

Article

Bayesian Calibration with Augmented Stochastic State-Space Models of District-Heated Multifamily Buildings

Lukas Lundström ^{1,*}  and Jan Akander ² 

¹ School of Business, Society and Engineering, Mälardalen University, 72123 Västerås, Sweden

² Department of Building Engineering, Energy Systems and Sustainability Science, University of Gävle, 80176 Gävle, Sweden; jan.akander@hig.se

* Correspondence: lukas.lundstrom@mdh.se

Received: 18 November 2019; Accepted: 17 December 2019; Published: 22 December 2019



Abstract: Reliable energy models are needed to determine building energy performance. Relatively detailed energy models can be auto-generated based on 3D shape representations of existing buildings. However, parameters describing thermal performance of the building fabric, the technical systems, and occupant behavior are usually not readily available. Calibration with on-site measurements is needed to obtain reliable energy models that can offer insight into buildings' actual energy performances. Here, we present an energy model that is suitable for district-heated multifamily buildings, based on a 14-node thermal network implementation of the ISO 52016-1:2017 standard. To better account for modeling approximations and noisy inputs, the model is converted to a stochastic state-space model and augmented with four additional disturbance state variables. Uncertainty models are developed for the inputs solar heat gains, internal heat gains, and domestic hot water use. An iterated extended Kalman filtering algorithm is employed to enable nonlinear state estimation. A Bayesian calibration procedure is employed to enable assessment of parameter uncertainty and incorporation of regulating prior knowledge. A case study is presented to evaluate the performance of the developed framework: parameter estimation with both dynamic Hamiltonian Monte Carlo sampling and penalized maximum likelihood estimation, the behavior of the filtering algorithm, the impact of different commonly occurring data sources for domestic hot water use, and the impact of indoor air temperature readings.

Keywords: building energy performance; energy models; Bayesian calibration; augmented stochastic state-space modeling; Iterated Extended Kalman Filtering; uncertainty

1. Introduction

Reliable energy models are needed to determine flexibility in energy demand of buildings and to estimate energy savings resulting from energy conservation measures. Senave et al. [1] defined three key elements for which thorough insight is required to assess energy performance of existing buildings: (i) the thermal performance of the building fabric, (ii) the efficiency of the technical systems, and (iii) the behavior of the users. Energy modeling of urban buildings seeks to expand conventional whole building energy modeling to the scale of neighborhoods, cities, or even entire building stocks [2]. Creating reliable models at these scales is difficult, as they require large amounts of detailed input data that are lacking in digitized form for most existing buildings. However, development within aerial imaging or LiDAR (light detection and ranging) [3] makes 3D shape representation of existing buildings possible. Such data sources enable relatively detailed energy models to be constructed, especially geometrical aspects of the building and its surroundings [4]. However, parameters describing thermal

performance of the building fabric, the technical systems, and occupant behavior are usually not readily accessible in a digitized format. Calibration with on-site measurements is needed to obtain reliable estimates of these aspects [1,5]. Today, metering data that provides the basis for utility billing are often available in decent resolution and sampling rates [6]. These metering data coming from on-site measurements, such as supplied energy and domestic cold water, indirectly contain or mirror the buildings' actual energy performance.

In a review on methodologies and recent advancements in the calibration of building energy models, Fabrizio and Monetti [5] concluded that automated models are often simplified in order to reduce computational time, and due to that, more complex models are difficult to handle in the calibration process. They also highlighted the importance of assessment of occupancy behavior, "since the occupancy related to the building usage is one of the main sources of uncertainty in the building simulation models" [5]. A method of introducing and handling parameter uncertainty in energy models is with a Bayesian approach. Several studies aiming at understanding and showing the potentials of this approach are listed in Chong and Menberg [7], where energy models performance, based on Bayesian calibration, are tested on and compared with deterministic model results. Based on these experiences, guidelines for Bayesian calibration of building energy models are presented. Accordingly, a test of the guidelines in a case study [7] also suggests that Bayesian calibration has limitations depending on the information content within datasets in terms of how many uncertain parameters can be introduced in the model and the quality of prior knowledge about these.

Previous energy model calibration research can be grouped into those working with relatively simple state-space models with 2 to 4 state variables (for example Bacher and Madsen [8], Coffman and Barooah [9], Raillon and Ghiaus [10], Rouchier et al. [11]) and those working with calibration of detailed engineering-based models (for example Chong and Menberg [7], Tian et al. [12], Bandera and Ruiz [13]). The benefits of simple state-space models are that little prior information about the building is needed and that these can be efficiently computed and used directly in Monte Carlo Markov Chain (MCMC)-based inference. However, such models do not reveal much details about the energy performance. Engineering-based simulation tools such as EnergyPlus [14] are based on extensive building physics research [15]. When such models are calibrated, they can provide detailed knowledge about the energy performance of the building and its components. Nonetheless, these engineering-based models require detailed input and are too computationally expensive to be directly used in MCMC-based inference. The typical Bayesian approach is to approximate the detailed model with a simpler surrogate or emulator model [7].

Weather-dependent space heating use can be estimated reasonably well deterministically [4]. However, even the most sophisticated deterministic models contain approximation errors and, for example, Bacher and Madsen [8] and Rouchier et al. [11] suggest using stochastic state-space modeling to account for such approximation errors. Solar heat gains, internal heat gains, and heat use for domestic hot water (DHW) are inputs that contain high levels of time-varying uncertainty and are challenging for both stochastic and deterministic modeling approaches. Internal heat gains and DHW use are inherently uncertain due to dependency on occupant behavior [16], whereas solar heat gains are uncertain both due to epistemic uncertainty of occupant behavior (i.e., shading operation) and aleatory uncertainty due to the complexity in modeling solar irradiance and shading from the surroundings [4,17,18]. Augmented stochastic state-space models have been suggested as a suitable approach to deal with such highly nondeterministic inputs: Kim and Park [19] used a detailed augmented state-space model (consisting of 15 states) to estimate time-varying process disturbances attributed to uncertainty in internal heat sources and airflow, and Coffman and Barooah [9] augmented a simplified two state-space model with a third disturbance state to account for unmeasured disturbances attributed to occupant behavior.

In this paper, we focus on Swedish district-heated multifamily buildings. Figure 1 exemplifies the heat balance of a district heating substation serving multiple multifamily buildings. According to Kensby et al. [20], there is untapped potential in using thermal mass as short-term thermal energy

storage to shift heat demand from times when the district heat is produced at high cost and with negative environmental impact to more favorable hours. In Sweden, energy and water bills must be based on actual consumption [21]. Consequently, most Swedish district heating operators have automatic meter-reading systems installed, which gather hourly or sub-hourly readings, storing them in centralized databases. Most Swedish cities have extensive district heating networks which provide heat to around 90% of all multifamily buildings [22]. Such on-site measurements make calibration of energy models on large scales possible.

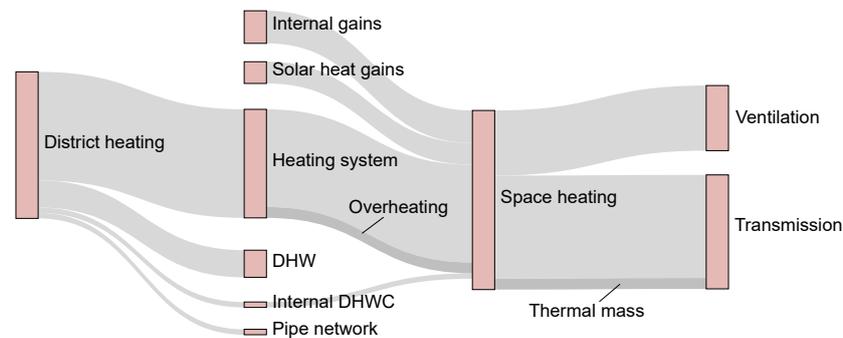


Figure 1. Typical heat balance of a district heating-connected Swedish multifamily building: The height of the gray areas indicates relative share of total heat flow over the heating season. Darker gray areas show 1. overheating due to imperfect control and 2. net contribution from charging and discharging the thermal mass.

Kim and Park [19] and Lundström et al. [4] have proposed detailed state-space models that are able to produce simulation results similar to engineering-based simulation tools and that have potential to reveal detailed knowledge about the energy performance of buildings and their systems/components. To the authors' knowledge, there are no previous published works that have conducted calibration with such detailed space-models with data from actual buildings. With increased availability of digitized information, automated collection, and integration of data into an energy model, representing one or multiple similar buildings is feasible today [4]. A deterministic model in the form of a thermal network [4] (a 14-node implementation based on that proposed by the ISO 52016-1:2017 [23] standard) is transformed into a stochastic state-space model and further augmented with four additional disturbance variables. Furthermore, we present a Bayesian calibration procedure to assess parameter uncertainty, to incorporate prior knowledge, and to test parameter estimation with both dynamic Hamiltonian Monte Carlo (HMC) sampling and penalized Maximum Likelihood Estimation (MLE) optimization by utilizing the probabilistic programming language Stan [24]. The proposed model framework was tested in a case study, and lessons learned are reported and discussed.

The objective is a framework for assessing actual energy performance: A framework that works on most Swedish multifamily buildings, that is sufficiently detailed to result in usable insights, and that can make use of the varying data sources that are available today while still resulting in decent results in situations of less data availability. The main original contributions of our work are an iterated extended Kalman filter to handle nonlinearity of DHW use; an augmentation that implements a probabilistic approach for handling uncertain input data; a direct implementation of the model in the Stan language, thus enabling computationally efficient (including automatic differentiation) inference with both HMC and penalized MLE; and a case study based on real buildings and original data.

The work is presented as follows: Section 2 presents the proposed Bayesian calibration procedure with an augmented stochastic state-space model. In Section 3, we briefly describe our previous modeling work [4] and extend the deterministic model to include heat losses in the piping, manual venting, DHW heat use, and internal heat gains and the uncertainty modeling of these four uncertain inputs. Finally, Section 4 presents a case study to demonstrate the performance of the developed model framework: parameter estimation with both dynamic HMC sampling and penalized MLE

optimization, the behavior of the filtering algorithm, the impact of different data sources for DHW use, and the impact of uncertainty in indoor temperature sensor readings.

2. Bayesian Calibration with Augmented Stochastic State-Space Models

Section 2.1 gives an overview of the proposed Bayesian calibration procedure with an augmented stochastic state-space model. Section 2.2 presents the discretization and augmentation procedures. Section 2.3 describes the nonlinear state estimation using an Iterated Extended Kalman Filtering (IEKF) algorithm.

2.1. Overview

Let θ denote the set of unknown/uncertain parameters. Further, let $\theta \subset \Theta$, where Θ contains all of the system-describing parameters, including parameters considered as known/certain. In Bayesian inference, the unknown parameters θ are given a prior distribution $\pi(\theta)$, which represents the prior belief before calibrating with data. Information available in the data is summarized in the likelihood function $\pi(\mathbf{y}_{1:N}|\theta)$, which is the conditional distribution over the output $\mathbf{y}_{1:N}$ given parameters θ . The prior belief and the evidence provided by the data are combined in the Bayes' theorem to compute the posterior distribution:

$$\pi(\theta|\mathbf{y}_{1:N}) \propto \pi(\mathbf{y}_{1:N}|\theta)\pi(\theta) \quad (1)$$

A penalized maximum likelihood (MLE) estimate is obtained by optimizing the posterior density, and a full Bayesian inference is obtained by sampling from the posterior density. In this paper, we employ the dynamic Hamiltonian Monte Carlo (HMC) sampling algorithm implemented in the statistical modeling language Stan [24]. Stan also handles the calculation of the partial derivatives of the likelihood function through its automatic differentiation capabilities. In this work, the likelihood is computed with an augmented stochastic state-space model using an IEKF algorithm for the nonlinear state estimation. Figure 2 shows an overview of the proposed procedure.

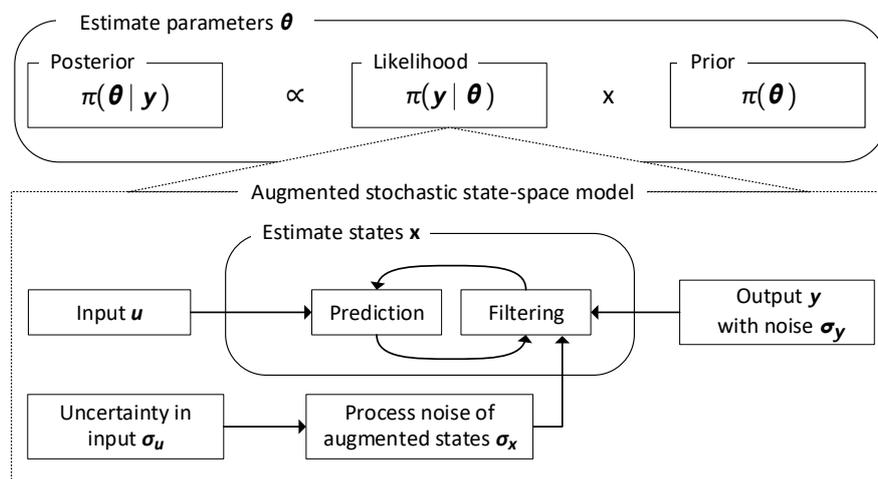


Figure 2. Bayesian calibration using an augmented stochastic state-space model to compute the likelihood.

Figure 3 visualizes the proposed augmented stochastic state-space model in the form of an energy-flow diagram. The cyan-filled circles show the four inputs that, through state augmentation, are converted into uncertain inputs, i.e., inputs declared with a location value μ_k and accompanying an uncertainty scale σ_k for each time step k . Most of the energy flowing into the conditioned space (the middle section of the figure) are treated as uncertain, whereas the heat losses (energy leaving the conditioned space) are treated as deterministic. The heating system energy flow is also treated as

uncertain—the heat use is estimated at each time step based on deterministic inputs and the output internal air temperature θ_{int} from the previous time step and then forwarded as an uncertain input to the correction stage.

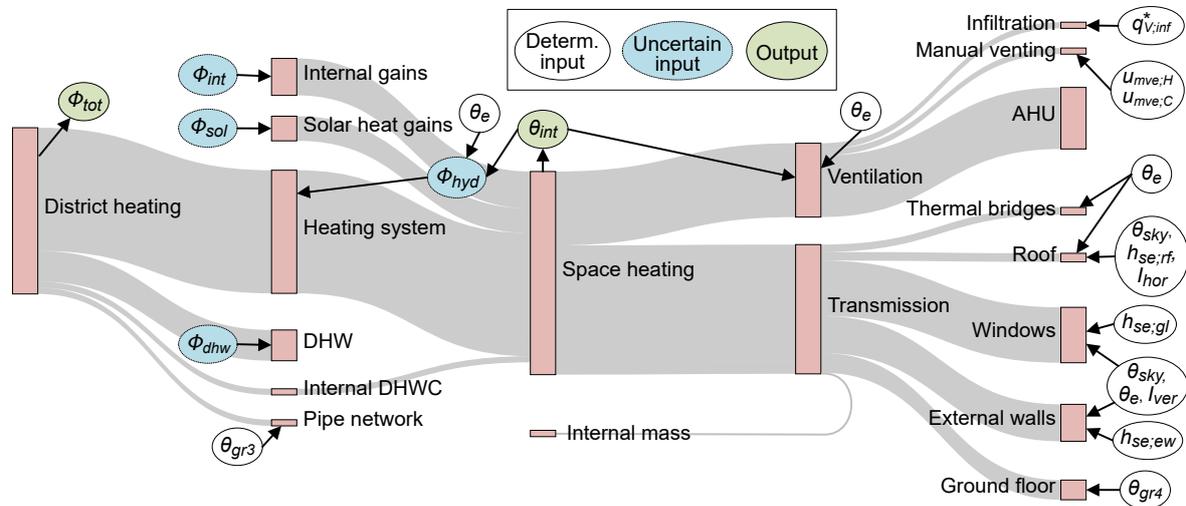


Figure 3. An energy flow diagram showing where the deterministic inputs, uncertain inputs, and uncertain outputs enters the stochastic system.

The proposed filtering algorithm is summarized as follows: at each time step k , prediction/time update and filtering/measurement update stages are carried out. During the prediction/time update stage, a deterministic prediction of the outputs is calculated based on the states of the previous time step $k - 1$, the deterministic inputs, and the most probable value (the mode) of the uncertain inputs. Next, during the filtering/measurement update stage, any discrepancy between predicted and measured outputs results in a correction of the four augmented state variables and a selection of temperature-state variables of the thermal network (the five internal surface nodes and the internal air node). The magnitude of correction depends on the ratio between measurement and process noise; a small ratio will render a large correction and vice-versa. The relative share of the correction received by each state variable depends on the ratios of process noises; more uncertain state variables receive a relatively larger share of the correction. For example, the input for solar heat gains during night time is zero with no uncertainty ($\mathcal{N}(0, 0)$), and therefore, any correction is directed to state variables with encoded uncertainty during this time step.

2.2. Augmented State-Space Modeling

In Section 2.2.1, the thermal network of the ISO14N model (presented in Section 3) is converted from a continuous-time system into a discrete-time system. In Section 2.2.2, the system is further augmented with four new state variables, which will later be used to enable time-varying uncertainty for selected inputs.

2.2.1. Discretization

The ISO14N thermal network is expressed as a deterministic continuous-time state-space model:

$$dx_t^\theta = A_t^\theta x_t^\theta dt + B_t^\theta u_t^\theta dt \tag{2}$$

where $x^\theta \in \mathbb{R}^{14 \times 1}$, $A^\theta \in \mathbb{R}^{14 \times 14}$, $B^\theta \in \mathbb{R}^{14 \times 10}$, and $u^\theta \in \mathbb{R}^{10 \times 1}$ are the states (the temperature nodes), continuous-time transition matrix, continuous-time input coefficient matrix, and input data vector of the ISO14N thermal model, respectively, and t denotes time. The states of the thermal network are expressed in degrees centigrade, and the thermal network part of the system is therefore denoted with the superscript θ . Let Δt denote the sampling interval, and assume that the inputs u are constant

during this interval. A continuous-time system is then translated into a discrete system using the following set of equations [25]:

$$F = \exp(A\Delta t), \quad G = (F - I)A^{-1}B \quad (3)$$

where I is the identity matrix. Applying Equation (3) results in the following discrete-time system (where k is the discrete-time index):

$$\mathbf{x}_k^\theta = \mathbf{F}_k^\theta \mathbf{x}_{k-1}^\theta + \mathbf{G}_k^\theta \mathbf{u}_k^\theta \quad (4)$$

2.2.2. Augmentation

To enable encoding of time-varying uncertainty in selected inputs, the system is augmented with four additional state variables. The augmented states are expressed as thermal power ($\text{W}/\text{m}_\text{fl}^2$) and denoted with the superscript ϕ :

$$\begin{bmatrix} \mathbf{x}_k^\theta \\ \mathbf{x}_k^\phi \end{bmatrix} = \begin{bmatrix} \mathbf{F}_k^\theta & \mathbf{G}^\phi & 0 \\ \mathbf{0} & \mathbf{I} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{k-1} \\ \boldsymbol{\phi}_{k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_k^\theta & 0 \\ \mathbf{0} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_k^\theta \\ \ln(\phi_{dhw;k} + 1) \end{bmatrix} \quad (5)$$

where $\mathbf{x}^\phi \in \mathbb{R}^{4 \times 1}$ contains four augmented state variables describing solar heat gains, internal heat gains, heat from the hydronic heating system, and DHW heat use and where $\mathbf{G}^\phi \in \mathbb{R}^{14 \times 3}$ is a subset of \mathbf{G}^θ containing the three columns that model how the inputs $\mathbf{u}_{\phi;sol}^\theta$, $\mathbf{u}_{\phi;int}^\theta$, and $\mathbf{u}_{\phi;hyd}^\theta$ enter the thermal network. The fourth augmented state variable $\mathbf{x}_{\phi;dhw;ln}^\phi$ models the DHW heat use with a logarithmic scale (see Section 3.5 for rationale). No interaction between DHW use and the the space heating (the thermal network) is assumed: the input coefficient matrix simply maintains an identity link to the input $\ln(\phi_{dhw;k} + 1)$. The system in Equation (5) can now be written in the following simplified form:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{G}_k \mathbf{u}_k \quad (6)$$

where $\mathbf{x} \in \mathbb{R}^{18 \times 1}$, $\mathbf{F} \in \mathbb{R}^{18 \times 18}$, $\mathbf{G} \in \mathbb{R}^{18 \times 10}$, and $\mathbf{u} \in \mathbb{R}^{10 \times 1}$ are the state vector, transition matrix, input coefficient matrix, and input data vector, respectively, of the augmented system.

2.3. Nonlinear State Estimation of a Stochastic System

Kalman and Bayesian filters blend noisy and limited knowledge of how a system behaves with noisy and limited sensor readings to produce an optimal estimate of the state of the system at each time step. In Section 2.3.1, the system is formulated as a nonlinear stochastic system and the algorithm for the proposed filtering method is presented. Section 2.3.2 details the process and measurement noise of the filter, whereby the process noise of the augmented state variables are used to encode the time-varying uncertainty for the selected inputs. A state-space formulation of a discrete-time nonlinear stochastic system is given by the following pair of equations:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{w}_k \quad (7a)$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k \quad (7b)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the state vector of the system, $\mathbf{y}_k \in \mathbb{R}^{n_y}$ is the measurement vector, $f: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$ is the state transition function, $h: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$ is the measurement function, $\mathbf{w}_k \in \mathbb{R}^{n_x}$ is the process noise vector, and $\mathbf{v}_k \in \mathbb{R}^{n_y}$ is the measurement noise. \mathbf{w}_k and \mathbf{v}_k are zero-mean Gaussian sequences with covariance $\mathbf{Q}_k \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{R}_k \in \mathbb{R}^{n_y \times n_y}$, respectively. Due to the nonlinear augmented state variables introduced in Equation (5), the system under study is no longer linear and a nonlinear filtering method is required.

2.3.1. Iterated Extended Kalman Filter

An Extended Kalman Filter (EKF) handles nonlinearity by linearizing the system at the point of the current estimate [26]. However, due to the often high and non-normally distributed uncertainty in the DHW use (see Section 3.5 for detailed investigation), regular EKF does not always suffice. Therefore, Iterated Extended Kalman Filtering (IEKF) is used. IEKF improves the reference trajectory and, consequently, the estimate by local iterations of the EKF measurement update step [27]. The IEKF procedure is outlined in Algorithm 1. The log-likelihood is obtained by

$$-\ln \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^N \ln |S_k(\boldsymbol{\theta})| + \frac{1}{2} \sum_{k=1}^N \boldsymbol{\varepsilon}_k(\boldsymbol{\theta})^T S_k(\boldsymbol{\theta})^{-1} \boldsymbol{\varepsilon}_k(\boldsymbol{\theta}) \quad (8)$$

where $\boldsymbol{\varepsilon}$ is the Kalman innovation and S is the Kalman innovation covariance. The Kalman filter applies Bayes' rule at each time step k , where the updated state $p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_{k|k}, \mathbf{P}_{k|k})$ is a posterior distribution, obtained by compromising between a prior output $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k|k-1}, \mathbf{P}_{k|k-1})$ and the evidence from the measurement \mathbf{y}_k . Their relative weight is expressed by the Kalman gain \mathbf{K}_k , which measures the relative confidence put on both the prediction and the measurement.

Algorithm 1: Iterated Extended Kalman Filter (IEKF).

```

1 Initialize  $\hat{\mathbf{x}}_0$  and  $\mathbf{P}_0$ ;
2 for  $k$  in 1:n do
    // prediction, time update
3    $\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k, \boldsymbol{\Theta}_k)$ ; // predict state mean values
4    $\mathbf{F} = \partial f(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u} = 0, \boldsymbol{\Theta}_k) / \partial \hat{\mathbf{x}}$ ; // linearize state transition function  $f$ 
5    $\mathbf{P}_{k|k-1} = \mathbf{F} \mathbf{P}_{k-1} \mathbf{F}^T + \mathbf{Q}_k$ ; // predict state covariance matrix
    // filtering, measurement update
6    $\hat{\mathbf{x}}^1 = \hat{\mathbf{x}}_{k|k-1}$ ; // initialization
7   for  $i$  in 0: $N_{iter}$  do
8      $\mathbf{H}^i = \partial h(\hat{\mathbf{x}}^i, \mathbf{u} = 0) / \partial \hat{\mathbf{x}}$ ; // linearize measurement function  $h$ 
9     if  $i > 0$  &  $\|\mathbf{H}^i - \mathbf{H}^{i-1}\| < \varepsilon$  then stop; // stop if negligible change in  $\mathbf{H}$ 
10     $\hat{\mathbf{y}}_k = h(\hat{\mathbf{x}}^i, \mathbf{u}_k)$ ; // predict output
11     $\boldsymbol{\varepsilon}_k = \mathbf{y}_k - h(\hat{\mathbf{x}}^i, \mathbf{u}_k)$ ; // Kalman innovation
12     $\mathbf{S}_k = (\mathbf{H}^i)^T (\mathbf{H}^i \mathbf{P}_{k|k-1} (\mathbf{H}^i)^T + \mathbf{R}_k)$ ; // Kalman innovation covariance
13     $\mathbf{K} = \mathbf{P}_{k|k-1} \mathbf{S}_k^{-1}$ ; // Kalman gain
14     $\hat{\mathbf{x}}_k^i = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}(\boldsymbol{\varepsilon}_k - \mathbf{H}^i(\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}_k^i))$ ; // update states
15  end
16   $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}^i$ ; // final state update
17   $\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K} \mathbf{H}^i) \mathbf{P}_{k|k-1}$ ; // update state covariance
18 end
19  $\hat{\mathbf{y}}_k = h(\hat{\mathbf{x}}^i, \mathbf{u}_k)$ ; // final filtered output

```

The state transition function f of the system under study is given in Algorithm 2, where line 2 adds a prediction of the heat use ($\hat{\phi}_{hyd}$) to the input vector \mathbf{u} and line 4 equals the linear state transition of Equation (6). For the system under study, the linearization step of the transition function (line 4 in Algorithm 1) is not needed, as the Jacobian matrix of $f(\cdot)$ equals the linear transition matrix \mathbf{F} of Equation (6).

Algorithm 2: The state transition function

```

1 Function  $f(x, u, \Theta)$ 
2    $u_{\phi;hyd} = f_{hyd}(x, u, \Theta);$  // calculate space heating use as a nonlinear function
3    $\Theta \rightarrow F, G;$  // map parameter set  $\Theta$  to matrices  $F$  and  $G$ 
4    $\hat{x} = Fx + Gu;$  // linear state transition
5   return  $\hat{x};$  // return the predicted states

```

The measurement function h is given in Algorithm 3, where $x_{\phi;sol}$, $x_{\phi;int}$, and $x_{\phi;hyd}$ are the augmented state variables containing the correction for the corresponding inputs solar heat gains, internal heat gains, and hydronic heating system. $x_{\phi;dhw;ln}$ is heat use for the DHW in logarithmic scale and u contains the inputs, including the predicted heat use of the hydronic heating system. \hat{y}_1 is the predicted total heat use, corresponding to the total heat use metered by the DH-meter y_1 . The augmented states $x_{\phi;hyd}$ and $x_{\phi;dhw;ln}$ are added as these are directly part of the measured total heat use y_1 , while $x_{\phi;sol}$ and $x_{\phi;int}$ are subtracted as these correspond to passive heat input that is not measured by the y_1 (i.e., 1 kWh of passive heat gain is assumed to offset 1 kWh of actively provided heat input).

Algorithm 3: The measurement function.

```

1 Function  $h(x, u)$ 
2    $\hat{y}_1 = u_{\phi;hyd} + u_{\phi;dhw;c} + u_{\phi;p} - x_{\phi;sol} - x_{\phi;int} + x_{\phi;hyd} + \exp(x_{\phi;dhw;ln}) - 1;$ 
3    $\hat{y}_2 = x_{\theta;int};$ 
4   return  $\hat{y};$ 

```

Local linearization at time step k of the measurement function h is given below in Equation (9), where $H \in \mathbb{R}^{2 \times 18}$ is the linearized measurement matrix and u_{trv} denotes average control signal from the thermostatic radiator valves.

$$H = \frac{\partial h(x_k, u)}{\partial x} = \begin{bmatrix} (0)_{\times 13} & 0 & -1 & -1 & 1 & \exp(x_{\phi;dhw;ln}) \\ (0)_{\times 13} & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

2.3.2. Process and Measurement Noise

The time-varying process noise covariance matrix is given as

$$Q_k = \text{diag}([0 \ 0 \ \sigma_{\theta} \ 0 \ 0 \ \sigma_{\theta} \ 0 \ 0 \ 0 \ \sigma_{\theta} \ 0 \ 0 \ \sigma_{\theta} \ \sigma_{\theta} \ \sigma_{\phi;sol;k} \ \sigma_{\phi;int;k} \ \sigma_{\phi;hyd;k} \ \sigma_{\phi;dhw;ln;k}]^2) \quad (10)$$

where σ_{θ} is the time-invariant uncertainty of the interior surfaces and internal air states of the thermal network and where $\sigma_{\phi;sol;k}$, $\sigma_{\phi;int;k}$, $\sigma_{\phi;hyd;k}$, and $\sigma_{\phi;dhw;ln;k}$ are the time-varying uncertainty of the augmented state variables (Section 3 details the modeling of the uncertainty). Setting σ_{θ} to 0.05 was found heuristically to give reasonable results, in the sense that it resulted in the measurement noise being on the same scale as the process noise of the augmented states. If heat input to the conditioned space is overestimated at time step k , it is clear that the internal air temperature node will immediately increase, causing the temperature nodes of the interior surface nodes to increase. However, it takes time for the overestimated heat input to propagate to the temperature nodes further away from the internal air node. Therefore, only the interior surface nodes (except the low-mass windows) and the internal air nodes of the thermal network are given process noise, which results in a sparse matrix, enabling less computation (the Kalman filtering can be conducted with 9×9 matrices instead of 18×18 matrices).

The discrete-time measurement noise covariance matrix is given as

$$R_k = \text{diag}([\sigma_{\phi;tot;k} \ \sigma_{\theta;int;k}]^2 / \Delta t) \quad (11)$$

where $\phi_{tot;\sigma;k}$ is the standard deviation of the heat meter readings and $\theta_{int;\sigma;k}$ is the standard deviation of the internal air temperature measurements. When the total heat use measurements come from a billing heat meter, the time-varying uncertainty can be determined by the method developed by Lundström and Dahlquist [6].

3. Modelling District-Heated Multifamily Buildings

Figure 4 shows a schematic example of a district heating substation serving multiple buildings with both space heating and heat for DHW use (including DHW circulation and local pipe network). A complete energy model for such building(s) has to model everything after the meter (new buildings tend to have one substation per building). In our previous work [4], we developed a deterministic energy model consisting of a thermal network, a hydronic radiator heating system, and simulation and data handling procedures. The thermal network consists of 14 nodes and is a lumped and simplified version of the ISO 52016-1:2017 [23] standard; see Figure 5. The data handling consists of preprocessing procedures to acquire and make use of satellite-based solar radiation data, meteorological reanalysis data (air temperature, ground temperature, wind, albedo, and thermal radiation) and preprocessing procedures of boundary conditions to account for the effect of shading objects, window blinds, wind- and stack-driven air leakage, and variable exterior surface heat transfer coefficients. In this work, we extend our previous modeling work to also include network pipe heat losses, manual venting, DHW heat use, and internal heat gains and modeling of the uncertainty of the four uncertain inputs. The procedures are detailed in the sections below.

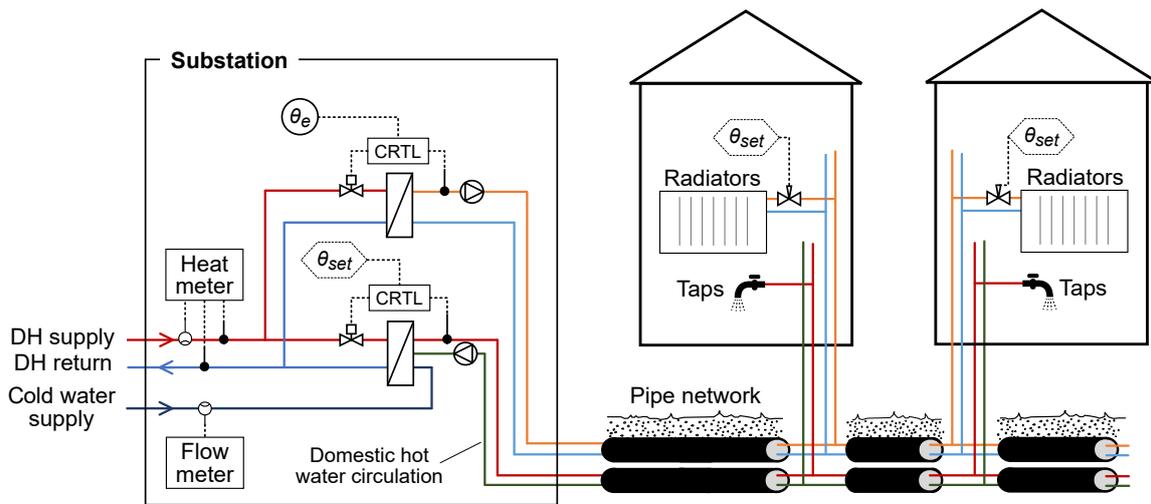


Figure 4. District heating substation serving multiple buildings.

3.1. Network Pipe Heat Losses

This section details the modeling of two additional major sources of heat losses: the local piping network and DHW circulation (DHWc). In the case of a district heating substation serving several buildings, a local piping network is needed to connect all buildings with the substation. The heat losses from this local piping network can comprise a substantial part of the heat use that is metered at the substation [28]. Here, we model a four-pipe configuration with separate pairs of pipes for DHW and space heating:

$$\phi_{p;k} = \Psi_p \cdot L_p / A_{fl} \cdot ((\theta_{dhw;set} - \theta_{gr3;k}) + u_{trv;k} \cdot (\theta_{hyd;sup;k} + \theta_{hyd;ret;k} - 2 \cdot \theta_{gr3;k})) \quad (12)$$

where Ψ_p (W/(m · K)) is the linear heat loss coefficient for the ground-buried pipes, L_p (m) is the length of the piping network, A_{fl} is the floor area for normalization, $\theta_{dhw;set}$ is the set-point temperature for DHW (in Sweden, set above 55 °C to avoid legionella growth and below 60 °C to avoid scolding), u_{trv} is

the control signal from the thermostatic radiator valves (modeling that the flow is turned off when there is no heat demand), $\theta_{hyd;sup}$ and $\theta_{hyd;ret}$ are the supply and return temperatures of the hydronic heating system, respectively, and θ_{gr3} represents ground temperature of soil layer at 0.28–1.0 m depth.

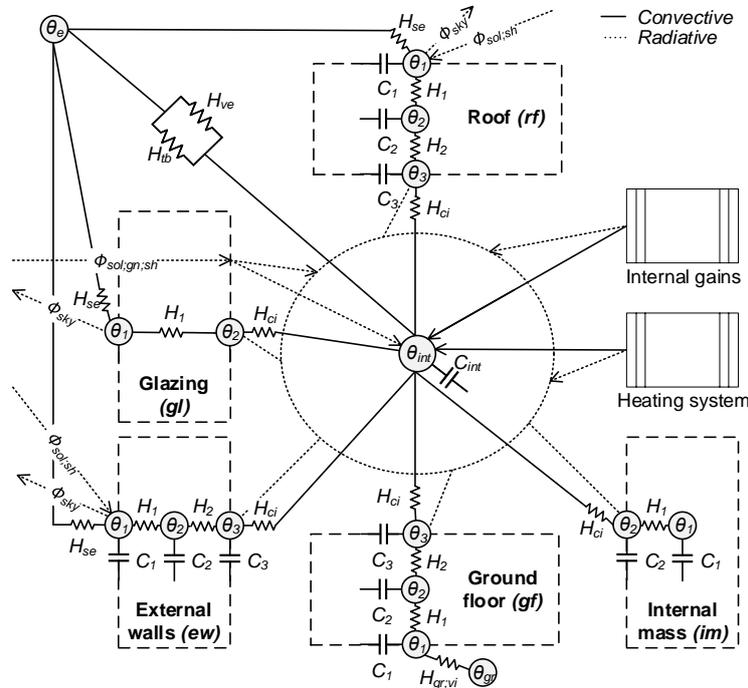


Figure 5. The thermal network used to represent the conditioned space.

The purpose of DHWC is to ensure that waiting time at the tap is short. Heat losses due to DHWC can be substantial and can depend on factors such as design, craftsmanship, and insulation level of the DHWC system. For energy calculations, a standard value of $4 \text{ kWh}/(\text{m}_{fl}^2 \cdot \text{year})$ is often used. However, a Swedish study [28] showed that heat losses varied between 2 and $28 \text{ kWh}/(\text{m}_{fl}^2 \cdot \text{year})$ in 12 studied multifamily buildings. The DHWC heat losses can be divided into two parts: internal heat losses occurring within the building ($\phi_{dhwc;i}$) and external heat losses occurring in the eventual post-substation piping network ($\phi_{dhwc;e}$). Heat losses within the building contribute to the heating of the building and are assumed to be constant over the year. In this paper, external DHWC heat losses are modeled as part of the heat losses from the piping network as described in Equation (12). In the rest of the paper, the symbol ϕ_{dhwc} refers only to the internal heat losses.

3.2. Ventilation

Heat transfer due to ventilation is modeled as a sum of controlled ventilation in the air handling unit, infiltration, and occupant-induced manual venting:

$$\phi_{ve;k} = \kappa\rho_a \cdot (\hat{\theta}_{int;k} - \theta_{e;k}) \cdot (q_{V;ahu} \cdot (1 - \eta_{ahu}) + q_{V;inf;k} + q_{V;mve;k}) \tag{13}$$

where $\kappa\rho_a$ is the heat capacity of air per volume $1.21 \text{ W} \cdot \text{s}/(\text{l} \cdot \text{K})$, $\hat{\theta}_{int;k}$ is the estimated internal air temperature (from the previous time step in the first iteration), $\theta_{e;k}$ is the external air temperature, $q_{V;ahu}$ is the specific air flow rate ($1/(\text{s} \cdot \text{m}_{fl}^2)$) for the air handling unit, η_{ahu} is the temperature transfer efficiency of the heat recovery unit, $q_{V;inf;k}$ is the specific air flow rate due to infiltration, and $q_{V;mve;k}$ is the specific air flow rate due occupant-induced manual venting. The infiltration calculation accounts for wind and stack-driven pressure difference and is described in Lundström et al. [4] (adapted from the ASHRAE enhanced infiltration model [29]).

For Swedish multifamily buildings, occupant-induced manual venting mainly arises when occupants open windows or balcony doors or force ventilate with the kitchen hood or bathroom exhaust air outlets. Manual venting contributes with a relatively small share to the total air exchange and is typically accounted for by adding between 2.3 to 4.6 l/s per apartment in continuous air flow rate [30]. Based on field surveys, Rijal et al. [31] proposed the use of a logistic function for describing the relationship between the external and internal globe temperature and the fraction of open windows in office buildings. Hedegaard et al. [32] adopted the model to describe the effective amount of natural ventilation as a fraction of a design flow for Danish single-family buildings. In this study, we further adopted logistic function approach and used the following equation to model occupant induced manual venting:

$$q_{V;mve;k} = q_{V;mve;H} \cdot u_{mve;H;k} + q_{V;mve;C} \cdot u_{mve;C;k} \quad (14)$$

where $q_{V;mve;H}$ and $q_{V;mve;C}$ are the nominal manual venting rates during heating season and cooling season respectively, and $u_{mve;H;k}$ and $u_{mve;C;k}$ are the control signals describing how much of the available nominal venting rates are utilized at each time step k . The control signal for manual venting during heating season is modeled with the following logistic function:

$$u_{mve;H;k} = 0.3 + \frac{0.7}{1 + \exp(-0.25 \cdot (\theta_{e;mwa;k} - 1))} \quad (15)$$

where $\theta_{e;mwa;k}$ is computed as a moving weighed average of the external air temperature from the last 48 h (similar to what is used for calculating the position of the window blinds [4]). The original version in Rijal et al. [31] only studied the relationship with regards to current external air temperature by using a moving, weighed average of the recent temperatures' results in a control signal that models occupants as acting on both current conditions (active control) and recent conditions (delayed control), which seems to agree better with the results presented in the extensive review article by Fabi et al. [33]: 70% of the venting is assumed related to the external air temperature, and the other 30% is assumed constant. However, a time-related component is not accounted for, which has been reported to be present in typical venting patterns [34,35].

Swedish multifamily buildings typically rely on manual window venting to avoid overheating during hot summer days. The control signal for manual venting during cooling season is modeled with the following logistic function:

$$u_{mve;C;k} = \frac{1}{1 + \exp(-2 \cdot (\theta_{op;k} - 24))} \quad (16)$$

where $\theta_{op;k}$ is the operative temperature (computed accordingly to the ISO 52016-1:2017 [23] standard), which is, for practical purposes, equal to the globe temperature measurements used by Rijal et al. [31].

3.3. Hydronic Heating System

The heat supplied by the hydronic heating system is predicted at each time step using the model developed in [4] (the thermostatic radiator valve equation was updated to a inverse logistic-function smoothed PD-controller, which is more stable and better suited for automatic differentiation). The uncertainty of the heat supplied by the hydronic heating system, $\sigma_{\phi;hyd;k}$, is represented with following empirically derived equation:

$$\sigma_{\phi;hyd;k} = 0.02 \cdot (\hat{\phi}_{hyd;k} + 10 \cdot u_{trv;k}) \quad (17)$$

where $\hat{\phi}_{hyd}$ is the estimated supplied heat and u_{trv} is the control signal from the thermostatic radiator valves. The rationale for this equation is that the uncertainty should increase with $\hat{\phi}_{hyd;k}$, but there is also often uncertainty when $\hat{\phi}_{hyd;k}$ is close to zero. The inclusion of $u_{trv;k}$ ensures that the uncertainty

approaches zero as indoor temperature rises above the set-point temperature, e.g., during warm summer days.

3.4. Internal Heat Gains

Several bottom-up approaches [16,36,37] for modeling internal heat gains induced by occupants and electrical equipment have been reported in literature. For brevity, we used a simple reference metering-based model to estimate the internal heat gains:

$$\hat{\phi}_{int;k} = f_{dhe}(wn_k, wd_k, h_k) + \phi_b \quad (18)$$

where f_{dhe} is a function that matches hourly metered domestic household electricity from reference buildings based on week number wn , working day or non-working day wd , and hour of the day h for each time step k and where ϕ_b is constant base load, which is assumed to be 2 W/m^2 to account for heat load that is not included in domestic household electricity (mainly heat dissipation from occupants): 80% of household electricity was assumed to end up as useful heat gains [38]. The reference metering is based on 3 years of hourly metering data from 121 apartments in multifamily buildings. This model can be assumed to include major patterns: the diurnal pattern of the occupants, the difference between working and non-working days, and the seasonal variation in available daylight (affecting heat-dispatch from electrical lighting).

The uncertainty is modeled as normally distributed:

$$\sigma_{\phi;int;k} = \sigma_{\%int} \cdot \hat{\phi}_{int;k} \quad (19)$$

where $\sigma_{\%int}$ is the relative standard deviation calculated according to Equation (20). The assumption of normal distribution holds quite well when the number of apartments is large, but when there are fewer than approximately 10 apartments, a right-tailed distribution (e.g., log-normal) starts to better match observations from the dataset. The probability density distributions for hourly domestic household electricity resemble those for DHW in Figure 6 (to be described in Section 3.5) but are more normally distributed. Based on the observation of both datasets, the following relationship was derived to model the relative standard deviation as a function of the number of apartments, N :

$$\sigma_{\%} = 1.5/\sqrt{N} \quad (20)$$

3.5. Domestic Hot Water

The following equation is used to estimate the heat required for DHW use:

$$\hat{\phi}_{dhw;k} = q_{V;dhw;k} \cdot 4180 \cdot (\theta_{dhw;set} - \theta_{gr4;k}), \quad (21)$$

where $q_{V;dhw}$ ($l/(s \cdot m^2_{fl})$) is the DHW water volume flow averaged per hour, 4180 is the heat capacity of water in $W \cdot s/(l \cdot K)$, $\theta_{dhw;set}$ is the desired set-point temperature of the DHW (typically around $55 \text{ }^\circ\text{C}$, as that is the minimum temperature required by Swedish building regulations), and θ_{gr4} is the ground temperature representative of the soil layer at 1.0–2.89 m (assumed to reflect incoming cold water temperature).

The availability of data and quality on $q_{V;dhw}$ varies greatly within the building stock. Recent Swedish building regulations require that new buildings and buildings that undergo a major renovation have separate DHW metering (individually per apartment for multifamily buildings). However, in practice, hourly measured DHW use is still often not available either for new or older buildings. In Sweden, domestic cold water (DCW) use is often gathered with the same infrastructure that is used for gathering district heating (DH) use and such data is therefore available for large building stocks in a systematic way. However, the resolution of the hourly DCW values can be of poor quality (much of the current billing metering infrastructure was built in a time when only monthly

values were required [39]). Also, the placement of the DCW meters can be quite different than the placement of DH meters, which is the case in our case study (see Figure 7). The share of DCW used for preparation of DHW can generally be estimated as:

$$q_{V,dhw;k} = \alpha_k \cdot q_{V,dcw} \quad (22)$$

where α_k varies, but has been reported to reach an average of 44–50% over time [30].

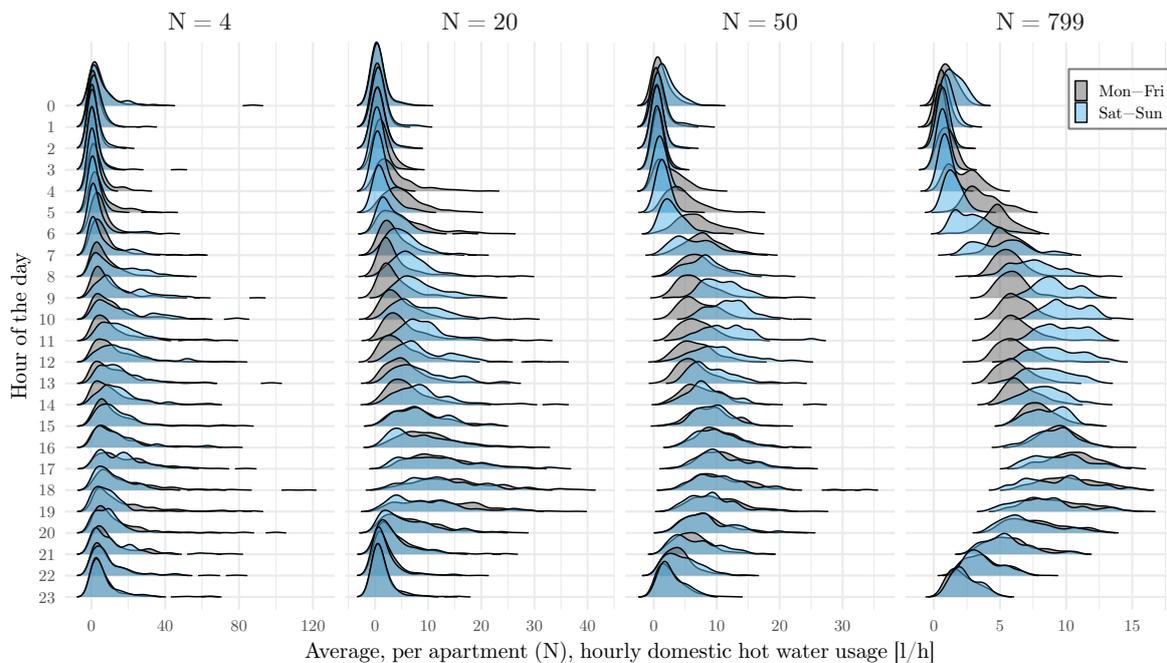


Figure 6. One year of domestic hot water (DHW) usage probability density distributions, grouped per hour and weekday: N denotes the number of apartments over which the measurements are averaged.

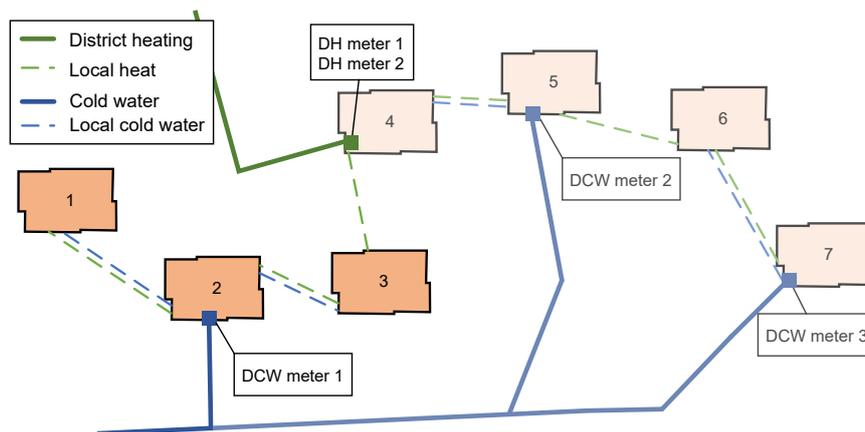


Figure 7. Overview of the case study, where buildings 1–3 are the ones being modeled.

We identified three commonly occurring situations of data availability: (i) no hourly data, the DHW needs to be modeled (for example, by using reference metering from similar buildings); (ii) useful hourly domestic cold water metering exists; or (iii) actual hourly DHW use metering exists (metered for the whole building or aggregated from individual, per apartment, metering). A common factor is that usually only flow is measured; therefore, Equation (21) needs to be applicable to all three cases above. In order to test methods that are applicable for the three situations mentioned above, individual billing metering data for DHW use was gathered. The data consist of hourly time-series,

grouped per apartment (799 in total), from 38 multifamily buildings located in Eskilstuna, Sweden and was gathered from 2016 to 2018. Figure 6 shows probability density distributions for hourly DHW use based on that data. From the figure, it is apparent that there are diurnal and workday-related patterns. When averaging over a large number of apartments, the probability density tends to follow a close-to-normal distribution. However, when averaging over a smaller number of apartments, the long right-tailed distribution is better described as log-normal. Based on observation of the dataset, the following equation was derived to model the uncertainty (log-scaled) standard deviation of a log-normal distribution:

$$\sigma_{\phi;dhw;ln;k} = \sigma_{\%;dhw} \cdot \left(\frac{a \cdot \bar{\phi}_{dhw}}{(\hat{\phi}_{dhw;k} + 1)} + 1 - a \right) \quad (23)$$

where $\sigma_{\%;dhw}$ is the nominal relative standard deviation Equation (20) used to relate it to number of apartments, $\bar{\phi}_{dhw}$ is the average DHW heat use, and a is a constant between 0 and 1 that determines whether the uncertainty is a constant or proportional. Setting $a = 1$ results in a constant uncertainty of $\sigma_{\%;dhw} \cdot \hat{\phi}_{dhw}$, whereas setting $a = 0$ result in a proportional uncertainty of $\sigma_{\%;dhw} \cdot \phi_{dhw;k}$. Setting $a = 0.7$ was found to approximately reflect the diurnal pattern in uncertainty shown in Figure 6. Equation (23) can be somewhat counterintuitive due to the log-normal transformation; it can be helpful to think of the standard deviation in logarithmic scale as corresponding to the relative standard deviation in normal scale.

3.6. Solar Heat Gains

Solar heat gains are estimated with following deterministic equation:

$$\phi_{sol;k} = r_{gl} \cdot g_{gl} \cdot g_{bl;k} \cdot I_{tot;ver;sh;k} \quad (24)$$

where r_{gl} is the glazing-to-floor area ratio, g_{gl} is the solar energy transmittance of the glazing, $g_{bl;k}$ is the time-varying solar energy transmittance of the window blinds, and $I_{tot;ver;sh;k}$ is the surface orientation weighted hemispherical solar irradiance on vertical surfaces including shading effects. Satellite-based solar radiation is acquired from Copernicus Atmosphere Monitoring Service (CAMS) radiation service [40]. Calculation procedures developed for Reference [4] based on the ISO 52010:2017 [41] standard are used to calculate the solar irradiance at an arbitrarily tilted and oriented surface (code available at <https://github.com/lukas-rokka/solarCalcISO52010>). The shading and window blind operation calculation are described in Lundström et al. [4].

The solar heat gain uncertainty is assumed to be normally distributed, and the standard deviation is modeled as follows:

$$\sigma_{\phi;sol;k} = \sigma_{\%;sol} \cdot \bar{\phi}_{sol} \cdot \frac{\ln(\hat{\phi}_{sol;k} + 1)}{\ln(\bar{\phi}_{sol} + 1)} \cdot \frac{3 - r}{2} \quad (25)$$

where $\sigma_{\%;sol}$ is the relative standard deviation, $\bar{\phi}_{sol}$ is the long-term average solar heat gain (for hours with nonzero heat gain), $\hat{\phi}_{sol;k}$ is the estimated solar heat gain for time step k , and r is a reliability code in the CAMS dataset that ranges between 0 and 1 and encodes cloud information availability for the particular time step [40]. The log-transform term, $\ln(\hat{\phi}_{sol;k} + 1) / \ln(\bar{\phi}_{sol} + 1)$, models a standard deviation that rises quickly towards the expected standard deviation of $0.3 \cdot \bar{\phi}_{sol}$ and then flattens out when $\hat{\phi}_{sol;k} > \bar{\phi}_{sol}$. According to the CAMS validation report [42], the relative standard deviation (relative to the mean of one year of measurements) is in the order of 15–30% for hourly global horizontal irradiance for stations located in northern Europe. In this paper, we use the higher value of 30% to account for the uncertainty added due to shading and window blinds.

4. Case Study

In this section, the proposed modeling framework is tested on a group of three buildings sharing the same district heating substation. Section 4.1 describes the buildings and used data sources. Section 4.2 presents results from conducting parameter estimation (Section 4.2.1), the behavior of the proposed filtering method (Section 4.2.2), and four experiments that address different sources of data regarding DHW use and the uncertainty in measured indoor air temperature (Section 4.2.3).

4.1. Setup

Buildings 1–3 in Figure 7 share a common district heating meter (“DH meter 1”) and were chosen for the case study. Buildings 4–7 are not explicitly modeled, but one of the experiments use metered domestic cold water use as input, measurements that are shared with all seven buildings. Description of the buildings is as follows: four floors (bottom floor has both apartments and common space) built in 1967 with brick facade, flat roof with retrofitted extra insulation, windows retrofitted with an insulating inner glazing pane, 59 apartments, mechanical exhaust ventilation, and a hydronic heating system consisting of the original plate radiators (in Sweden known as HE radiators) with thermostatic radiator valves on the return pipes. Among tenants who responded to a customer satisfaction questionnaire, only 20% responded that they were satisfied with the indoor thermal comfort (typically, a satisfaction rate of at least 70% is targeted). This can be attributed to cold drafts from the air inlets, varying indoor temperatures, and temporal faults in heat supply. The radiator type is notoriously difficult to control, and the ventilation is suspected to be unbalanced due to tenants tampering air inlets (cold draft causes tenants to seal air inlets, which tends to result in a vicious circle of increased drafts in other apartments and air inlets in other apartments getting sealed as well).

Figure 8 shows the selection of parameters that were chosen as uncertain (i.e., part of θ) and which energy flows they impact). The prior model parameterization is shown in Table 1. $\mathcal{N}(\mu, \sigma)$ denotes a normal probabilistic density function (PDF), in which μ and σ are the mean and standard deviation, respectively. The $\mathcal{N}_{ln}(\mu, \sigma)$ denotes a log-normal PDF, where μ and σ are the mean and standard deviation of the natural logarithm of the variable, respectively. $\text{EMG}(\mu, \sigma, \lambda)$ denotes an exponentially modified Gaussian (normal) PDF, where μ and σ are the mean and standard deviation of a normal distribution and λ is the inverse of the exponential decay (high λ results in a slowly decaying probability density in the upper tail). $\mathcal{N}_{MV}(\mu, \Sigma)$ denotes a multivariate normal PDF, where μ is a vector of mean values and Σ is a covariance matrix. The $\Sigma(\sigma, \rho)$ denotes a covariance matrix with standard deviations σ and a correlation of ρ .

Two time periods of data were used for the calibrations, 5 July 2016 to 10 August 2016 and 21 December 2016 to 1 March 2017. The first ten days of both periods were discarded from the likelihood calculation and were therefore only used to initialize inert state variables (ground and internal mass). Two periods were used so that different seasonal conditions would be represented in the data (e.g., thermal transmittance parameters need the heating season to be learned well, whereas the heat losses for DHW circulation are better learned during the non-heating season).

Shading and sheltering impact of the surroundings were modeled using the category method developed in Lundström et al. [4]. As the site of the case study is well sheltered, a shading/sheltering class 4 was used (ranging from (1) no shelter to (5) a highly dense urban area). Class 4 results in a reduction of approximately 35% in annual solar irradiance, compared to no shading. Also, the effect of air infiltration and exterior surface convective heat transfer are substantially reduced compared to an open field site.

Table 1. Input parameters of the case study: Parameters considered as uncertain are expressed as probability density functions.

Symbol	Description	Value/Distribution	Unit
	Geographic coordinates	N 59.36, E 16.50	°
N_b	Number of buildings	3	-
H_{fl}	Ceiling height	2.6	m
N_{fl}	Number of floors	4	-
P	Building perimeter (per building)	$(90)_{\times 3}$	m
A_{fl}	Floor area (all buildings)	5754	m_{fl}^2
r	Surface to floor area ratios for roof, external walls, windows, internal mass, ground floor	$(0.25, 0.36, 0.13, 3.5, 0.25)$	m_s^2/m_{fl}^2
γ	Azimuth angles for the external walls	$(165, 75, -15, 105)_{\times 3}$	°
F_γ	Surface area fraction for the external walls	$(0.19, 0.31, 0.31, 0.19)_{\times 3}$	-
xy	Supply temperature look-up table	$\begin{bmatrix} -17 & -10 & 0 & 18 \\ 60 & 55 & 43 & 18 \end{bmatrix}$	°C
L_p	Length of piping network	110	m
$q_{V,mve;H}$	Nominal manual venting, heating season	0.03	$1/(s \cdot m_{fl}^2)$
$q_{V,mve;C}$	Nominal manual venting, cooling season	0.25	$1/(s \cdot m_{fl}^2)$
Ψ_p	Linear heat loss coefficient for ground buried pipes	$\mathcal{N}_{ln}(\ln(0.25), 0.05)$	$W/(m \cdot K)$
ϕ_{dhwc}	Heat losses DHW circulation	$\mathcal{N}_{ln}(\ln(0.7), 0.15)$	W/m_{fl}^2
$\theta_{int;set}$	Internal air set-point temperature	$\mathcal{N}(22.5, 0.1)$	°C
$q_{V,ahu}$	Specific air flow rate of the air handling unit	$\mathcal{N}(0.35, 0.01)$	$1/(s \cdot m_{fl}^2)$
C_{inf}	Specific infiltration coefficient	EMG(0.06, 0.001, 200)	$1/(s \cdot m_{fl}^2 Pa^n)$
g_{gl}	Solar heat gain coefficient	$\mathcal{N}(0.65, 0.02)$	-
H_{hyd}	Specific radiator constant	$\mathcal{N}(0.45, 0.02)$	$W/(K \cdot m_{fl}^2)$
F_{int}	Bias factor internal heat gains	$\mathcal{N}(1, 0.05)$	-
F_{dhw}	Bias factor DHW	$\mathcal{N}(1, 0.05)$	-
U_{rf}	Thermal transmittance roof	$\mathcal{N}_{ln}(\ln(0.1), 0.15)$	$W/(K \cdot m_s^2)$
U_{ew}	Thermal transmittance external walls	$\mathcal{N}_{ln}(\ln(0.25), 0.15)$	$W/(K \cdot m_s^2)$
U_{gl}	Thermal transmittance windows	$\mathcal{N}(1.5, 0.2)$	$W/(K \cdot m_s^2)$
U_{gf}	Thermal transmittance ground floor	$\mathcal{N}_{ln}(\ln(0.25), 0.15)$	$W/(K \cdot m_s^2)$
H_{tb}	Specific heat transfer thermal bridges	$\mathcal{N}_{ln}(\ln(0.06), 0.1)$	$W/(K \cdot m_{fl}^2)$
κ_m	Areal heat capacity for roof, external walls, internal mass, ground floor	$\mathcal{N}_{MV}((69, 31, 49, 69), \Sigma((1)_{\times 4}, \rho = 1))$	$Wh/(K \cdot m_s^2)$

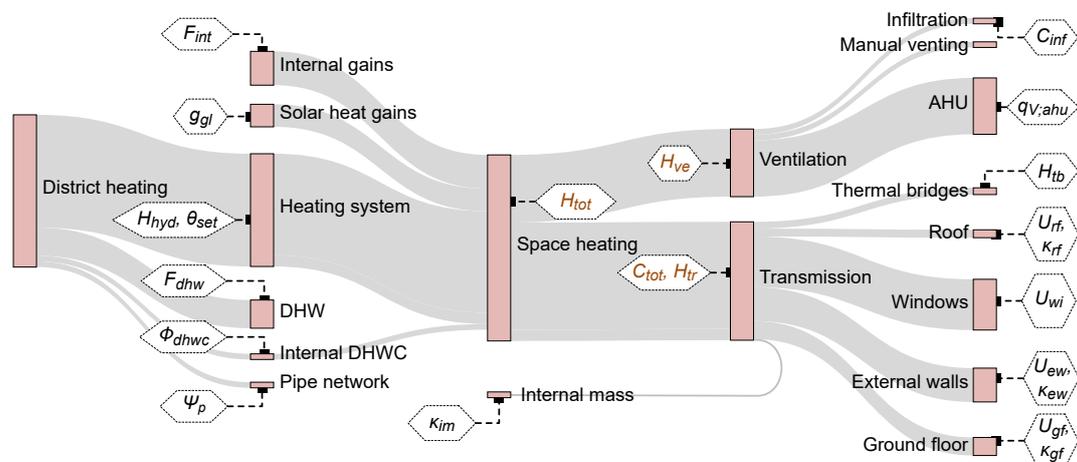


Figure 8. Parameters selected as uncertain, placed at the energy flow they influence: Brown colored names show parameters that are calculated from lower levels and, thus, not explicitly given prior probability densities. Height of the gray areas indicates relative share of the yearly heat flow, based on the prior model.

4.1.1. Data Availability Experiments

To study the effect of commonly occurring data availability situations, four experiments were conducted: experiments 1–3 compare different sources of DHW input (see Section 3.5 for motivation), and experiment 4 studies the impact of less uncertain measurements on indoor temperature. Results are presented in Section 4.2.3. The original plan for experiment 4 was to compare using actual on-site measurements with modeled input (using the same reference metering approach as used for DHW in experiment 1). The measured indoor temperature is based on six temperature meters—two for each building. The uncertainty is periodically very large, resulting in a small influence on the likelihood calculation. Using on-site measurements resulted in a barely noticeable difference compared to using modeled indoor temperatures. Therefore, the fourth experiment was modified and constructed as specified below.

Experiment 1: This case assumes no on-site measurements are available for DHW and, instead, makes use of reference input composed of hourly measured DHW flow rates from similar buildings and exactly the same time period. See Section 3.3 for the calculation procedure. The nominal relative standard deviation is estimated at 20% for this case study with its 59 apartments ($1.5/\sqrt{59} = 0.20$).

Experiment 2: This experiment uses hourly domestic cold water (DCW) billing metering data. The share of DCW used for preparation of DHW can generally be estimated with Equation (22). Unfortunately, the DCW meter structure for our case study is complex and the DCW for our three buildings could only be estimated. The following weighed sum from two meters was used: $q_{V;dcw} = q_{V;dcw1} + q_{V;dcw2} \cdot 3/9$. “DCW meter 1” in Figure 7 measures DCW in buildings 1–3, while “DCW meter 2” measures DCW for DHW preparation for all 7 buildings and DCW in buildings 4 and 5. “DCW meter 3” measures DCW for buildings 6 and 7 and can thus be omitted. Equation (23) was used to model the uncertainty, assuming a relative standard deviation of 10%.

Experiment 3: In this experiment, individual billing metering data is used. Hourly measured DHW flow rates values were measured at 129 tapping places of the 59 apartments of buildings under study. The resolution of the water flow metering data is 10 liters; one meter had faulty values and was removed. Equation (23) was used to model the uncertainty, assuming a relative standard deviation of 5%.

Experiment 4: To simulate a situation where more certain indoor temperature measurements are available and to get insight of the influence of such data, the measured standard deviations were divided by 3 for this experiment. DHW is the same as in experiment 1.

4.2. Results

4.2.1. Bayesian Parameter Estimation

Bayesian inference, as described in Section 2, was conducted on the demonstration of the case presented in Section 4.1. Resulting parameter estimates are shown in Figure 9 (the energy flows controlled by the parameters are visualized in Figure 8). Induced prior density distributions are shown for parameters H_{tot} , H_{tr} , H_{ve} , and C_{tot} , while the prior distributions for the other parameters are explicitly defined by their PDFs, as presented in Table 1. Figure 10 shows the correlations of the HMC inferred parameter posteriors.

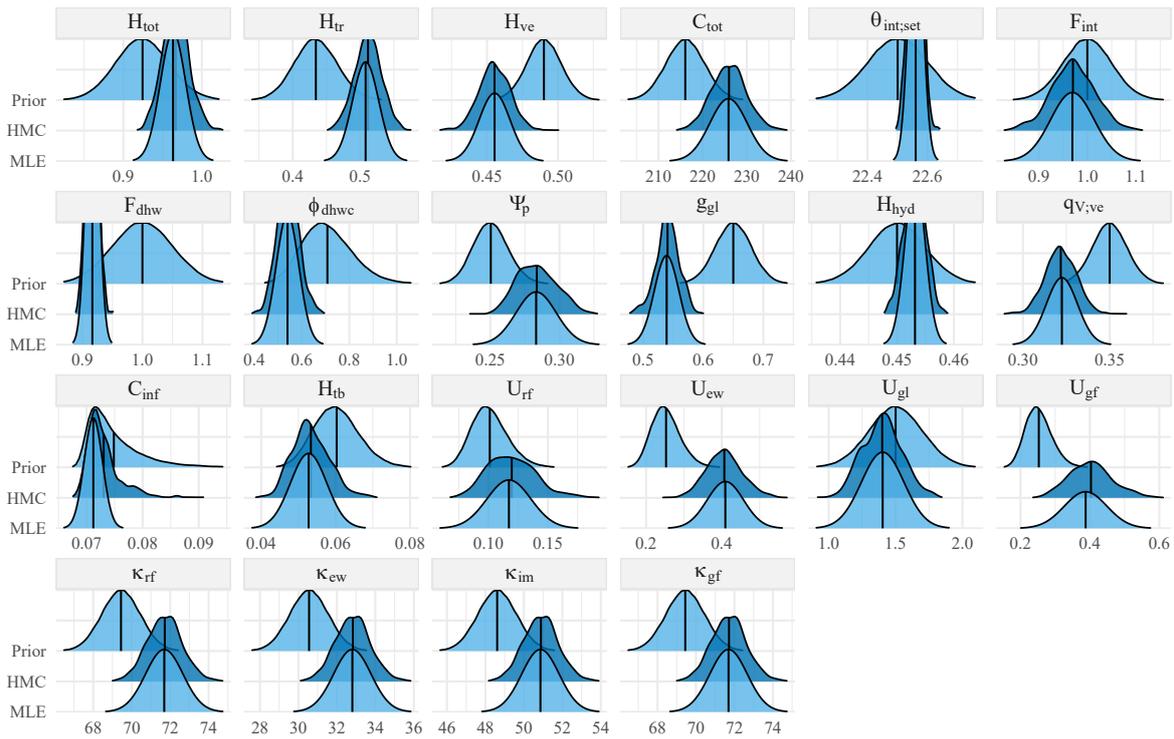


Figure 9. Parameter probability density distributions. Prior: the prior model; HMC: posterior inferred with Hamiltonian Monte Carlo sampling; MLE: posterior inferred with penalized maximum likelihood estimation. Vertical lines show mean values. Densities are scaled per parameter such that the mode of the prior has a relative height of 200%; densities below 0.01% are filtered out.

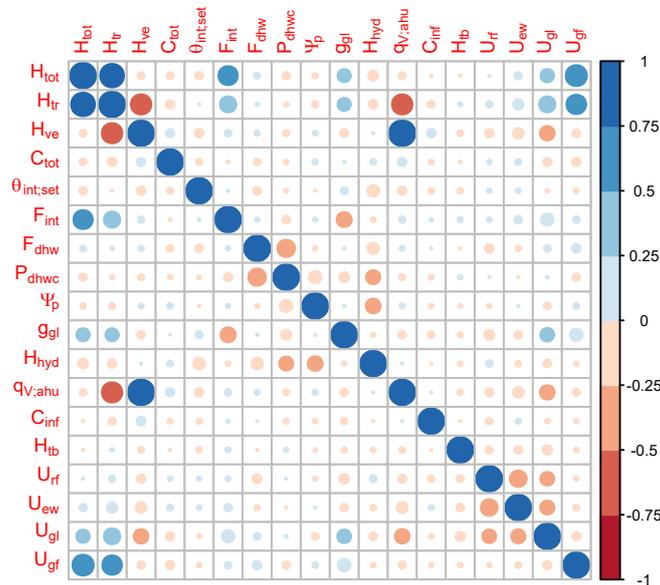


Figure 10. Parameter correlations based on the HMC samples: Color saturation and size of the circles indicate the magnitude of correlation.

Equation (24) deterministically models solar heat gains by considering glazing-to-floor area ratio, solar energy transmittance of the glazing, occupant operated window blinds, shading from the surroundings, and orientation and areas of the external envelope. However, only the solar transmittance parameter g_{gl} was chosen as uncertain, although it could be argued that the other solar gain governing parameters are just as uncertain. The strong downward push on the posterior of the g_{gl}

parameter suggests that the calculated solar heat gains are still overestimated. That deterministically modeled solar gains tend to be overestimated is consistent earlier results reported by Sjögren et al. [17].

The correlations seen in Figure 10 follow what can be expected. The relatively strong negative correlation between H_{tr} and H_{ve} confirms that heat loss occurring as air-exchange or as heat transmission through the building fabric is difficult to separate. This is also the case with the F_{dhw} , $F_{dhw,c}$ and Ψ_p parameters, which all have a similar “base load”-affecting effect. The relatively strong positive correlation between H_{tot} and the heat gains governing parameters F_{int} and g_{gl} confirms that uncertainty in these heat gains inputs will affect how well the energy performance of the building fabric can be assessed.

Separating the transmission heat loss through the building fabric H_{tr} parameters is even more challenging. In a mathematical sense, these parameters are identifiable—the probability density of the likelihood will start to decrease if the parameter estimate gets far from the true value. However, in the region close to the true value (the region we, in practice, are interested in) the likelihood flattens out—e.g., practically no change in likelihood is obtained by increasing $r_{rf} \cdot U_{rf}$ with a small number and decreasing $r_{ew} \cdot U_{ew}$ with the same small number. This effect manifests as negative correlations between U_{rf} , U_{ew} , and U_{gl} in Figure 10. The data (through the likelihood function) usually carry strong evidence about the total heat loss but has much less information about individual U-values. On the other hand, our prior knowledge is typically more informed and intuitive to specify on the individual building element level.

The Gaussian approximation approach of the penalized MLE matches the result of the HMC sampling. The MLE method was on the order of 100 times faster than HMC sampling. The MLE parameter estimation took approximately a half minute to compute on a standard laptop computer, which largely can be attributed to the automatic differentiation and fast linear algebra enabled by Stan.

4.2.2. The Behavior of the Filter

The behavior of the filter is illustrated by running both filtered and deterministic (filtering turned off) simulations, based on the penalized MLE parameter modes of Section 4.2.1. At every time step, the measurement-update step of the filter conducts a correction to state variables according to the Kalman gain matrix, which in turn depends on the specified time-varying uncertainties of the states and measurements. The correction (Figure 11d) is conducted on the uncertain inputs (Figure 11c) (note that the interior surface temperature nodes of the thermal network are also corrected, but only indoor air temperature is shown in the figure).

Figure 11c shows the details of the uncertain inputs (viewing the figures online is recommended as they are vector-based and can be zoomed in). The augmented states heating system, internal heat gains, and solar heat gains are modeled as normally distributed. Their uncertainty band is symmetrical in both directions, and the corrected estimates can occasionally go below zero. This is alleviated by setting the uncertainty to zero when the predicted mean is clearly zero (e.g., solar gain during nights or heating when indoor temperature is much higher than the set-point). Indoor temperatures close or below the set-point and predicted heating close to zero do occasionally result in corrected estimates below zero for heating. During the extreme cold days of 5–7th of January, there was a relatively large deviation between the model outputs and the measured values. The heat use measurement suggests lower energy use than predicted by the model, whereas the measured indoor air temperature suggests higher temperatures than predicted; the filtered estimates are a compromise between these contradicting signals.

During the nights of 11–12th of August, there was a fault in the heating system and no heat was delivered (the measured low indoor temperature is probably correct). The model predicts that the heating system should deliver heat, but this is not supported by the measurements and a correction is applied by the filter. However, the correction is not strong enough (in the light of the knowledge that the heating system was faulty). Such events could be dealt with by using distributions with longer tails or adaptive filters.

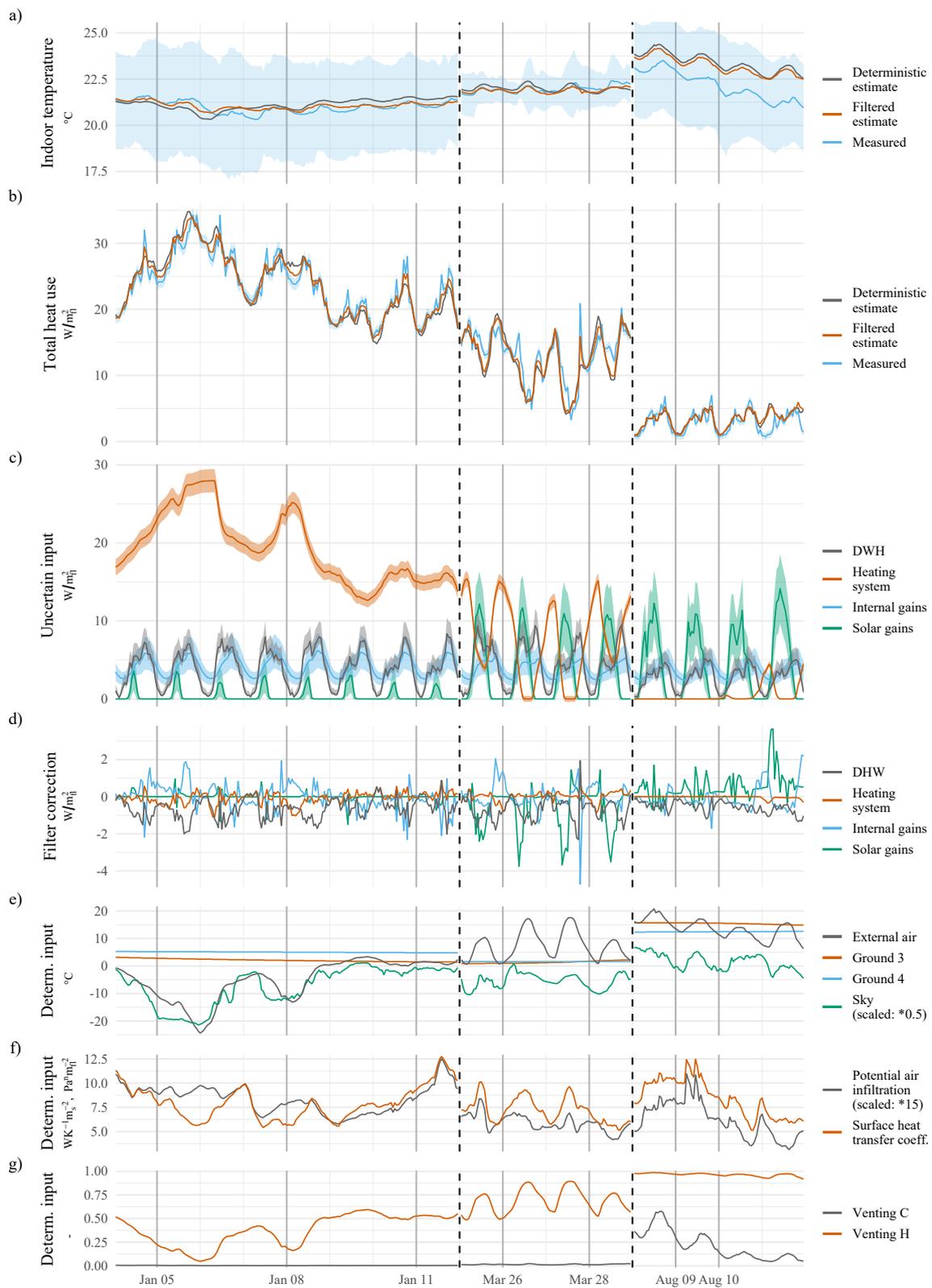


Figure 11. Hourly inputs and outputs of the Bayesian filtering simulation: (a,b) Comparison of measured values to both filtered and deterministic simulation outputs for indoor air temperature and total heat use; (c) normally scaled uncertain inputs; (d) normally scaled corrections (i.e., augmented states); (e–g) the deterministic inputs: 95% confidence interval is shown for the measured outputs and uncertain inputs.

4.2.3. Data Availability Experiments

To study the effect different sources of input and output data, four experiments were conducted: experiments 1–3 compare different sources for the DHW input, and experiment 4 studies the impact of less uncertain measurements on indoor temperature output (see Section 4.1.1 for details and motivation for the experimental setup). The resulting parameter estimates are shown in Figure 12 and yearly energy flow estimates are given in Table 2.

Due to the different uncertainty models, the posterior log-likelihood is not comparable between the experiments. To get an indication of which source of DHW input carries most information, experiments 1–3 were rerun using the uncertainty model of experiment 1 for all experiments. The resulting posterior log-likelihoods are -1785 , -1472 , and -1394 , with a standard deviation of 3. This suggests that using actual metering data from the building (experiments 2 and 3) substantially improves calibration compared with using a reference input (experiment 1). A more surprising result is that the log-likelihood of experiment 2 is relatively close to that of experiment 3; using cold water meter data to predict DHW heat use thus appears to be a promising approach, especially if a cleaner metering signal is available (in our case study, the meter was shared with neighboring buildings). Parameter estimates are fairly consistent between the experiments, but experiment 3 shows a push upward for parameters which captures a pattern of delayed heat use (thermal mass and the transmittance parameters of building elements with thermal mass).

Experiment 4 simulates a situation where more certain indoor temperature measurements are available. The result shows much smaller estimated solar heat gains, less “base load” and more space heating. The heat transmittance of the ground floor (U_{gf}) and the total heat capacity (C_{tot}) are also affected, which are parameters that influence time delays in energy use and indoor temperatures.

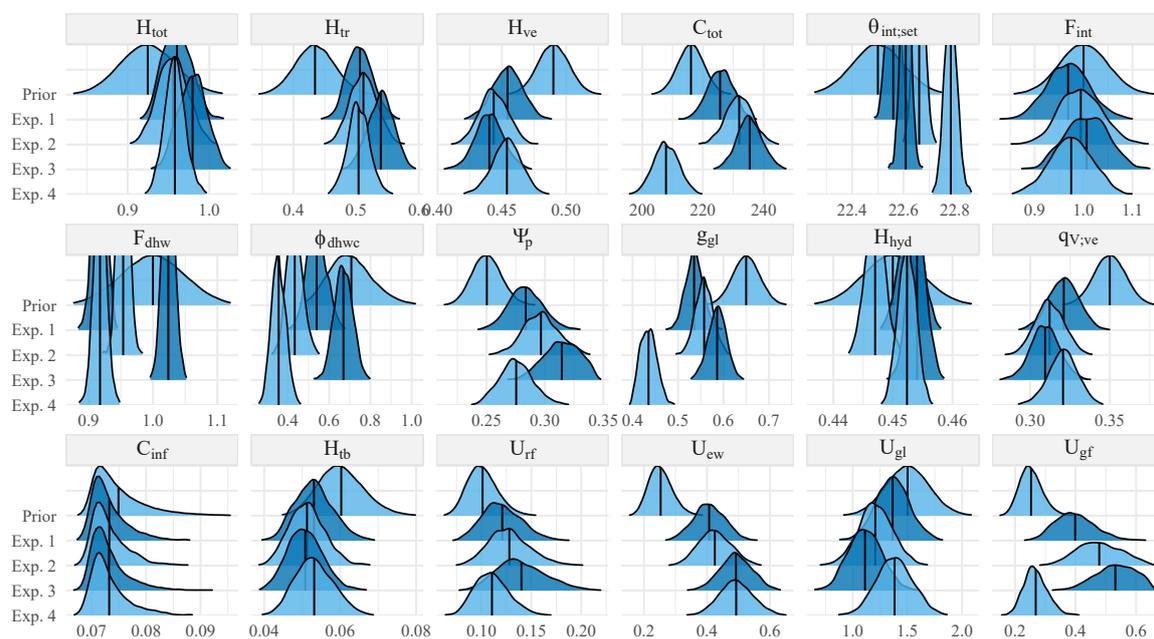


Figure 12. Parameter probability density distributions showing the prior model and HMC sampled posteriors for the four experiments. Experiment 1: DHW based on reference measurements. Experiment 2: DHW based on measured total cold water use. Experiment 3: DHW based on individual DHW measurements per apartment. Experiment 4: The standard deviation of the measured indoor air temperature divided with 3 (DHW is the same as in Experiment 3). Vertical lines show mean values. Densities are scaled per parameter such that the mode of the prior has a relative height of 200%.

Table 2. Estimated specific thermal energy flows in kWh/(m_{fl}² · year): Estimated as the hourly average of the training period multiplied by 8760 h (thus not representative for a typical year).

Model	Q_{tot}	Q_{hyd}	Q_{dhw}	Q_p	Q_{dhw_c}	Q_{sol}	Q_{int}
Prior	140.6	98.1	32.0	4.4	6.1	22.5	33.9
Exp. 1	137.9	100.6	27.6	5.0	4.7	16.7	32.3
Exp. 2	138.2	99.7	30.1	5.2	3.8	18.1	33.0
Exp. 3	139.5	100.0	28.1	5.5	5.9	19.6	34.6
Exp. 4	137.6	101.7	27.9	4.9	3.1	13.2	32.6

5. Discussion

5.1. Bayesian Calibration

The developed energy model is relatively detailed, and identifying the full set of parameters Θ with only likelihood function is not feasible: the likelihood might not carry enough information for some parameters, and it is difficult to separate the effects of other parameters. An informative prior model is needed to constrain the parameter space. The data (through the likelihood function) usually carries strong evidence about the total heat loss but has much less information about the individual parameters. On the other hand, our prior knowledge is typically more informed and intuitive to specify on the individual building element level. The selection of parameters that are predetermined as uncertain θ should be seen as part of the prior model specification. The used parameterization was based on a qualitative assessment by the authors, where aspects such as ease of prior specification, anticipated influence, and insight gain were considered. More formal parameter screening such sensitivity analysis (e.g., Senave et al. [1] or Chong and Menberg [7]) could be used for a more quantitative assessment.

The following ideal goals were formulated as a basis for the development work: 1. each parameter should describe one property; 2. the posterior should encompass the true value; 3. the prior should encompass the posterior; and 4. the prior model should be informative enough to be useful on its own. Goal 1 might be unreachable as there are interacting effects that are difficult to fully separate into a single parameter. If we are too far from goal 1, then goal 2 becomes more difficult to guarantee. Goal 3 could be reached by defining a very noninformative prior model, but this would breach goal 4 and leave us with a useless model in the absence of calibration data. Furthermore, a sufficiently informative prior model can, due to its regulating impact, help us in achieving goal 1 and thereby goal 2 as well. In essence, the entire process is somewhat iterative: reliable output is dependent on decent input which, in turn, is dependent on statistics from previous outputs and case-specific input data.

External air temperature is the strongest driver for energy use, but its effect manifests in different time lags. There is an instantaneous correlation that mainly affects heat loss through glazing, thermal bridges, and ventilation; heat losses through the envelope with thermal mass have more delayed effects (hours and up to a few days) while heat loss through the ground floor or external piping network has a delayed effect that is best described as seasonal. The occupant-dependent DHW preparation and internal heat gains have strong diurnal patterns and a seasonal pattern. Solar heat gains are most instantaneously correlated to the solar irradiance, but there are some delayed effects due to heat stored on the interior surface layers and heat conducted through the envelope. Also, solar irradiance has both seasonal and diurnal patterns. The calibration procedure can be seen as pattern matching. The data carries information about the abovementioned patterns, but the prior model needs to be relatively informative to enable separation of the various effects into specific heat flows.

The parameters of the prior model are set as independent, i.e., specified without any correlation. Specifying a correlation structure for a prior model with nonnormal parameters is a complex task, and as long the likelihood is sufficiently rich, there is no need to specify correlations. As seen in Figure 10, the correlation structure is identified in the posterior HMC sampled parameters. However, if the prior model is to be used by itself, the assumption of independent parameters might result in overly

optimistic uncertainty intervals, summing independent parameters results in a tighter uncertainty interval than summing correlated parameters. The total heat transfer coefficient, H_{tot} is a weighted sum of seven parameters that we expect to correlate and is therefore more sensitive for such an effect.

Kristensen et al. [2] used hierarchical modeling to pool information between energy models through the likelihood function. Such an approach could also be used with our proposed models, thus allowing energy models of buildings that lack measurement data to be informed by data from other buildings that can be reliably measured.

5.2. Bayesian Filtering with Time-Varying Uncertainty on Inputs

We propose an augmented stochastic state-space model and iterated extended Kalman filtering. Both Kim and Park [19] and Coffman and Barooah [9] proposed using augmented state-space modeling to identify non-measured process disturbance. They use one augmented state to model the non-measured process disturbance and to identify it as part of the estimation procedure. We propose four augmented states to enable definition of predetermined time-varying uncertainty separately for major uncertainty sources. The increased number of disturbance states does not necessarily provide increased prediction power. Rather, the main motivation is that it is conceptually more appealing and more natural to incorporate prior knowledge when these sources of uncertainty are kept separate. The augmentation allows incorporation of prior knowledge, such as solar heat gains, which are highly uncertain during day time but certainly zero during night time or that DHW use is often close to zero (but has right-skewed outliers) at night. Another effect of using uncertain inputs is that parameters are learned with more weight during periods of low uncertainty in the inputs (e.g., night time) than during periods of high uncertainty (e.g., when occupant activity peaks and high insolation is plausible).

The DHW is modeled as log-normally distributed while the other inputs are modeled as normal distributions. The log-normal distribution has a multiplicative uncertainty, i.e., right-tailed and always positive, which describes the nature of uncertain energy flows very well. Modelling the other augmented states as log-normal was also considered, but a normal scale was chosen for its conceptual simplicity and computational efficiency. DHW can be modeled as nonnormal with relatively simple alteration of the measurement function h , while the other uncertain inputs are connected to the thermal network and thus require a nonlinear state transition function f (discarding the benefits of computationally fast linear algebra). The more apartments that are served by a DH substation, the better a normal distribution approximates the occupant-dependent inputs (an effect known as the central limit theorem). Thus, single family and smaller multifamily buildings benefit more from nonlinear modeling and/or actual measurements on the inputs.

5.3. Data Availability

The results of Section 4.2.3 show that DHW inputs based on on-site measurements are substantially more informative than modeled inputs. In practice, measurement-based input data are frequently not available. We have shown that reference metering from similar buildings can substitute measured data. Such reference metering will be noisy on a hourly time-scale but captures the general correlations to weather and time and possibly to other aspects such as demographics. In our case study, we had access to reference metering for DHW and indoor air temperatures for the exact same period as used for the training but not for the domestic household electricity. The matching method described in Section 3.4 was used for internal heat gains; this is a relatively simple modeling approach compared with existing bottom-up based approaches in the literature [16,36,37]. Using a modeling approach for all the inputs would also enable calibration when measurements are lacking and for out-of-sample predictions with, e.g., forecast or normal year weather data.

A few weeks of data during the heating season is enough to identify the main parameters describing thermal behavior. However, different weather conditions contain different types of information, and using a longer period of data is beneficial for learning more aspects of the model/building. For example, heat-use data during the summer holds information about DHW

use and heat losses due to DHW circulation. If there exist indoor air temperature measurements, the summer season captures information about the insulation and thermal heat capacity (which is difficult to identify during the heating season for buildings that hold the set-point temperature well). Our method of using different time periods for training has the advantage of learning different aspects about the buildings in a computationally efficient way and allows skipping of periods with faulty or missing data.

5.4. Energy Modeling

In this paper, we have presented *one* energy model that we anticipate as general and accurate enough to sufficiently approximate a large proportion of the targeted Swedish district-heated multifamily buildings. However, there are also buildings within this category that would not be well approximated by the current model, for example, buildings with exhaust air heat pumps, thermal solar panels, active thermal storage, or advanced control strategies. The generality of the model applied to various building types requires further development work.

Model selection is a frequently advocated approach, for example, by Bacher and Madsen [8], Raillon and Ghiaus [10], Rouchier [43]. The strongly regulating prior used for the air infiltration parameter C_{inf} can be viewed as model selection. In the case study, there is very little evidence of air infiltration in the data (the buildings are well sheltered and located in an inland region of generally low wind exposure), and the posterior distribution of C_{inf} keeps close to the specified prior. However, the used exponentially modified Gaussian prior distribution would allow for higher C_{inf} values in case of strong enough evidence in the data.

In the case study, the internal mass element (modeling internal walls and the intermediate floors) accounts for some 79% of the total thermal mass. Similar shares would be expected for other Swedish multifamily buildings. Thus, this element is important for the thermal charging–de-charging pattern, which is essential for predicting heat demand flexibility [20] and thermal comfort assessments. The internal mass is modeled with a two-node element [4], its thermal resistance is set to $1 \text{ m}_{fl}^2 \cdot \text{K/W}$, and the thermal mass is split equally between surface and inside nodes. This part of the thermal network is likely to benefit from more detailed investigation and validation to ensure realistic thermal mass behavior.

6. Conclusions

Reliable energy models are needed to determine building energy performance. Relatively detailed energy models can be auto-generated based on 3D shape representations of existing buildings. However, parameters describing the dynamic thermal behavior of buildings are seldom available in digitized form and these parameters have to be estimated. This paper presents a Bayesian calibration framework and qualitatively evaluates results produced from a case study.

The proposed energy model is detailed, and identifying the full set of parameters with only likelihood function is not feasible. An informative prior model is needed to constrain the parameter space to ensure identifiability. The Bayesian calibration approach enables incorporation of such regulating information in a formal way. A benefit of the relatively detailed model specification is that it allows incorporation of prior knowledge on a level at which such information tends to exist, e.g., thermal transmittance per building element type rather than total heat loss. An informative prior model also allows meaningful inference in cases where there is no measurement data to calibrate with.

The results of the Gaussian approximation approach of the penalized MLE match well with the result of the HMC sampling. The MLE was on the order of 100 times faster than HMC sampling. For applications where short computation time is essential, penalized MLE is adequate. The MLE parameter estimation, in the case study, took approximately a half minute to compute on a standard laptop computer (much thanks to the automatic differentiation and fast linear algebra enabled by Stan).

DHW use is a major source of uncertainty for the targeted district-heated multifamily buildings. Incorporating hourly billing metering for domestic cold water substantially decreases this uncertainty.

The benefit would be even greater for buildings with fewer apartments than our case study with 59 apartments.

The proposed state augmentation enables a probabilistic approach for handling uncertain input data, which allows more prior knowledge to be incorporated. We provide uncertainty models for these inputs that can be utilized for both model- and measurement-based inputs.

A limitation of using measurement-based inputs is that these cannot be applied in cases lacking data or for out-of-sample predictions. A seamless utilization of both measurement-based and model-based inputs is needed to achieve a framework that can learn from all available data while still result in decent enough results in situations of less or no data availability. We also anticipate that a hierarchical modeling approach can be used to pool information between models, thus allowing energy model of buildings that lack measurement data to be informed from other building that do have reliable measurements.

Author Contributions: L.L. contributed to conceptualization, modeling, coding, analysis, and original draft preparation; J.A. contributed to conceptualization, knowledge of building physics, reviewing, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Eskilstuna Kommunfastigheter, Eskilstuna Energy and Environment, and the Knowledge Foundation (KK-stiftelsen).

Acknowledgments: This work was carried out under the auspices of the industrial post-graduate school Reesbe. We are grateful to two anonymous referees for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AHU	Air Handling Unit
DH	District Heating
DCW	Domestic Cold Water
DHW	Domestic Hot Water
DHWC	Domestic Hot Water Circulation
HMC	Hamiltonian Monte Carlo
LiDAR	Light Detection And Ranging
MLE	Maximum Likelihood Estimation
PDF	Probability Density Function

Symbols

Δt	time interval (h)
γ	azimuth angle ($^{\circ}$)
κ	areal heat capacity ($\text{Wh}/(^{\circ}\text{C} \cdot \text{m}_s^2)$)
μ	location parameter, the mean value in a normal distribution
ϕ	specific thermal power (W/m_n^2)
Ψ	linear heat loss coefficient ($\text{W}/(\text{m} \cdot \text{K})$)
σ	standard deviation
$\sigma_{\%}$	relative standard deviation
θ	centigrade temperature ($^{\circ}\text{C}$)
Θ	all parameters, including known/certain parameters
θ	set of uncertain/unknown parameters
A	System dynamics matrix, continuous time
B	Input coefficient matrix, continuous time
F	Transition matrix, discrete time
G	Input coefficient matrix, discrete time
H	Measurement model matrix, discrete time
K	Kalman gain
P	Covariance matrix
Q	Process noise covariance matrix
R	Measurement noise covariance matrix
S	Kalman filter innovation covariance
u	Input vector
w, v	White noise vector
x	State vector

y	Measurement vector
ϵ	Kalman filter innovation/error term
A	Area (m^2)
C	specific heat capacity ($Wh/(^{\circ}C \cdot m_{fl}^2)$)
F	factor
H	specific heat transfer coefficient ($W/(^{\circ}C \cdot m_{fl}^2)$)
L	length (m)
I	solar or thermal radiation (W/m_s^2)
N	number/amount
Q	specific quantity of heat (kWh/m_{fl}^2)
U	thermal transmittance ($W/(^{\circ}C \cdot m_s^2)$)
g	total solar energy transmittance (-)
h	surface coefficient of heat transfer ($W/(^{\circ}C \cdot m_s^2)$)
ln	natural logarithmic
n	exponent
qv	specific (floor area normalized) volumetric flow rate ($l/(s \cdot m_{fl}^2)$)
r	ratio (-)
u	control signal (-)

Subscripts

C, H	cooling, heating
ahu	air handling unit
bl	(window) blinds
c, ci	convective, convective interior surface
$dhw, dhwc$	domestic hot water and domestic hot water circulation
e	external (as in outdoor)
gf	ground floor
gl	glazing (windows, doors etc)
gr	ground
hor	horizontal
hyd	hydronic heating system
im	internal mass (internal walls, intermediate floors and adiabatic external walls)
inf	infiltration (uncontrolled air leakage)
int	internal (as in indoor)
k	discrete-time time index
mve	manual venting
op	operative temperature
p	pipe
rc	resistance and capacitance
r, ri	radiative, radiative interior surface
rf	roof
se, si	surface exterior, surface interior
set	set-point
sh	shading or sheltering
sky	sky temperature or sky thermal radiation
sol	solar radiation/heat gain
sup	supply
t	continuous-time time index
tb	thermal bridges
tot	total
tr	transmission
trv	thermostatic radiator valve(s)
ve	ventilation
ver	vertical
vi	virtual ground layer
wi	windows

References

1. Senave, M.; Roels, S.; Verbeke, S.; Lambie, E.; Saelens, D. Sensitivity of characterizing the heat loss coefficient through on-board monitoring: A case study analysis. *Energies* **2019**, *12*, 3322. [[CrossRef](#)]
2. Kristensen, M.H.; Hedegaard, R.E.; Petersen, S. Hierarchical calibration of archetypes for urban building energy modeling. *Energy Build.* **2018**, *175*, 219–234. [[CrossRef](#)]

3. Wang, R. 3D building modeling using images and LiDAR: A review. *Int. J. Image Data Fusion* **2013**, *4*, 273–292. [CrossRef]
4. Lundström, L.; Akander, J.; Zambrano, J. Development of a Space Heating Model Suitable for the Automated Model Generation of Existing Multifamily Buildings—A Case Study in Nordic Climate. *Energies* **2019**, *12*, 485. [CrossRef]
5. Fabrizio, E.; Monetti, V. Methodologies and advancements in the calibration of building energy models. *Energies* **2015**, *8*, 2548–2574. [CrossRef]
6. Lundström, L.; Dahlquist, E. Uncertainty in Hourly Readings from District Heat Billing Meters. In Proceedings of the SIMS 2019: 60th International Conference of Scandinavian Simulation Society, Västerås, Sweden, 13–16 August 2019.
7. Chong, A.; Menberg, K. Guidelines for the Bayesian calibration of building energy models. *Energy Build.* **2018**, *174*, 527–547. [CrossRef]
8. Bacher, P.; Madsen, H. Identifying suitable models for the heat dynamics of buildings. *Energy Build.* **2011**, *43*, 1511–1522. [CrossRef]
9. Coffman, A.R.; Barooah, P. Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings. *Build. Environ.* **2018**, *128*, 153–160. [CrossRef]
10. Raillon, L.; Ghiaus, C. An efficient Bayesian experimental calibration of dynamic thermal models. *Energy* **2018**, *152*, 818–833. [CrossRef]
11. Rouchier, S.; Rabouille, M.; Oberlé, P. Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling. *Build. Environ.* **2018**, *134*, 181–190. [CrossRef]
12. Tian, W.; Yang, S.; Li, Z.; Wei, S.; Pan, W.; Liu, Y. Identifying informative energy data in Bayesian calibration of building energy models. *Energy Build.* **2016**, *119*, 363–376. [CrossRef]
13. Bandera, C.F.; Ruiz, G.R. Towards a New Generation of Building Envelope Calibration. *Energies* **2017**, *10*, 2102. [CrossRef]
14. U.S. Department of Energy. *EnergyPlus Version 9.0—Engineering Reference*; U.S. Department of Energy: 2019. Available online: <https://bigladdersoftware.com/epx/docs/9-0/engineering-reference> (accessed on 20 December 2019).
15. Coakley, D.; Raftery, P.; Keane, M. A review of methods to match building energy simulation models to measured data. *Renew. Sustain. Energy Rev.* **2014**, *37*, 123–141. [CrossRef]
16. Baetens, R.; Saelens, D. Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. *J. Build. Perform. Simul.* **2016**, *9*, 431–447. [CrossRef]
17. Sjögren, J.U.; Andersson, S.; Olofsson, T. Sensitivity of the total heat loss coefficient determined by the energy signature approach to different time periods and gained energy. *Energy Build.* **2009**, *41*, 801–808. [CrossRef]
18. Hoyer-Klick, C.; Lefèvre, M.; Schroedter-Homscheidt, M.; Wald, L. *MACC-III Deliverable D57.5: USER'S GUIDE to the MACC-RAD Services on Solar Energy Radiation Resources*; Technical Report; Copernicus Atmosphere Monitoring Service (CAMS); Copernicus, European Union's Earth Observation Programme: Brussels, Belgium, 2015. [CrossRef]
19. Kim, D.W.; Park, C.S. Application of Kalman Filter for Estimating a Process Disturbance in a Building Space. *Sustainability* **2017**, *9*, 1868. [CrossRef]
20. Kensby, J.; Trüschel, A.; Dalenbäck, J.O. Potential of residential buildings as thermal energy storage in district heating systems—Results from a pilot test. *Appl. Energy* **2015**, *137*, 773–781. [CrossRef]
21. Swedenergy. *Energimätare för termisk energi—Tekniska branschkrav och råd om mätarhantering och Leverans av mätvärden (Energy Meter for Thermal Energy—Technical Industry Requirements and Advice on Meter Management and Delivery of Measured Values)*; Technical report; Swedenergy: Stockholm, Sweden, 2019.
22. Werner, S. District heating and cooling in Sweden. *Energy* **2017**, *126*, 419–429. [CrossRef]
23. ISO 52016-1:2017. *Energy Performance of Buildings—Energy Needs for Heating and Cooling, Internal Temperatures and Sensible and Latent Heat Loads—Part 1: Calculation Procedures (ISO 52016-1:2017)*; ISO: Geneva, Switzerland, 2017.
24. Carpenter, B.; Gelman, A.; Hoffman, M.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **2017**, *76*, 1–32. [CrossRef]
25. Simon, D. *Optimal State Estimation: Kalman, H, and Nonlinear Approaches*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006. [CrossRef]

26. Brown, R.; Hwang, P. *Introduction to Random Signals and Applied Kalman Filtering*, 4th ed.; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2012.
27. Havlík, J.; Straka, O. Performance evaluation of iterated extended Kalman filter with variable step-length. *J. Phys. Conf. Ser.* **2015**, *659*, 012022. [[CrossRef](#)]
28. BeBo. *Kartläggning av VVC-förluster i Flerbostadshus (Mapping of DHWC Losses in Multifamily Buildings)*; Technical Report; BeBo (Swedish Energy Agency's Network for Residential Property Owners); BeBo: Stockholm, Sweden, 2015.
29. ASHRAE. *ASHRAE Handbook: Fundamentals (SI Edition)*; ASHRAE: Atlanta, GA, USA, 2017.
30. Sveby. *Brukarindata Bostäder (Input Data for Residential Buildings)*; Technical Report; Sveby (Swedish Industry Standard for Energy Calculations in Buildings); Sveby: Stockholm, Sweden, 2012.
31. Rijal, H.; Tuohy, P.; Humphreys, M.; Nicol, J.; Samuel, A.; Clarke, J. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. *Energy Build.* **2007**, *39*, 823–836. [[CrossRef](#)]
32. Hedegaard, R.E.; Kristensen, M.H.; Pedersen, T.H.; Brun, A.; Petersen, S. Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response. *Appl. Energy* **2019**, *242*, 181–204. [[CrossRef](#)]
33. Fabi, V.; Andersen, R.V.; Corgnati, S.; Olesen, B.W. Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models. *Build. Environ.* **2012**, *58*, 188–198. [[CrossRef](#)]
34. Pereira, P.F.; Ramos, N.M.; Almeida, R.M.; Simões, M.L.; Barreira, E. Occupant influence on residential ventilation patterns in mild climate conditions. *Energy Procedia* **2017**, *132*, 837–842. [[CrossRef](#)]
35. Carlander, J.; Trygg, K.; Moshfegh, B. Integration of Measurements and Time Diaries as Complementary Measures to Improve Resolution of BES. *Energies* **2019**, *12*, 837–842. [[CrossRef](#)]
36. Widén, J.; Wäckelgård, E. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Appl. Energy* **2010**, *87*, 1880–1892. [[CrossRef](#)]
37. Fischer, D.; Wolf, T.; Scherer, J.; Wille-Hausmann, B. A stochastic bottom-up model for space heating and domestic hot water load profiles for German households. *Energy Build.* **2016**, *124*, 120–128. [[CrossRef](#)]
38. Sandberg, E. *Metodrapport—Underlag för Kriteriedokument (Method Report—Basis for Criteria Documents)*; Technical Report; Swedish Energy Agency/FEBY: Stockholm, Sweden, 2009.
39. Lundström, L. Mesoscale Climate Datasets for Building Modelling and Simulation. In Proceedings of the CLIMA 2016: 12th REHVA World Congress, Aalborg, Denmark, 22–25 May 2016.
40. CAMS. *User's Guide to the CAMS Radiation Service (CRS)*; Technical Report; Copernicus Atmosphere Monitoring Service (CAMS); Copernicus, European Union's Earth Observation Programme: Brussels, Belgium, 2018.
41. ISO 52010:2017. *Energy Performance of Buildings — External Climatic Conditions — Part 1: Conversion of Climatic Data for Energy Calculations (EN ISO 52010:2017)*; ISO: Geneva, Switzerland, 2017.
42. CAMS. *Supplementary Services; Evaluation and Quality Assurance Reports*; Technical Report; Copernicus Atmosphere Monitoring Service (CAMS). Copernicus, European Union's Earth Observation Programme: Brussels, Belgium, 2018. Available online: <https://atmosphere.copernicus.eu/supplementary-services> (accessed on 1 December 2019).
43. Rouchier, S. Solving inverse problems in building physics: An overview of guidelines for a careful and optimal use of data. *Energy Build.* **2018**, *166*, 178–195. [[CrossRef](#)]

