the following, suggested by [37], which relies on the current cloud cover and the global radiation

$$R_n(c(t), \phi_s(t), t) = K_c + k_c\phi_s(t) + k\alpha(t)^2 + \epsilon(t) \ , \tag{40}$$

where K_c and k_c are constants that are dependent on the present cloud cover and α is the solar elevation. The model in (40) is linear in its parameters. This makes linear least squares estimation useful to fit the parameters

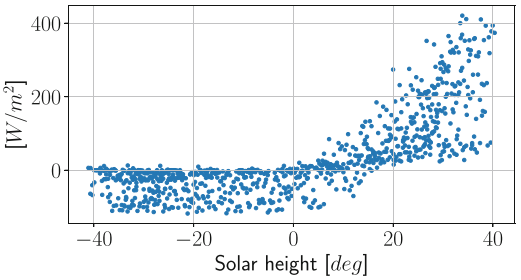

Fig. 14 The net radiation during March

$$\min_{\boldsymbol{x}} \|\mathsf{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \ , \tag{41}$$

having the unique solution

$$\hat{\boldsymbol{x}}_{LS} = (\mathsf{A}^T\mathsf{A})^{-1}(\mathsf{A}^T\boldsymbol{b}) \ . \tag{42}$$

A and $\boldsymbol{b}$ in (41) have the forms

$$\mathsf{A} = \begin{bmatrix} \boldsymbol{e}_{c_0}^T & \boldsymbol{e}_{c_0}^T\phi_s(0) & h(0)^2 \\ \boldsymbol{e}_{c_1}^T & \boldsymbol{e}_{c_1}^T\phi_s(1) & h(1)^2 \\ & \vdots & \\ \boldsymbol{e}_{c_i}^T & \boldsymbol{e}_{c_i}^T\phi_s(i) & h(i)^2 \\ & \vdots & \\ \boldsymbol{e}_{C_N}^T & \boldsymbol{e}_{c_N}^T\phi_s(N) & h(N)^2 \end{bmatrix} \ , \qquad \boldsymbol{b} = \begin{bmatrix} R_n(0) \\ R_n(1) \\ \vdots \\ R_n(i) \\ \vdots \\ R_n(N) \end{bmatrix} \ , \tag{43}$$

where N is the total number of observations and $\boldsymbol{e}_i$ is a vector of zeros with a one in the ith entry. c_i is the cloud cover index of the ith observation—recall the modified cloud cover data state space is $\{0, 0.5, \ldots, 8.5, 9\}$. $\boldsymbol{x}$ is thus the parameters, $\boldsymbol{x} = (K_0, K_{0.5}, K_1, \ldots, K_9, k_0, k_{0.5}, k_1, \ldots, k_9, k)$. Table 6 shows the estimated parameters in (40). The increasing trend in constant net radiation, $\hat{K}_c$, with the increasing amount of cloud cover indicates that the clouds 'contain' the net radiation (net energy flux) within the atmosphere. This can be recognised by the phenomenon that colder nights typically appear when the skies are completely clear.

5.4 Ambient Air Temperature

The only missing piece in the puzzle now in the advanced disturbance model is the ambient air temperature. While the net radiation describes the net flux at the boundary of the atmosphere, the following fundamental relationship describes the heat fluxes: *close to Earth's surface*

Table 6 The estimated parameters of the model in (40). Even though we have not imposed any regularisation or other things that tie the estimates together, the values are somewhat nicely distributed

Cloud cover okta	$\hat{K}_c$	$\hat{k}_c$	$\hat{\sigma}_\varepsilon$	$\hat{k}$
0	–69.2	0.549	40.1	0.0418
0.5	–68.2	0.551	38.5	0.0418
1	–74.1	0584	38.3	0.0418
1.5	–74.9	0.576	37.7	0.0418
2	–74.1	0.571	39.7	0.0418
2.5	–75.5	0.566	45.2	0.0418
3	–73.6	0.565	48.4	0.0418
3.5	–71.5	0.566	54.4	0.0418
4	–73.6	0.589	58.1	0.0418
4.5	–67.5	0.606	57.9	0.0418
5	–69.5	0.661	59.8	0.0418
5.5	–63.6	0.695	60.7	0.0418
6	–56.5	0.717	59.7	0.0418
6.5	–46.5	0.699	56.7	0.0418
7	–29.5	0.595	49.4	0.0418
7.5	–13.5	0.461	43.5	0.0418
8	2.5	0.150	31.8	0.0418

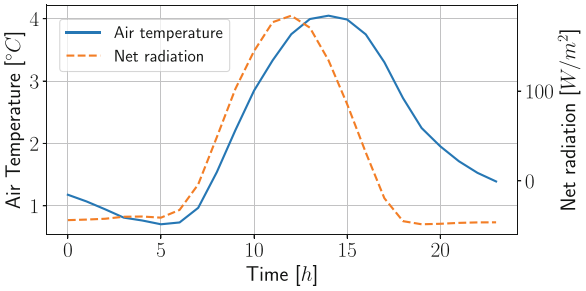

Fig. 15 The hourly mean values of the ambient air temperature and the net radiation during a day. The delayed response from the ambient air temperature suggests that a dynamic model is necessary

$$R_n = L_f + S_f + G_f \ , \tag{44}$$

where R_n is the net radiation described in Sect. 5.3 and G_f is the soil heat flux. L_f and S_f are the latent and sensible heat fluxes. The latent heat flux is heat gradients related to absorbed or released heat due to phase changes by matter—e.g., when water evaporates, it absorbs heat in order to decrease the molecule density. The sensible heat flux is all energy required to change the temperature of matter without phase changes taking place. The latent and sensible heat fluxes thus relate to the gradients of the air temperature. To get an idea of the kind of model needed to describe the air temperature, Fig. 15 shows the diurnal mean value variations of the net radiation and ambient air temperature. It supports the fact that the net radiation is an important explanatory variable to describe the air temperature. It further suggests that a dynamical model is needed due to the time lag between the peak values of 3–4 h.

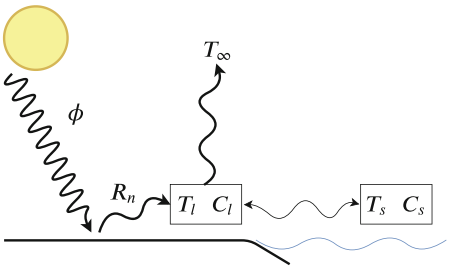

Fig. 16 An illustration of the dynamical model describing the ambient air temperature. The states are indicated by the squares containing the heat capacities

The atmospheric air directly above the Earth's surface has a relatively small heat capacity, making it quick to respond to level changes in the net radiation. Water in contrast has a very large heat capacity. The temperature of the seas thus highly regulates the temperature of the air above it. The air masses above sea and land interact due to climatic motions, and the sea consequently regulates the land air temperature. For instance, it is well known that the largest annual temperature difference occurs in the middle of large continents. Hence, the level of regulation by the sea depends on geographical location and local climate. Using the above knowledge about the behaviour and balances of the air temperatures, we are ready to formulate the stochastic dynamical model describing the ambient air temperature above land

$$C_w \mathrm{d}T_w(t) = \left(\frac{1}{R_{wl}} \left(T_l(t) - T_w(t)\right) \right) \mathrm{d}t + \sigma_w \mathrm{d}\omega_w(t) \,, \tag{45a}$$

$$C_l \mathrm{d}T_l(t) = \left(\frac{1}{R_{wl}} \left(T_w(t) - T_l(t)\right) + \frac{1}{R_{l\infty}} \left(T_\infty - T_l(t)\right) + R_n(t) \right) \mathrm{d}t + \sigma_l \mathrm{d}\omega_l(t) \,, \tag{45b}$$

$$T(t_k) = T_l(t_k) + v_k, \quad v_k \sim N(0, R_v) \,, \tag{45c}$$

where C_w and C_l are the heat capacities for sea and land, R_{wl} and $R_{l\infty}$ are the resistances against the heat flows between the states, ω_w and ω_l are the standard Wiener processes, and v_k is the observation noise. The model is also illustrated in Fig. 16: The solar radiation influences the net radiation, which in turn acts as an input to the land air temperature. The land air temperature interacts with the sea temperature and a constant outflow of energy, T_∞, to counteract heat inputs and to ensure stability of the model. Equation (45) thus uses the sea temperature as a *hidden state* to describe the land air temperature.

Since this is a continuous-time model, we use the continuous–discrete Kalman filter to calculate one-step predictions and estimate the observation variance to compute one-step predictions and filter the estimates. Let $\hat{x}_{k|k-1}(\boldsymbol{\theta})$ and $R_k(\boldsymbol{\theta})$ be the one-step prediction and observation variance for x_k at time t_k, calculated using

Table 7 The parameter estimates in (45) and the corresponding standard errors

Parameter	$\hat{C}_w$	$\hat{C}_l$	$\hat{R}_{wl}$	$\hat{R}_{l\infty}$	$\hat{\sigma}_w$	$\hat{\sigma}_l$
Estimate	534.56	58.99	0.0145	0.1017	0.913	0.0003
Std. Error	8.44	0.49	0.0001	0.0026	0.0123	0.0001

a given set of parameters $\boldsymbol{\theta} = (C_w, C_l, R_{wl}, R_{l\infty}, \sigma_w, \sigma_l)$. Then, the ML estimate is the solution to the problem in (9).

Using the data from Højbakkegård, Table 7 shows the estimation results for the model in (45). As expected, the heat capacity for the sea is much larger compared to the land air. Also note the very small process noise for the land air temperature, ω_l, compared to the sea temperature. This indicates that it is primarily the sea temperature that drives the land air temperature.

6 Model-Based Predictive Control

The previous sections focused on establishing statistically determined dynamical models for the smart building and the most important disturbances. This section shows the potential benefits of using the advanced disturbance models for forecasting. We start by introducing model predictive control (MPC) and deriving the optimisation problem involved with computing the optimal control. Furthermore, we discuss how to incorporate and use the given disturbance forecasts in the MPC algorithm. Lastly, we present a more classical method for handling disturbances in an MPC setup, where the disturbances are not modelled but instead an integrator is introduced to estimate the current disturbances. Even though the method provides offset-free control, we discuss why it is not ideal when dealing with very fast dynamics (as with the solar radiation).

6.1 Constrained Model Predictive Control

Many variations of MPC exist and have gained high popularity for control purposes due to the framework's superiority over non-predictive control schemes such as PI/PID control [35] and its simplicity. In general, the MPC framework is given by the following (*Bolza*) problem:

$$J(\hat{\boldsymbol{x}}_{k|k}, \{\hat{\boldsymbol{d}}_{k+i|k}\}_{i\in\mathcal{N}}) = \min_{\boldsymbol{u}} \int_{t_k}^{t_{k+N_p}} \ell(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \boldsymbol{d}(\tau))\mathrm{d}\tau + \ell_b(\boldsymbol{x}(t_{k+N_p})), \tag{46a}$$

$$s.t. \quad \boldsymbol{x}(t_k) = \hat{\boldsymbol{x}}_{k|k}\ , \tag{46b}$$

$$\boldsymbol{d}(t) = \hat{\boldsymbol{d}}_{k+i|k},\ t \in [t_{k+i}, t_{k+i+1}[\ , \tag{46c}$$

$$\mathrm{d}\boldsymbol{x}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t))\mathrm{d}t\ , \tag{46d}$$

$$\boldsymbol{x}(t) \in \mathcal{X}(t)\ ,\ \boldsymbol{u}(t) \in \mathcal{U}(t)\ , \tag{46e}$$

where $\mathcal{N} = \{0, 1, \ldots, N_p - 1\}$ is the control times and N_p is the prediction horizon. J is the cost function, $\boldsymbol{x}$ is the system, $\boldsymbol{u}$ is the input, and $\mathcal{X}(t)$ and $\mathcal{U}(t)$ are the allowed sets for $\boldsymbol{x}$ and $\boldsymbol{u}$. $\hat{\boldsymbol{x}}_{k|k}$ is the filtered estimate of $\boldsymbol{x}$ at time t_k and acts as the initial condition. $\{\hat{\boldsymbol{d}}_{k+i|k}\}_{i\in\mathcal{N}}$ is the sequence of disturbances, which in general comes from outside of the MPC framework. In this text, we get it from the separate disturbance model developed in the previous sections. ℓ_b is a cost on $\boldsymbol{x}$ on the boundary of the time domain sometimes called a *cost-to-go term*.

The cost function in (46a) involves evaluation of an integral. In practice though, a computer can only deal with discrete time. Consequently, the problem in (46) is typically reformulated as a discrete problem (in the case of a linear system)

$$J(\hat{\boldsymbol{x}}_{k|k}, \{\hat{\boldsymbol{d}}_{k+i|k}\}_{i\in\mathcal{N}}) = \min_{\hat{\boldsymbol{u}}_k} \sum_{i\in\mathcal{N}} \left[\ell_k(\hat{\boldsymbol{x}}_{k+i+1}, \hat{\boldsymbol{u}}_{k+i}, \hat{\boldsymbol{d}}_{k+i})\right] + \ell_{N_p}(\hat{\boldsymbol{x}}_{k+N_p})\ , \tag{47a}$$

$$s.t.\quad \hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_{k|k}\ , \tag{47b}$$

$$\hat{\boldsymbol{d}}_{k+i} = \hat{\boldsymbol{d}}_{k+i|k}\ , \tag{47c}$$

$$\hat{\boldsymbol{x}}_{k+i+1} = \mathsf{A}_\mathsf{d}\hat{\boldsymbol{x}}_{k+i} + \mathsf{B}_\mathsf{d}\hat{\boldsymbol{u}}_{k+i} + \mathsf{E}_\mathsf{d}\hat{\boldsymbol{d}}_{k+i}\ , \tag{47d}$$

$$\hat{\boldsymbol{x}}_{k+i+1} \in \mathcal{X}_{k+i+1}\ ,\ \hat{\boldsymbol{u}}_{k+i} \in \mathcal{U}_{k+i}\ , \tag{47e}$$

$$i \in \mathcal{N}\ , \tag{47f}$$

where the subscript d in (47d) indicates that the matrices are discretised. We obtain such a discrete system using, e.g., *zero-order hold*. That is, we assume that the input variable is constant during each preferably small time sample $\boldsymbol{u}(t) = \boldsymbol{u}_k$, for $t \in [t_k, t_{k+1}]$, $k \in \mathcal{N}$. (47d) describes the dynamics of the system and provides the so-called *Kalman predictions* given by the recursion. The disturbances, $\{\hat{\boldsymbol{d}}_{k+i|k}\}_{i\in\mathcal{N}}$, are again obtained from the separate disturbance model. We let the cost-to-go term be zero $\ell_{N_p}(\hat{\boldsymbol{x}}_{k+N_p}) = 0$. But it can be very important to include in some cases. For example, when batteries are included, the controller will try to sell all stored electricity (which we do not immediately want) since it minimises the cost. Unless we include a cost-to-go term, that weights the value of the electricity left in the battery [57].

The cost function is of crucial importance in terms of defining the behaviour of the controller. It is important that it minimises a term that reflects the desired behaviour and ensures stability. The latter is usually not a problem when dealing with systems of slow dynamics such as the temperature of a building. Often the cost

function is minimising some distance between a control variable and a set point. For building climate control, the control variable can be the room air temperature and the set point can be the desired temperature. Two common examples of cost functions are the following:

$$\text{Quadratic cost} \qquad \boldsymbol{x}^T \mathsf{Q}\boldsymbol{x} + \boldsymbol{u}^T \mathsf{R}\boldsymbol{u} \ , \tag{48a}$$

$$\text{Economic (linear) cost} \qquad \boldsymbol{c}^T \boldsymbol{u} \ . \tag{48b}$$

The quadratic cost function typically minimises a relative weighting between the variables and can provide a trade-off between the input and the regulation of the system. The linear cost function measures an amount of some resource. In temperature regulation of a building, it is often the energy consumption or price. But in general, the resource is an abstract size and can also measure the CO_2 emission from electricity generation or even a generic penalty signal manually designed to force a certain behaviour. For our purpose, we use *economic MPC*, while also softening the constraints

$$J(\hat{\boldsymbol{x}}_{k|k}, \{\hat{\boldsymbol{d}}_{k+i|k}\}_{i\in\mathcal{N}}) = \min_{\hat{\boldsymbol{u}}_k, \hat{\boldsymbol{s}}_k} \sum_{i\in\mathcal{N}} \boldsymbol{c}_{k+i}\hat{\boldsymbol{u}}_{k+i} + \sum_{i\in\mathcal{N}^+} \boldsymbol{\rho}_{k+i}\hat{\boldsymbol{s}}_{k+i} \ , \tag{49a}$$

$$s.t. \quad \hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_{k|k} \ , \tag{49b}$$

$$\hat{\boldsymbol{d}}_{k+i} = \hat{\boldsymbol{d}}_{k+i|k} \ , \qquad i \in \mathcal{N} \ , \tag{49c}$$

$$\hat{\boldsymbol{x}}_{k+i+1} = \mathsf{A}_\mathsf{d}\hat{\boldsymbol{x}}_{k+i} + \mathsf{B}_\mathsf{d}\hat{\boldsymbol{u}}_{k+i} + \mathsf{E}_\mathsf{d}\hat{\boldsymbol{d}}_{k+i} \ , \qquad i \in \mathcal{N} \ , \tag{49d}$$

$$\hat{\boldsymbol{y}}_{k+i} = \mathsf{C}_\mathsf{d}\hat{\boldsymbol{x}}_{k+i} \ , \qquad i \in \mathcal{N}^+ \ , \tag{49e}$$

$$\hat{\boldsymbol{y}}_{k+i} - \hat{\boldsymbol{s}}_{k+i} \leq \boldsymbol{y}_{max,k+i} \ , \qquad i \in \mathcal{N}^+ \ , \tag{49f}$$

$$\boldsymbol{y}_{min,k+i} \leq \hat{\boldsymbol{y}}_{k+i} + \hat{\boldsymbol{s}}_{k+i} \ , \qquad i \in \mathcal{N}^+ \ , \tag{49g}$$

$$\Delta\boldsymbol{u}_{min,k+i} \leq \Delta\hat{\boldsymbol{u}}_{k+i} \leq \Delta\boldsymbol{u}_{max,k+i} \ , \qquad i \in \mathcal{N} \ , \tag{49h}$$

$$\boldsymbol{u}_{min,k+i} \leq \hat{\boldsymbol{u}}_{k+i} \leq \boldsymbol{u}_{max,k+i} \ , \qquad i \in \mathcal{N} \ , \tag{49i}$$

$$\boldsymbol{0} \leq \hat{\boldsymbol{s}}_{k+i} \ , \qquad i \in \mathcal{N}^+ \ , \tag{49j}$$

where $\mathcal{N}^+ = \{1, 2, \ldots, N_p\}$, $\Delta\hat{\boldsymbol{u}}_{min,i}$ and $\Delta\hat{\boldsymbol{u}}_{max,i}$ are the minimum and maximum allowed changes of input, and $\hat{\boldsymbol{s}}_i$ and $\boldsymbol{\rho}_i$ are the slack variable and slack penalty, respectively. $\boldsymbol{u}_{max,i}$ and $\boldsymbol{u}_{min,i}$ are the upper and lower constraints on the input, and $\hat{\boldsymbol{y}}_{max,i}$ and $\hat{\boldsymbol{y}}_{min,i}$ are the upper and lower constraints on the observed variables.

The slack variable has the purpose of *softening* the constraints. That is, it allows the solution to move outside of the constraints without making the problem infeasible—but at a cost!

6.1.1 Rewriting the State Equations of the Optimisation Problem

We are now familiar with the objective of MPC and what the purpose of the constraints is, but it is not directly clear how to write out the optimisation problem such that we can implement it. First, we need to recognise that the variables of the optimisation problem are the input and the slack variables, $[\boldsymbol{u}_k, \boldsymbol{s}_k]$. That is, we need to write each constraint in (49) as an equation using $\boldsymbol{u}_k$ and $\boldsymbol{s}_k$. To do this, we use the Kalman predictions of the system to obtain a matrix expression for the states for all prediction times in $\mathcal{N}$. Writing out the observed system using the Kalman predictions is

$$\begin{aligned}\hat{\boldsymbol{y}}_k = \mathsf{C_d}\hat{\boldsymbol{x}}_k &= \mathsf{C_d}(\mathsf{A_d}\hat{\boldsymbol{x}}_{k-1} + \mathsf{B_d}\hat{\boldsymbol{u}}_{k-1} + \mathsf{E_d}\hat{\boldsymbol{d}}_{k-1}) \,, \\ &= \mathsf{C_d}\mathsf{A_d}\hat{\boldsymbol{x}}_{k-1} + \mathsf{C_d}\mathsf{B_d}\hat{\boldsymbol{u}}_{k-1} + \mathsf{C_d}\mathsf{E_d}\hat{\boldsymbol{d}}_{k-1} \,.\end{aligned} \tag{50}$$

The state development for $\hat{\boldsymbol{x}}_k$ is again given by the Kalman predictions, where $\hat{\boldsymbol{x}}_{k-1} = \mathsf{A_d}\hat{\boldsymbol{x}}_{k-2} + \mathsf{B_d}\hat{\boldsymbol{u}}_{k-2} + \mathsf{E_d}\hat{\boldsymbol{d}}_{k-2}$. Inserting this into (50) yields

$$\begin{aligned}\hat{\boldsymbol{y}}_k &= \mathsf{C_d}\mathsf{A_d}\left(\mathsf{A_d}\hat{\boldsymbol{x}}_{k-2} + \mathsf{B_d}\hat{\boldsymbol{u}}_{k-2} + \mathsf{E_d}\hat{\boldsymbol{d}}_{k-2}\right) + \mathsf{C_d}\mathsf{B_d}\hat{\boldsymbol{u}}_{k-1} + \mathsf{C_d}\mathsf{E_d}\hat{\boldsymbol{d}}_{k-1} \,, \\ &= \mathsf{C_d}{\mathsf{A_d}}^2\hat{\boldsymbol{x}}_{k-2} + \mathsf{C_d}(\mathsf{A_d}\mathsf{B_d}\hat{\boldsymbol{u}}_{k-2} + \mathsf{B_d}\hat{\boldsymbol{u}}_{k-1}) + \mathsf{C_d}(\mathsf{A_d}\mathsf{E_d}\hat{\boldsymbol{d}}_{k-2} + \mathsf{E_d}\hat{\boldsymbol{d}}_{k-1}) \,.\end{aligned}$$

Continuing this approach until an initial state is reached (and shifting the time to start at t_k and end at t_{k+Np}), the result is

$$\hat{\boldsymbol{y}}_{k+N_p} = \mathsf{C_d}{\mathsf{A_d}}^k\hat{\boldsymbol{x}}_k + \mathsf{C_d}\sum_{i=0}^{N_p-1}{\mathsf{A_d}}^i\mathsf{B_d}\hat{\boldsymbol{u}}_{k+N_p-1-i} + \mathsf{C_d}\sum_{i=0}^{k-1}{\mathsf{A_d}}^i\mathsf{E_d}\hat{\boldsymbol{d}}_{k+N_p-1-i} \,. \tag{51}$$

Let $\hat{\boldsymbol{Y}}_{k+1}$ be a vector containing the predictions N_p steps ahead starting from t_{k+1}, $\hat{Y}_{k+1} = [\hat{\boldsymbol{y}}_{k+1}^T, \hat{\boldsymbol{y}}_{k+2}^T, \ldots, \hat{\boldsymbol{y}}_{k+N_p}^T]^T$. Then (51) shows how to formulate an expression for $\hat{\boldsymbol{y}}_{k+i}$ for $i \in \mathcal{N}^+$ using a convenient matrix-vector notation

$$\boxed{\hat{\boldsymbol{Y}}_{k+1} = \Phi\hat{\boldsymbol{x}}_k + \Gamma\boldsymbol{U}_k + \Pi\hat{\boldsymbol{D}}_k}, \tag{52}$$

where

$$\Phi = \begin{bmatrix} C_d A_d \\ C_d A_d{}^2 \\ C_d A_d{}^3 \\ \vdots \\ C_d A_d{}^{N_p} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} C_d B_d & 0 & \dots & \dots & 0 \\ C_d A_d B_d & C_d B_d & \ddots & & 0 \\ C_d A_d{}^2 B_d & CAB & C_d B_d & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ C_d A_d{}^{N_p-1} B_d & C_d A_d{}^{N_p-2} B_d & \dots & \dots & C_d B_d \end{bmatrix},$$

$$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{u}_{k+1} \\ \boldsymbol{u}_{k+2} \\ \vdots \\ \boldsymbol{u}_{k+N_p-1} \end{bmatrix}, \quad \Pi = \begin{bmatrix} C_d E_d & 0 & \dots & \dots & 0 \\ C_d A_d E_d & C_d E_d & \ddots & & 0 \\ C_d A_d{}^2 E_d & C_d A_d E_d & C_d E_d & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ C_d A_d{}^{N_p-1} E & C_d A_d{}^{N_p-2} E_d & \dots & \dots & C_d E_d \end{bmatrix},$$

$$\hat{\boldsymbol{D}}_k = \begin{bmatrix} \hat{\boldsymbol{d}}_k \\ \hat{\boldsymbol{d}}_{k+1} \\ \hat{\boldsymbol{d}}_{k+2} \\ \vdots \\ \hat{\boldsymbol{d}}_{k+N_p-1} \end{bmatrix}. \tag{53}$$

6.1.2 Rewriting the Constraints in the Optimisation Problem

Now that we have an expression for $\hat{\boldsymbol{y}}_{k+i}$, $i \in \mathcal{N}^+$, we are able to eliminate the dependence on $\hat{\boldsymbol{y}}_{k+i}$ in the constraints. Starting with (49f)

$$\begin{aligned} & \hat{\boldsymbol{Y}}_k - \boldsymbol{S}_{k+1} \leq \boldsymbol{Y}_{max} , \\ \implies \quad & \Phi \hat{\boldsymbol{x}}_k + \Gamma \boldsymbol{U}_k + \Pi \hat{\boldsymbol{D}}_k - \boldsymbol{S}_{k+1} \leq \boldsymbol{Y}_{max} , \\ \implies \quad & \Gamma \boldsymbol{U}_k - \boldsymbol{S}_{k+1} \leq \boldsymbol{Y}_{max} - \Phi \hat{\boldsymbol{x}}_k - \Pi \hat{\boldsymbol{D}}_k , \end{aligned}$$

where $\boldsymbol{S}_{k+1} = \{\boldsymbol{s}_{k+i}\}_{i \in \mathcal{N}^+}$ is a vector with the slack variables. We can do the same thing with the lower constraint for $\hat{\boldsymbol{y}}_{k+i}$,

$$-\Gamma \boldsymbol{U}_k - \boldsymbol{S}_{k+1} \leq -\boldsymbol{Y}_{min} + \Phi \hat{\boldsymbol{x}}_k + \Pi \hat{\boldsymbol{D}}_k . \tag{54}$$

To rewrite (49h), we need the following transcription:

$$\begin{bmatrix} \Delta \boldsymbol{u}_{min} + \boldsymbol{u}_{-1} \\ \Delta \boldsymbol{u}_{min} \\ \vdots \\ \Delta \boldsymbol{u}_{min} \end{bmatrix} \leq \begin{bmatrix} \mathsf{I} & \mathsf{0} & \cdots & \mathsf{0} \\ -\mathsf{I} & \mathsf{I} & \mathsf{0} & \\ \mathsf{0} & -\mathsf{I} & \mathsf{I} & \mathsf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ & & \mathsf{0} & -\mathsf{I} & \mathsf{I} & \mathsf{0} \\ \mathsf{0} & \cdots & & \mathsf{0} & -\mathsf{I} & \mathsf{I} \end{bmatrix} \boldsymbol{U}_k \leq \begin{bmatrix} \Delta \boldsymbol{u}_{max} + \boldsymbol{u}_{-1} \\ \Delta \boldsymbol{u}_{max} \\ \vdots \\ \Delta \boldsymbol{u}_{max} \end{bmatrix}, \tag{55}$$

where $\boldsymbol{u}_{-1}$ is the input given to the system at time t_{k-1}, and I is the identity matrix with the same size as the length of $\boldsymbol{u}_k$. Denoting the matrix in (55) by Λ, then the constraint in (49h) is

$$\begin{aligned} -\Lambda \boldsymbol{U}_k &\leq -\Delta \boldsymbol{U}_{min} \,, \\ \Lambda \boldsymbol{U}_k &\leq \Delta \boldsymbol{U}_{max} \,. \end{aligned} \tag{56}$$

(49i) is straightforward in the sense that it requires no further notational introduction

$$\begin{aligned} -\boldsymbol{U}_k &\leq -\boldsymbol{U}_{min} \,, \\ \boldsymbol{U}_k &\leq \boldsymbol{U}_{max} \,. \end{aligned} \tag{57}$$

Finally, we demand the slack variables to be non-negative, $-\boldsymbol{S}_{k+1} \leq 0$. We are now able to write the problem in (49) as an expression of the input and slack variables

$$\begin{aligned} J = \min_{\boldsymbol{U}_k, \boldsymbol{S}_{k+1}} \quad & \begin{bmatrix} \boldsymbol{C}_u^T & \boldsymbol{P}_s^T \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_k \\ \boldsymbol{S}_{k+1} \end{bmatrix} \\ s.t. \quad & \begin{bmatrix} -\Gamma & -\mathsf{I} \\ \Gamma & -\mathsf{I} \\ -\mathsf{I} & \mathsf{0} \\ \mathsf{I} & \mathsf{0} \\ -\Lambda & \mathsf{0} \\ \Lambda & \mathsf{0} \\ \mathsf{0} & -\mathsf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_k \\ \boldsymbol{S}_{k+1} \end{bmatrix} \leq \begin{bmatrix} -\boldsymbol{Y}_{min} + \Phi \hat{\boldsymbol{x}}_k + \Pi \hat{\boldsymbol{D}}_k \\ \boldsymbol{Y}_{max} - \Phi \hat{\boldsymbol{x}}_k - \Pi \hat{\boldsymbol{D}}_k \\ -\boldsymbol{U}_{min} \\ \boldsymbol{U}_{max} \\ -\Delta \boldsymbol{U}_{min} \\ \Delta \boldsymbol{U}_{max} \\ \mathbf{0} \end{bmatrix}, \end{aligned} \tag{58}$$

where $\boldsymbol{C}_u = [\boldsymbol{c}_k^T, \ldots, \boldsymbol{c}_{k+N_p-1}^T]^T$ and $\boldsymbol{P}_s = [\boldsymbol{\rho}_{k+1}^T, \ldots, \boldsymbol{\rho}_{k+N_p}^T]^T$ are the electricity costs and the slack variable penalty, respectively. We have now written the optimisation problem in (49) as a constrained *linear program* that gives us the optimal input, $\hat{\boldsymbol{U}}_{\boldsymbol{k}}{}^* = [\hat{\boldsymbol{u}}_k^{*T}, \ldots, \hat{\boldsymbol{u}}_{k+N_p-1}^{*T}]^T$, that minimises the cost J based on initial conditions for the system, $\hat{\boldsymbol{x}}_{k|k}$, and disturbance forecasts, $\{\hat{\boldsymbol{d}}_{k+i}\}_{i \in \mathcal{N}}$. This is referred to as *optimal control*. Figure 17 displays the overall MPC framework and how the elements interact. Very often, for systems governed by uncertainty, it is necessary to use a *moving horizon scheme*, where only the current optimal input

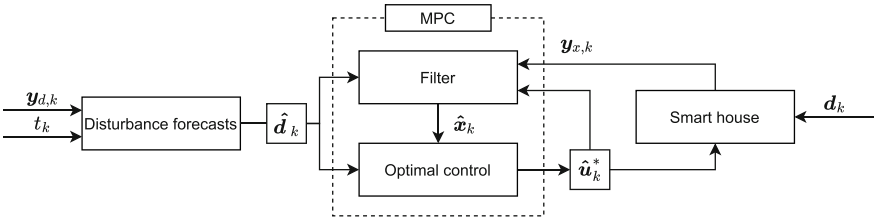

Fig. 17 The MPC framework for the smart building and how the disturbance model is incorporated

is applied to the system, $\boldsymbol{u}_k^*$. When arriving at the next time step, t_{k+1}, the optimal control problem is computed again and the current input is applied. This is known as *closed-loop feedback control* and ensures stability of the system as the controller can account for unforeseen uncertainty in the system between time steps. As this is a linear in-equality constrained problem, a closed-form solution to (58) does not exist in general. We therefore use numerical optimisation to find the unique solution that exists due to convexity as long as the problem is well posed.

6.2 Offset-Free Control Without Separate Disturbance Model

In more conventional MPC setups where separately modelled disturbance models are not feasible, there exist ways to deal with unforeseen disturbances. In practice, parameter uncertainties, lack of model accuracy, and non-modelled disturbances all usually necessitate some kind of action; otherwise, *offsets* can arise. For example, if the disturbances act with a constant (or slowly varying) force, we can obtain a non-zero distance between the system state and the desired set point. The literature suggests multiple ways to deal with this [51, 52, 55], and this is still an active research area. A popular method (among others) is known as the *augmented disturbance model*. In practice, two variants are widely used and well studied: the *input* and *output* disturbance models.

Morari and Stephanopoulos [47] derive some important concepts and results regarding disturbance modelling for continuous-time systems in the deterministic case. Consider a continuous-time linear state-space system of the form

$$\begin{aligned} \mathrm{d}\boldsymbol{x}(t) &= (\mathsf{A}\boldsymbol{x}(t) + \mathsf{B}\boldsymbol{u}(t))\,\mathrm{d}t\ , \\ \boldsymbol{y}_k &= \mathsf{C}\boldsymbol{x}(t_k)\ , \end{aligned} \tag{59}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\mathsf{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{u} \in \mathbb{R}^{n_u}$, $\mathsf{B} \in \mathbb{R}^{n\times n_u}$, and $\mathsf{C} \in \mathbb{R}^{n_y}$. Note that the disturbances are not a part of the model in the first place. The augmented disturbance model approach assumes that the disturbances act on the system as integrated white

noise; that is, we can add an *integrator* as an independent state, $\boldsymbol{\eta} \in \mathbb{R}^{n_d}$ in the system by

$$\begin{aligned} \mathrm{d}\boldsymbol{x}(t) &= \left(\mathsf{A}\boldsymbol{x}(t) + \mathsf{B}\boldsymbol{u}(t) + \bar{\mathsf{B}}\boldsymbol{\eta}(t)\right)\mathrm{d}t \ , \\ \mathrm{d}\boldsymbol{\eta}(t) &= \bar{\Sigma}\mathrm{d}\bar{\boldsymbol{\omega}}(t) \ , \\ \boldsymbol{y}_k &= \mathsf{C}\boldsymbol{x}(t_k) + \bar{\mathsf{C}}\boldsymbol{\eta}(t_k) \ , \end{aligned} \tag{60}$$

where $\bar{\mathsf{B}} \in \mathbb{R}^{n \times n_d}$ and $\bar{\mathsf{C}} \in \mathbb{R}^{n \times n_d}$ are the disturbances on the input and output (hence the name). We shall assume $C_d = \mathbf{0}$ in the rest of this section. The case when the disturbances only act on the system is called *input disturbances* because it acts as an input on the system. We can augment the disturbance and obtain the *augmented* system

$$\begin{aligned} \mathrm{d}\begin{bmatrix}\boldsymbol{x}(t)\\ \boldsymbol{\eta}(t)\end{bmatrix} &= \left(\begin{bmatrix}\mathsf{A} & \bar{\mathsf{B}}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}(t)\\ \boldsymbol{\eta}(t)\end{bmatrix} + \begin{bmatrix}\mathsf{B}\\ 0\end{bmatrix}\boldsymbol{u}(t)\right)\mathrm{d}t + \begin{bmatrix}0\\ \bar{\Sigma}\end{bmatrix}\mathrm{d}\bar{\boldsymbol{\omega}}(t) \ , \\ \boldsymbol{y}_k &= \begin{bmatrix}\mathsf{C} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}(t)\\ \boldsymbol{\eta}(t)\end{bmatrix} \ . \end{aligned} \tag{61}$$

Note that we are (obviously) not able to influence the disturbances and (in general) know nothing about them. However, it is crucial for us to *estimate* them in order to obtain offset-free control. *Observability* is an important concept that relates to whether we are able to estimate all states in the given system, $\boldsymbol{x}$, based on the observed information we have, $\boldsymbol{y}$. We say that the system (C, A) in (59) is observable if

$$\operatorname{rank}\begin{bmatrix}\mathsf{CA}\\ \mathsf{CA}^2\\ \vdots\\ \mathsf{CA}^{n-1}\end{bmatrix} = n \ . \tag{62}$$

In general, we are able to estimate all states in a system if and only if it is observable. That is, we need to make sure that the augmented system in (61) is observable—otherwise we cannot estimate the disturbances and in turn not obtain offset-free control. The system $\left(\begin{bmatrix}\mathsf{C} & 0\end{bmatrix}, \begin{bmatrix}\mathsf{A} & \bar{\mathsf{B}}\\ \mathsf{C} & 0\end{bmatrix}\right)$ is observable if and only if the following requirements are fulfilled [47]:

1. The system (C, A) is observable.
2. $\operatorname{rank}\begin{bmatrix}\mathsf{A} & \bar{\mathsf{B}}\\ \mathsf{C} & 0\end{bmatrix} = n + n_d$.

This implies that we are able to insert at most n_y (the number of independently observed variables) integrators into the system while ensuring observability of the augmented system. We can estimate the disturbance states simply by using the Kalman filter or Luenberger observer [32] (treating them as any other hidden state). This method also supplies disturbance forecasts by computing the predictions supplied by the system. It is easy to see that it corresponds to zero-order disturbance forecasts also called *persistent forecasts*, see, e.g., [38, p. 333] and [11]. For this reason, the integrator approach works best when the disturbance dynamics are slow—and not very well for faster dynamics such as the solar radiation for a smart building. We will show this in the next section.

7 Predictive Control with Embedded Disturbance Models

We now combine the individual weather models in the previous sections into a combined disturbance model framework. Ultimately, we want to show that by modelling the disturbances, we can obtain more accurate control than using, e.g., augmented integrators. The advanced disturbance model should return a vector, $\boldsymbol{d}(t) = \left[d_{T_a}(t), d_{\phi_s}(t)\right]^T$, containing the solar radiation and ambient air temperature. Writing up all equations for the individual disturbances gives the following complete description:

$$
\begin{aligned}
\text{Cloud cover model} &\begin{cases} \mathrm{d}Z_\kappa = f_\psi(Z_\kappa)\mathrm{d}t + \sigma_\psi \mathrm{d}\omega_\kappa \\ c = \zeta^{-1}(\psi^{-1}(Z_\kappa)) \end{cases} \\
\text{Solar radiation model} &\begin{cases} \phi_s = I_N(c,t) + I_D(c,t) \end{cases} \\
\text{Net radiation model} &\begin{cases} R_n = R_n(c, \phi_s, t) \end{cases} \\
\text{Air temperature model} &\begin{cases} \mathrm{d}T_w = f_{T_w}(T_l, T_w)\mathrm{d}t + \sigma_w \mathrm{d}\omega_w \\ \mathrm{d}T_l = f_{T_l}(T_l, T_w, R_n)\mathrm{d}t + \sigma_l \mathrm{d}\omega_l \end{cases} \\
\text{Observations} &\begin{cases} d_{\phi_s} = \phi_s + v_{\phi_s}, \quad v_{\phi_s} \sim N(0, R_{\phi_s}) \\ d_{T_a} = T_l + v_{T_a}, \quad v_{T_a} \sim N(0, R_{T_a}) \\ \boldsymbol{d} = \left[d_{T_a}, d_{\phi_s}\right]^T. \end{cases}
\end{aligned}
\tag{63}
$$

Since the disturbance model in (63) is based on SDEs, we are able to use the CDEKF to compute *certainty equivalent* Kalman predictions—which MPC requires. This procedure requires numerical solutions of coupled differential equations, which in turn requires initial conditions preferably from observations coming from the building site in order to ensure accuracy. Due to the one-way coupling of

the individual weather models, the computation of the predictions becomes much easier, as it can be split into smaller and simpler calculations:

1. Compute the cloud cover predictions $\{\hat{c}_{k+i|k}\}_{i\in\mathcal{N}}$.
2. Compute the solar radiation predictions $\{\hat{\phi}_{s,k+i|k}\}_{i\in\mathcal{N}}$.
3. Compute the net radiation predictions $\{\hat{R}_{n,k+i|k}\}_{i\in\mathcal{N}}$.
4. Compute the ambient air temperature predictions $\{\hat{T}_{a,k+i|k}\}_{i\in\mathcal{N}}$.

7.1 Comparison of Advanced Disturbance Forecasts and Persistent Forecasts

All the necessary elements are now introduced for us to demonstrate how to control the room air temperature of a smart building presented in Sect. 4 using the advanced disturbance forecasts from (63). Additionally, we want to show that the great effort put into modelling the disturbances actually improves the quality of the smart building regulation. To do this, we compare the advanced disturbance forecasts with a more typical and conventional kind of offset-free control that is explained in Sect. 6.2. In this text, we use *persistent forecasts*, which are often used as a reference model for weather and energy forecasting models. It uses the following constant predictions:

$$\hat{\boldsymbol{d}}_{k+i|k} = \hat{\boldsymbol{d}}_{k|k}\,, \quad i \in \mathcal{N}\,. \tag{64}$$

That is, we assume that all future disturbances equal the disturbance at the present time, t_k, and we assume that we actually observe them.

Figure 18 shows the persistent and advanced disturbance forecasts using a prediction horizon of 96 h. The simplicity of the persistent forecasts becomes very

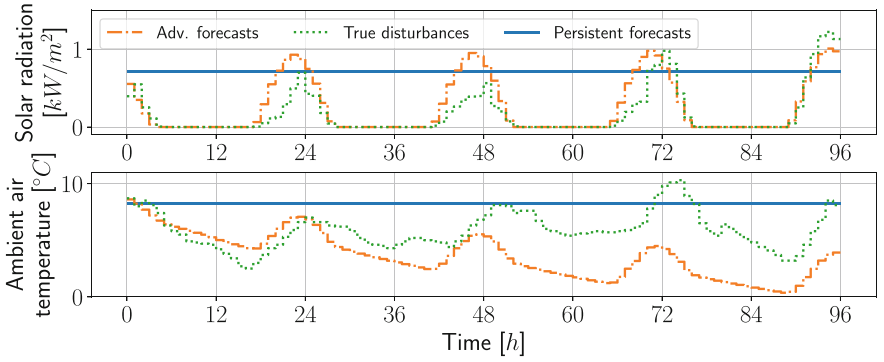

Fig. 18 The persistent forecasts against the advanced disturbance forecasts. The latter is computed by integrating (63) forward using the current observations as initial conditions

visible compared to the complex dynamics of the true disturbances. The advanced disturbance forecasts are of course most accurate in a short future time span due to the initial conditions. They then drift towards some stationary dynamics, highly dictated by the stationary points of the cloud cover that can be seen in Fig. 9. We therefore cannot hope to accurately forecast the disturbances 96 hours into the future using these methods—instead they give *expected disturbance values*. For this reason, the literature normally uses meteorological forecasts. They are, however, less accurate for short-term predictions. [66] suggest that in practice, the advanced disturbance forecasts work best 4–10 h into the prediction horizon and from that point on meteorological forecasts in general perform better. The latter is based on large systems of differential equations and is calculated using very powerful computers. In practice, it is believed that a combination of short- and long-term forecasts will be the best solution.

8 Simulation Results

As previously mentioned, we use data from March as the true disturbances acting on the smart building. This gives us 7 months of data to simulate control of the smart building using the two forecasting schemes. In this section, we show the results of controlling the smart buildings presented in Sect. 4. Furthermore, we present the results where the heat pump is combined with both electrical heaters and air conditioners (for cooling). Recall that the heat pump is a factor 3 more efficient compared to the electrical heaters, which makes it economically attractive and interesting to combine.

All simulations in this section use a prediction horizon of 96 h, a time sample of 1 hour, and temperature constraints $T_{r,min}$=20 C° and $T_{r,max}$=24 C° (which are softened). We use the slack penalty suggested by [57], $\rho_k = 5000$. The electricity price is taken from Nord Pool and is the average over all March data and equals to $c_k = 0.27 \cdot 10^{-3}$ EUR/Wh.

Figure 19 shows a 15-day sample of the 7 months of simulation for two smart buildings: one using electrical heaters and one using a heat pump. The smart building equipped with electrical heaters acts faster and is therefore more capable of adjusting to sudden changes from the disturbances. This is borne out by the electrical heaters that operate at a level that sets the room air temperature to the lower constraint to minimise costs—except when the sun shines and additional heat is not needed. The differences between two forecasting schemes are not greatly visible from this sample, however, due to the effect of the control feedback every hour.

The solution for the smart building equipped with a heat pump looks much different. The overall dynamics are much slower. In contrast to the case with the electrical heaters, the advanced disturbance forecasts seem to enable the controller

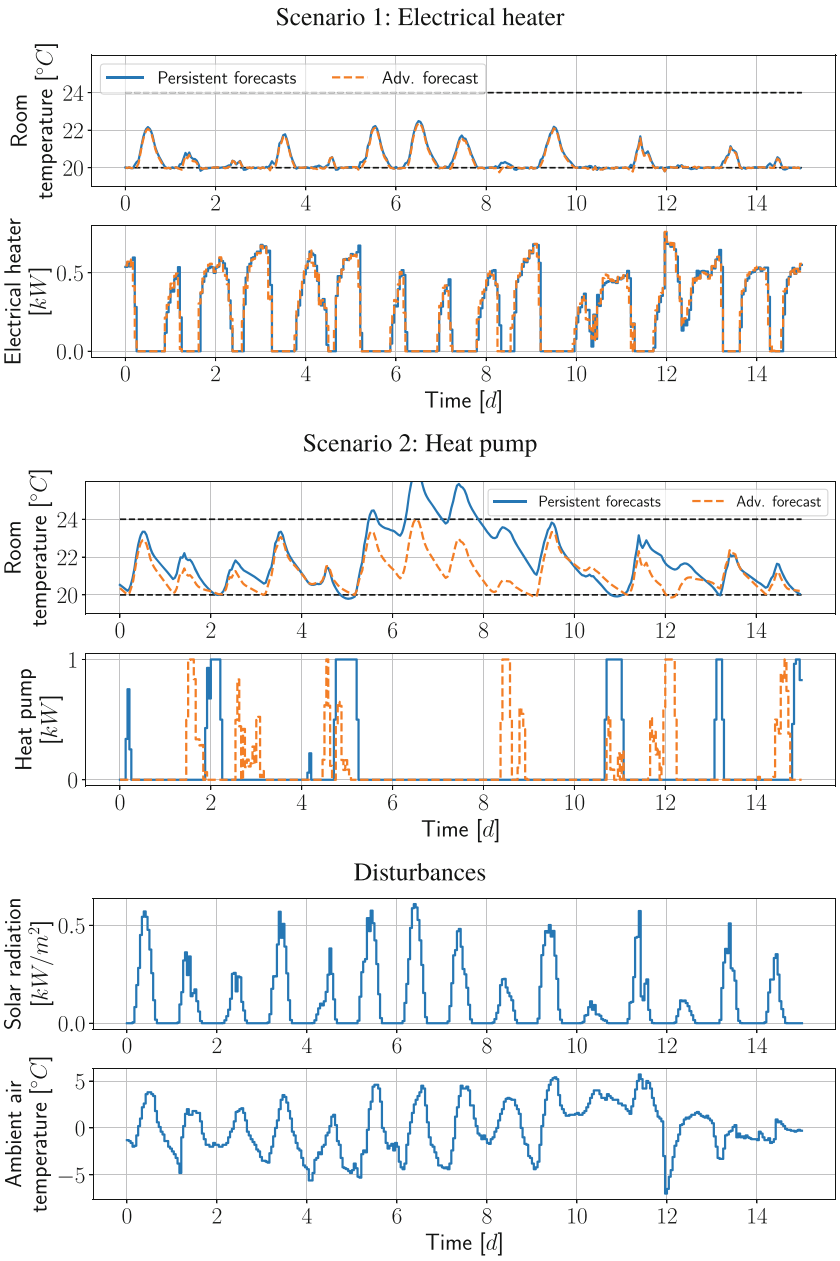

Fig. 19 A 15-day sample of the total 7 months of simulation. It shows the indoor air temperature and the heat input for the two scenarios at the same point in the time series of simulation. The black dashed lines are the constraints

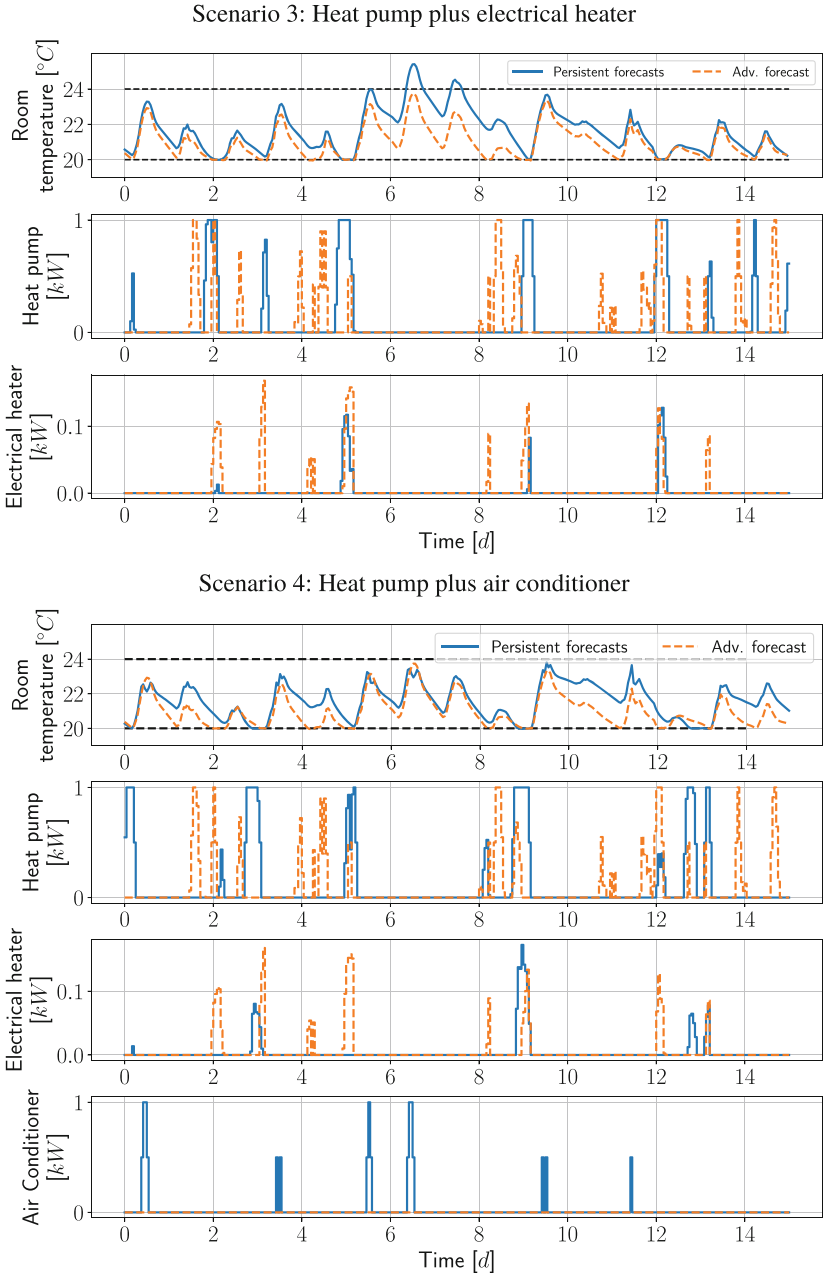

Fig. 20 A 15-day sample of the total 7 months of simulation. It shows the indoor air temperature and the heat input for the two scenarios where the heat pump is combined with faster heating inputs

to much better keep the room air temperature on the right side of the constraints—the baseline forecasts go above the upper constraint a couple of times during this sample.

Figure 20 shows a 15-day sample of two extended smart buildings. The 3rd scenario is a simulation of a smart building equipped with both a heat pump and an electrical heater. This enables the smart building to heat efficiently using the heat pump but also to make fast corrections using the electrical heaters. The 4th scenario considers a smart building equipped with a heat pump and an air conditioner such that it is also able to cool if necessary. Visible in both scenarios is the fact that the advanced disturbance forecasts use the expensive electrical heaters and air conditioners less often and are therefore able to obtain cheaper control.

In an attempt to draw asymptotic conclusions, we turn to consider how well the forecasting schemes minimise the actual cost function in (49) of the entire simulation of the 7 months, as this is what the solutions are based on. Table 8 shows the constraint violations of the entire simulations corresponding to the second term in (49a). Additionally, it shows the results for a controller that uses perfect forecasts: this gives a theoretical upper boundary on the performance using the settings in this chapter. The advanced disturbance forecasts seem to outperform the persistent forecasts in all scenarios. Especially in the case of the heat pump alone: this is perhaps the most realistic case—that houses equipped with a heat pump do not have addition heating or cooling (at least in Denmark).

Looking at the cost term in (49a), Table 9 shows the total electricity cost for all scenarios. It is obvious that the cost for the electrical heaters is almost identical for all scenarios since the total heat needed is the same. In the heat pump scenario, however, the advanced disturbance forecasts use much less electricity compared with the persistent forecasts. This is also the case for scenarios 3 and 4—the advanced disturbance forecasts seem to offer a significant decrease in electricity consumption and in general are very close to the perfect forecasts. This is also visible from the simulation samples in Figs. 19 and 20 where the advanced disturbance forecasts almost at all times lie below the persistent forecasts.

Table 8 The constraint violations (the second term in the cost function in (49a)) for all heating strategies for each forecasting scheme

Constraint violation of the control simulations			
Forecasting method	Persistent	Advanced disturbances	Perfect
Scenario 1: Electrical heater	48.5	39.6	25.11
Scenario 2: Heat pump	157.9	12.3	1.7
Scenario 3: Heat pump plus electrical heater	48.0	6.7	1.2
Scenario 4: Heat pump plus AC	4.4	2.4	0

9 Hierarchical Control

In the next section, we will first illustrate how the controllers described in the previous sections can be considered as the low-level controllers of a multi-level or hierarchical control setup for solving grid or ancillary service problems in future smart energy systems. Subsequently, we shall briefly outline how these principles can be generalised to multi-level and hierarchical control problems. This section will also outline how to establish a connection between the multi-level control problems and conventional electricity markets.

9.1 Two-level Control for Utilising Energy Flexibility

In the previous sections, it has been shown how to develop controllers for controlling smart buildings according to forecasts of prices, weather conditions, and indoor climate requirements. In this section, it will be explained how to leverage this by generating prices that are used indirectly to control the demand of the smart buildings. The basic concept is illustrated by Fig. 21, where a smart building, from an external perspective, takes an input (price) and gives an output (demand). Analysed in this way, a model, termed the Flexibility Function, can be developed that predicts demand as a dynamic function of price. The Flexibility Function could be any dynamic model. In [18], a linear model (finite impulse response model) is suggested, but in [19], it is shown that a grey-box model using stochastic differential equations is more appropriate.

Table 9 The electricity price in EUR (the first term in the cost function in (49a)) for all heating strategies for each forecasting scheme

Electricity cost of the simulations			
Forecasting method	Persistent	Advanced disturbances	Perfect
Scenario 1: Electrical heater	303.2	302.2	302.0
Scenario 2: Heat pump	117.3	110.4	107.7
Scenario 3: Heat pump plus electrical heater	113.0	108.2	107.5
Scenario 4: Heat pump plus AC	117.9	108.3	107.5

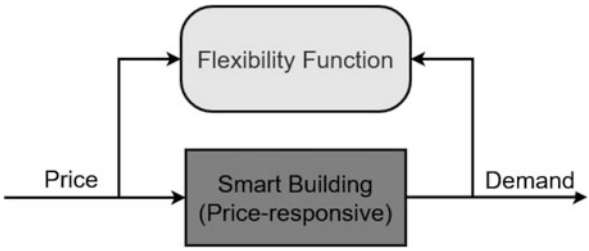

Fig. 21 The demand of a smart building can be predicted as a function of prices

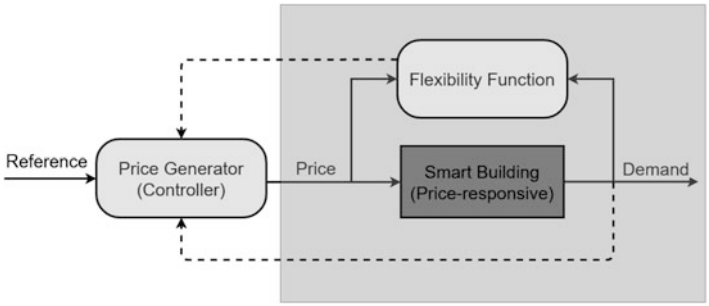

Fig. 22 Using a Flexibility Function to generate price signals and demand as control feedback

Once a Flexibility Function has been estimated, a second controller can be formulated where the objective is to control demand according to some criteria, and the decision variable is the price. As shown in Fig. 22, the Flexibility Function can be used to generate prices according to some reference. Notice how the demand acts as the feedback to the controller, closing the loop.

If FF is the Flexibility Function that takes prices as input and gives expected demand as output, while r_l is a reference load, then a naive upper-level optimisation problem can be written as

$$\min_{\boldsymbol{C}_u} \quad (\mathrm{FF}(\boldsymbol{C}_u) - r_l)^2. \tag{65}$$

Obviously, it might be necessary to impose limits on how much the price can change, requirements on the average value, and a more sophisticated optimisation problem than the minimum variance formulation as discussed in [17]. Combining this optimisation problem with the one presented in (58) reveals how the price signal, $\boldsymbol{C}_u$, couples the two in an elegant fashion

$$\begin{aligned}
&\min_{\boldsymbol{C}_u} \quad (\mathrm{FF}(\boldsymbol{C}_u) - r_l)^2 && \text{Upperlevel} \\
&\hline \\
&\min_{\boldsymbol{U}_k, \boldsymbol{S}_{k+1}} \quad \begin{bmatrix} \boldsymbol{C}_u{}^T & \boldsymbol{P}_s^T \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_k \\ \boldsymbol{S}_{k+1} \end{bmatrix} && \text{Lowerlevel} \\
&s.t. \quad \begin{bmatrix} -\Gamma & -\mathsf{I} \\ \Gamma & -\mathsf{I} \\ -\mathsf{I} & 0 \\ \mathsf{I} & 0 \\ -\Lambda & 0 \\ \Lambda & 0 \\ 0 & -\mathsf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_k \\ \boldsymbol{S}_{k+1} \end{bmatrix} \leq \begin{bmatrix} -\boldsymbol{Y}_{min} + \Phi \hat{\boldsymbol{x}}_k + \Pi \hat{\boldsymbol{D}}_k \\ \boldsymbol{Y}_{max} - \Phi \hat{\boldsymbol{x}}_k - \Pi \hat{\boldsymbol{D}}_k \\ -\boldsymbol{U}_{min} \\ \boldsymbol{U}_{max} \\ -\Delta \boldsymbol{U}_{min} \\ \Delta \boldsymbol{U}_{max} \\ \boldsymbol{0} \end{bmatrix} .
\end{aligned}$$

Notice how the two optimisation problems are solved independently from each other, thus preserving autonomy and privacy for the building owners while simultaneously allowing an aggregator to utilise the energy flexibility. In practice, there are going to be a lot of smart buildings for each aggregator that all have independent control problems. This method scales well to this case since the computational burden for the upper level remains constant—with the Flexibility Function simply representing the aggregated response from the smart buildings.

In [29], it is shown how the Flexibility Function can be used to generate a Flexibility Index for a building.

9.2 Multi-Level Control and Markets

Ultimately, the purpose of the future smart-energy system is to establish a connection between the controllers operating at local scales and high-level markets operating at large scales. Essentially, a spectrum of all relevant spatial aggregation levels (building, district, city, region, country, etc.) has to be considered. At the same time, control or market solutions must ensure that the power system is balanced at all future temporal scales. Consequently, data-intelligent solutions for operating flexible electrical energy systems have to be implemented on all spatial and temporal scales.

To address these issues, several solutions have been proposed in the literature in recent years. These major solutions are transactive energy, peer-to-peer, and control-based solutions, as described in [13].

Traditionally power systems are operated by sending bids to a market. However, in order to balance the systems on all relevant horizons, several markets are needed. Examples are day-ahead, intra-day, balancing, and regulation markets. The bids are typically static consisting of a volume and duration. However, we believe that we need a disruption related to principles for activating low-level flexibility.

Given all the bids, the so-called supply and demand curves for all the operated horizons can be found. Mathematically, these supply and demand curves are static and deterministic. Merit order dispatch is then used to optimise the cost of generation. However, if the production is from wind or solar power, then the supply curve must be stochastic, and the demand flexibility has to be described dynamically, by the introduced Flexibility Function. Consequently, we need to introduce new digitised markets that are dynamic and stochastic. And instead of using a large number of markets for different purposes (frequency, voltage, congestion, etc.) and on different horizons, we will suggest to use concepts based on the Flexibility Function and stochastic control theory, exactly as described in the previous section for the two-level case. We call this a Smart-Energy Operating System (SE-OS) [36, 44, 45].

If we zoom out in space and time, i.e., consider the load in a very large area on a horizon of days, or maybe next day, then both the dynamics and stochasticity can be eliminated, and hence, we can use conventional market principles as illustrated in

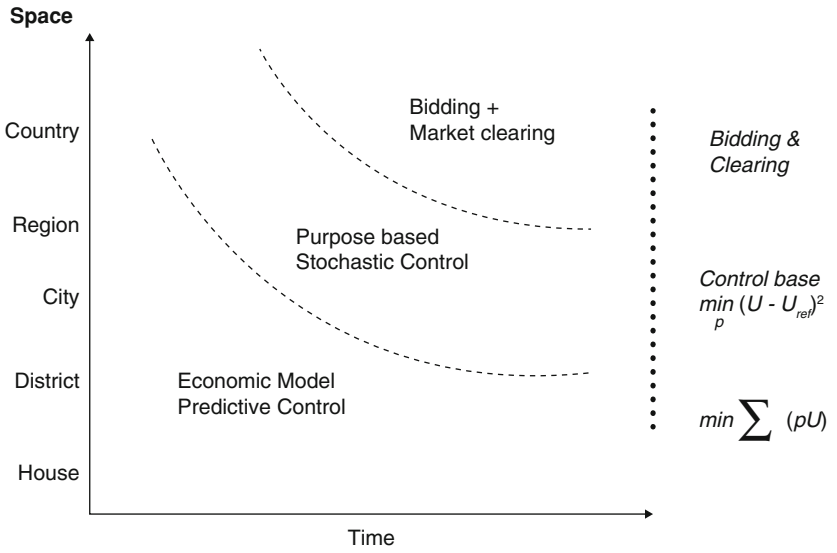

Fig. 23 Hierarchical control and markets

Fig. 23. If we zoom in on higher temporal and spatial resolutions (like for instance a house), the dynamics and stochasticity become important, and consequently, we will suggest to use the control-based methods for the flexibility as discussed in this chapter.

The total setup consists of a combination of all these options, and the best option depends on the zoom level. The conclusion is that we need new future digitised refined market principles, which operate as a hierarchy of conventional market-based bidding and clearing on the higher levels and control-based approaches on the lower level—see Fig. 23.

All these principles for forecasting, control, and optimisation are included in the so-called Smart-Energy Operating System (SE-OS), which is used to develop, implement, and test solutions (layers: data, models, optimisation, control, communication) for operating flexible electrical energy systems at all scales. See [14, 36, 44, 45] for further information.

10 Summary

In this chapter, we have presented methods for modelling relevant for the control of smart buildings. Specifically, we have introduced the grey-box modelling framework, and we have used this modelling framework to establish models for a building—as well as for some of the most important weather-related disturbances—namely cloud cover, solar radiation, net radiation, and ambient air temperature.

Most importantly, the grey-box principle bridges the gap between models based on first principles (physics) and models based on information obtained from the data (statistics).

Further, methods for model development are suggested. For parameter estimation, we suggest using the maximum likelihood method as this method allows for an integrated estimation of parameters related to the embedded description of the stochastic part.

Having models for the buildings and disturbance models as stochastic differential equations enables and promotes the use of model predictive control (MPC) as the regulation scheme for the indoor air temperature. MPC is widely described and used in the literature for building climate control problems. We introduced and formulated the mathematical optimisation problem involved with MPC and showed how to numerically compute the optimal control solution. We explained the problem of dealing with disturbances in control and showed how to incorporate them—both by simple means (using an augmented integrator) and by embedding the advanced disturbance models to supply forecasts. The last section presented simulation-based results of MPC applied to the presented smart building models using different heating strategies. The obtained results strongly suggest that the use of sophisticated disturbances models over conventional methods to supply weather forecasts can improve the building climate control.

Lastly, we have briefly explained how energy flexibility can be leveraged through price-based control, by utilising a two-level framework in which prices are generated by a controller to actuate the energy flexibility of the smart buildings. These principles are generalised to multi-level controllers for solving all types of ancillary and balancing service problems in future weather and data-driven energy systems.

Acknowledgments The authors are financed by a number of projects related to smart buildings, smart grids, and energy systems. Consequently, we would like to acknowledge support related to the following projects: *Centre for IT-Intelligent Energy Systems (IFD*[1] *CITIES)*, *Flexible Energy Denmark (IFD FED)*, *Digitally supported Smart District Heating (IFD HEAT 4.0)*, *Research Centre on Zero Emission Neighbourhoods in Smart Cities (FME*[2]*-ZEN)*, *Sustainable Energy Plus Neighbourhoods (H2020 syn.ikia)*, *Energy balancing and resilience solutions to unlock the flexibility and increase market options for distribution grid (H2020 ebalance-plus)*, *Demand response for energy cooperatives (H2020 FLEXCoop)*, *The value of end-use flexibility in the future Norwegian energy system (FlexBuild)*, *IEA EBC Annex 67: Energy Flexible Buildings*, *IEA EBC Annex 71: Building energy performance assessment based on in-situ measurements*, *IEA EBC Annex 81: Data-Driven Smart Buildings*, and *IEA EBC Annex 82: Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems*.[3] Finally, we would like to thank all the partners of this long list of projects for many very inspiring discussions.

[1] Innovation Fund Denmark.

[2] Norwegian Centres for Environment-friendly Energy Research.

[3] The participations in IEA Annex activities are supported by EUDP: a support scheme for development and testing of energy systems solutions under the Danish Energy Agency.

References

1. M. Alam et al., Applying extended Kalman filters to adaptive thermal modelling in homes. Adv. Build. Energy Res. **12**(1), 48–65 (2018). ISSN: 17562201, 17512549. https://doi.org/10.1080/17512549.2017.1325398
2. N. An, S. Hemmati, Y.-J. Cui, Assessment of the methods for determining net radiation at different time-scales of meteorological variables. J. Rock Mech. Geotech. Eng. **9**(2), 239–246 (2017)
3. B. Andersen et al., Meteorological Data for Design of Building and Installation: A Reference Year. Technical Report. 66. Technical University of Denmark. Thermal Insulation Laboratory (1977)
4. A.G. Azar et al., SmartNet: Results of Pilot B (2019)
5. P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings. Energy Build. **43**(7), 1511–1522 (2011)
6. J. Bak et al., Goodness of fit of stochastic differential equations, in *Symposium i anvendt statistik*. Conference 01 January 1999 (1999)
7. B.S. Bhangu et al., Nonlinear observers for predicting state-of-charge and state-of-health of lead-acid batteries for hybrid-electric vehicles. IEEE Trans. Vehic. Technol. **54**(3), 783–794 (2005)
8. C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, New York (2006)
9. O.M. Brastein et al., Parameter estimation for externally simulated thermal network models. Energy Build. **191**, 200–210 (2019). ISSN: 18726178, 03787788. https://doi.org/10.1016/j.enbuild.2019.03.018
10. N.L. Brok, H. Madsen, J.B. Jørgensen, Nonlinear model predictive control for stochastic differential equation systems. IFAC-PapersOnLine, in *6th IFAC Conference on Nonlinear Model Predictive Control NMPC 2018*, vol. 51(20) (2018), pp. 430–435. ISSN: 2405-8963. http://www.sciencedirect.com/science/article/pii/S2405896318327290
11. C.F.M. Coimbra, J. Kleissl, R. Marquez, Overview of solar-forecasting methods and a metric for accuracy evaluation, in *Solar Energy Forecasting and Resource Assessment*, chap. 8, ed. by J. Kleissl (Academic, Boston, 2013), pp. 171–194
12. O. Corradi et al., Controlling electricity consumption by forecasting its response to varying prices. IEEE Trans. Power Syst. **28**(1), 421–429 (2013). ISSN: 15580679, 08858950. https://doi.org/10.1109/TPWRS.2012.2197027
13. G. De Zotti et al., Ancillary services 4.0: a top-to-bottom control-based approach for solving ancillary services problems in smart grids. IEEE Access **6**, 11694–11706 (2018a). ISSN: 21693536. https://doi.org/10.1109/ACCESS.2018.2805330
14. G. De Zotti et al., Consumers' flexibility estimation at the TSO level for balancing services. IEEE Trans. Power Syst. **34**(3), 1918–1930 (2018b). ISSN: 15580679, 08858950. https://doi.org/10.1109/TPWRS.2018.2885933
15. J.S.G. Ehnberg, M.H.J. Bollen, Simulation of global solar radiation based on cloud observations. Solar Energy, in *ISES Solar World Congress 2003*, vol. 78(2) (2005), pp. 157–162. ISSN: 0038-092X. https://doi.org/10.1016/j.solener2004.08.016
16. C. Ghiaus, A. Chicinas, C. Inard, Grey-box identification of air-handling unit elements. Control Eng. Pract. **15**(4), 421–433 (2007). ISSN: 18736939, 09670661. https://doi.org/10.1016/jconengprac.2006.08.005
17. J.R. Grønborg, Characterisation and Integration of Energy Flexibility through Stochastic Modelling and Control (2019)
18. J.R. Grønborg et al., Characterizing the energy flexibility of buildings and districts. Appl. Energy **225**, 175–182 (2018)
19. J.R. Grønborg et al., Stochastic nonlinear modelling and application of price-based energy flexibility. Appl. Energy **275**, 115096 (2020). https://doi.org/10.1016/j.apenergy.2020.115096
20. R. Halvgaard et al., Economic model predictive control for building climate control in a smart grid, in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)* (2012a), pp. 1–6

21. R. Halvgaard et al., Electric vehicle charge planning using Economic Model Predictive Control, in *2012 IEEE International Electric Vehicle Conference* (2012b)
22. S. Iacus, *Simulation and Inference for Stochastic Differential Equations: With R Examples*, vol. 1 (2008). https://doi.org/10.1007/978-0-387-75839-8
23. A.H. Jazwinski, *Stochastic Processes and Filtering Theory*. Mathematics in Science and Engineering (Academic, New York, 1970)
24. S.Ø. Jensen et al., Examples of Energy Flexibility in Buildings (2019a)
25. S.Ø. Jensen, J. Parker, A.J. Marszal-Pomianowska, Principles of Energy Flexible Buildings (2019b)
26. M.J. Jimenez, H. Madsen, Models for describing the thermal characteristics of building components. Build. Environ. **43**(2), 152–162 (2008). ISSN: 1873684x, 03601323. https://doi.org/10.1016/j.buildenv.2006.10.029
27. M.J. Jimenez, H. Madsen, K.K. Andersen, "Identification of the main thermal characteristics of building components using MATLAB. Build. Environ. **43**(2), 170–180 (2008). ISSN: 1873684x, 03601323. https://doi.org/10.1016/j.buildenv.2006.10.030
28. R. Juhl, J.K. Møller, H. Madsen, *CTSMR - Continuous Time Stochastic Modelling in R* (2016). arXiv:1606.00242 [stat.CO]
29. R.G. Junker, R. Relan, H. Madsen, Designing individual penalty signals for improved energy flexibility utilisation. IFAC-PapersOnLine **52**(4), 123–128 (2019). https://doi.org/10.1016/j.ifacol.2019.08.166
30. N.R. Kristensen, H. Madsen, S.B. Jørgensen, "Parameter estimation in stochastic grey-box models. Automatica **40**(2), 225–237 (2004). ISSN: 0005-1098
31. K.B. Lindberg, Impact of Zero Energy Buildings on the Power System. Doctoral Thesis. 2017–35. NTNU (2017)
32. D.G. Luenberger, An introduction to observers. IEEE Trans. Autom. Control **16**(6), 596–602 (1971)
33. H. Lund, E. Münster, Integrated energy systems and local energy markets. Energy Policy **34**(10), 1152–1160 (2006)
34. Z. Ma, B.N. Jørgensen, J. Parker, Stakeholders' Perspectives on Energy Flexible Buildings (2019)
35. J.M. Maciejowski, *Predictive Control with Constraints* (Prentice Hall, London, 2002)
36. C. Madina et al., Technologies and protocols: the experience of the three smartNet pilots, in *TSO–DSO Interactions and Ancillary Services in Electricity Transmission and Distribution Networks*, ed. by G. Migliavacca (2019), pp. 141–183. https://doi.org/10.1007/978-3-030-29203-4
37. H. Madsen, Statistically Determined Dynamical Models for Climate Processes. PhD Thesis. Technical University of Denmark (1985)
38. H. Madsen, *Time Series Analysis* (Chapman & Hall, Boca Raton, 2008)
39. H. Madsen, CTSM-R. http://henrikmadsen.org/software/ctsm-r/. Accessed May 2020
40. H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building. Energy Build. **22**, 67–79 (1995)
41. H. Madsen, P. Thyregod, Inhomogenous Markov models for the variations in cloud cover, in *International Conference on Statistical Climatology*, vol. 3 (1986), pp. 71–76. http://imsc.pacificclimate.org/proceedings/3IMSC.pdf.
42. H. Madsen, P. Thyregod, *An Introduction to General and Generalized Linear Models* (Chapman & Hall, Boca Raton, 2011)
43. H. Madsen, H. Spliid, P. Thyregod, Markov models in discrete and continuous time for hourly observations of cloud cover. J. Clim. Appl. Climatol. **24**, 629–639 (1985)
44. H. Madsen et al., Control of electricity loads in future electric energy systems, in *Handbook of Clean Energy Systems*, ed. by A.J. Conejo, E. Dahlquist, J. Yan (Wiley, New York, 2015)
45. H. Madsen, J. Parvizi, P. Bacher, Smart-energy operating-system - a framework for implementing flexible electric energy systems in smart cities, in SUSTAIN 2016, DTU, November 2016 (2016)

46. J. Møller, H. Madsen, From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform (2010)
47. M. Morari, G. Stephanopoulos, Minimizing unobservability in inferential control schemes. Int. J. Control **31**(2), 367–377 (1980)
48. J.N. Nielsen, H. Madsen, Applying the EKF to stochastic differential equations with level effects, in *Automatica (Oxford)* (2000). ISSN: 18732836, 00051098. https://doi.org/10.1016/S0005-1098(00)00128-X
49. N. O'Connell et al., Benefits and challenges of electrical demand response: a critical review. Renew. Sustain. Energy Rev. **39**, 686–699 (2014). ISSN: 18790690, 13640321. https://doi.org/10.1016/j.rser.2014.07
50. M. O'Malley et al., Energy systems integration: defining and describing the value proposition. Contract **303**, 275–300 (2016)
51. G. Pannocchia, J. Rawlings, Disturbance models for offset-free model-predictive control. AIChE J. **49**, 426–437 (2003). https://doi.org/10.1002/aic.690490213
52. G. Pannocchia, M. Gabiccini, A. Artoni, Offset-free MPC explained: novelties, subtleties, and applications. IFAC-PapersOnLine **48**(23), 342–351 (2015)
53. M.A. Pinsky, S. Karlin, Continuous time Markov chains, in *An Introduction to Stochastic Modeling*, ed. by M.A. Pinsky, S. Karlin, 4th edn. (Academic, Boston, 2011), pp. 277–346. ISBN: 978-0-12-381416-6
54. C. Rasmussen et al., Semi-parametric modelling of sun position dependent solar gain using B-splines in grey-box models. Solar Energy **195**, 249–258 (2020). ISSN: 14711257, 0038092x. https://doi.org/10.1016/j.solener.2019.11.023
55. J. Rossiter, Model-based Predictive Control - A Practical Approach (2003)
56. A.Q. Santos, B.N. Jørgensen, Control Strategies and Algorithms for Obtaining Energy Flexibility in Buildings (2019)
57. H.A. Schløijter et al., Economic model predictive control for energy systems in smart homes, in *2019 IEEE Conference on Control Technology and Applications (CCTA)* (2019)
58. F.C. Schweppe et al., Homeostatic utility control. IEEE Trans. Power Appar. Syst. **99**, 1151–1163 (1980)
59. M.A. Shamim et al., An improved technique for global solar radiation estimation using numerical weather prediction. J. Atmos. Solar-Terrestrial Phys. **129**, 13–22 (2015). ISSN: 1364-6826. http://www.sciencedirect.com/science/article/pii/S1364682615000590
60. V. Sharma et al., Short term solar irradiance forecasting using a mixed wavelet neural network. Renew. Energy **90**, 481–492 (2016). ISSN: 0960-1481. http://www.sciencedirect.com/science/article/pii/S0960148116300209
61. C.A. Thilker, Optimization for Smart Energy Systems". Can be found at https://www.findit.dtu.dk. MA Thesis. Department of Applied Mathematics and Computer Science: Technical University of Denmark (2020)
62. T.O. Ting et al., State-space battery modeling for smart battery management system, in Lecture Notes Notes in Engineering and Computer Science, vol. 2210 (2014)
63. O. Tremblay, L.-A. Dessaint, A.-I. Dekkiche, A generic battery model for the dynamic simulation of hybrid electric vehicles, in *2007 IEEE Vehicle Power and Propulsion Conference* (2007), pp. 284–289
64. L. Wasserman, *All of Nonparametric Statistics*. Springer Texts in Statistics (Springer, Berlin, Heidelberg, 2006)
65. D. Yang, P. Jirutitijaroen, W.M. Walsh, Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy **86**(12), 3531–3543. Solar Resources (2012). ISSN: 0038-092X. http://www.sciencedirect.com/science/article/pii/S0038092X12003039
66. Y. Zhang, V.I. Hanby, Short-term prediction of weather parameters using online weather forecasts, in *Proceedings: Building Simulation 2007* (2007)