Estimating temperatures in a district heating network using smart meter data

Bergsteinsson, Hjörleifur G.; Vetter, Phillip B.; Møller, Jan Kloppenborg; Madsen, Henrik

Published in:
Energy Conversion and Management

Link to article, DOI:
10.1016/j.enconman.2022.116113

Publication date:
2022

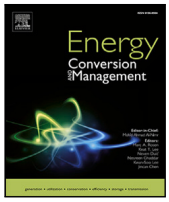
Document Version
Publisher's PDF, also known as Version of record

Energy Conversion and Management 269 (2022) 116113

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman



Estimating temperatures in a district heating network using smart meter data

Hjörleifur G. Bergsteinsson [a,*], Phillip B. Vetter [a], Jan Kloppenborg Møller [a], Henrik Madsen [a,b]

[a] Technical University of Denmark, Denmark
[b] Norwegian University of Science and Technology, Norway

ARTICLE INFO

Keywords:
Temperature optimisation
Estimating network temperature
Grey-box modelling
Kalman filter
Automatic differentiation

ABSTRACT

Smart meters at consumers create opportunities to improve operation of the district heating sector using data-driven methods. Information from these meter measurements carries the potential to increase the energy efficiency of both individual houses and the utility network, for example by identifying buildings with too high return temperature, or by detecting leakage in the network. This paper proposes a method for using meter data to estimate network temperatures. Network temperatures can subsequently be used to estimate the network characteristics, namely the nonlinear relationship between network temperature and the plants' temperature and flow. A description of the network characteristics is needed for most temperature-optimisation methods to keep the supply temperature as low as possible without violating the system constraints. Traditionally, measurement wells located in the network have been used. These wells are located at critical points in the network where the largest temperature losses occur. Since the lowest temperature often varies over time, multiple critical points are necessary. The method presented in this paper eliminates the need for these physical critical points in the network. It also makes it possible to change the location of the critical points if needed. The network temperature is estimated using a stochastic state–space model of the heat dynamics from the street level distribution pipe over the service pipe and into individual houses. The parameters in the model are estimated using a maximum likelihood approach, and the Kalman Filter is used to evaluate the likelihood function. The estimation process takes advantage of automatic differentiation using the R package Template Model Builder (TMB) to reduce the computational workload. The proposed method is validated by comparing the estimated temperature with the temperature measured from a measurement well.

1. Introduction

Future energy systems need to be flexible because of increasing shares of renewable energy sources that are typically intermittent due to their direct weather dependency. Furthermore, new regulation continues to emerge and awareness of greenhouse gas emissions is increasing, with a consequential transition to more renewable energy. For these reasons, more sophisticated methods are needed to deliver the required energy demand without using fossil fuels. Energy systems for gas, heat and electricity need to be integrated to phase out fossil fuels. Because of their unique capability to store energy, district heating systems play a key role in the transition towards more flexible energy systems [1]. An example is that wind power can be used for heating water to be delivered either directly or stored when the electricity demand is lower than the generated wind power. Hence, in order to fully maximise the flexibility potential of energy systems, they must be integrated. In addition, each energy system needs to perform efficiently to deliver an optimal energy integration. Mathiesen et al. [2] gives an extensive discussion on the integration of all energy systems to increase the flexibility of the system and reach 100% renewable energy supply.

The work presented here focuses on methods for optimising the temperature control in district heating systems based on information from smart meters installed at consumer level. Today, smart meters that can take very frequent readings are installed in many buildings. This paper suggests methods to take advantage of frequent readings by smart meters.

The role of district heating systems is to meet consumer heat demand while simultaneously minimising both production and operation costs. Optimisation of production planning is concerned with scheduling heating unit operation in order to produce the desired heat demand at the lowest cost [3]. In Denmark, heat is often co-generated (combined heat and power (CHP) production). CHP units typically run during periods with high electricity prices. Any heat that is not used during these periods is stored in the system. During low electricity price periods, district heating operators either use this stored heat or run heat-only units, e.g. gas boilers, heat pumps or solar heat [4].

* Correspondence to: Anker Engelunds vej 1, Building 101A, 2800 Kongens Lyngby, Denmark.
E-mail address: hgbe@dtu.dk (H.G. Bergsteinsson).

https://doi.org/10.1016/j.enconman.2022.116113
Received 4 May 2022; Received in revised form 6 August 2022; Accepted 8 August 2022
Available online 31 August 2022

Nomenclature

Abbreviations

SDE	Stochastic Differential Equation
TMB	Template Model Builder
CHP	Combined Heat and Power
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
GDPR	General Data Protection Regulation

Mathematical notation

ω	Wiener process
σ	Diffusion coefficient
V	Observation covariance matrix
e	Innovation
S	Logistics function
L	Likelihood function
$\mathcal{L}$	Log-likelihood function
U	Random effects
θ	Fixed effects
f	Drift term
g	Diffusion term
h	Observation function
x	System states
u	External inputs
$\mathcal{N}$	Normal distribution
K	Kalman gain
P	State covariance matrix
Σ	Observation covariance matrix

Physical parameters

Q	Mass flow rate [kg/s]
ρ	Mass density [kg/m^3]
c_v	Specific heat capacity [kJ/(kg K)]
A	Area [m^2]
v	Flow velocity [m/s]
p	Absolute pressure [Pa]
f_D	Darcy friction coefficient [–]
S	Circumference [m]
k	Thermal conductivity [W/(m K)]
T	Temperature [°C]
$\dot{q}_e$	Heat loss per meter [W/m]
C	Heat capacity per meter [J/m K]
λ_p	Thermal conductivity [W/(m K)]
τ	Time constant [s]
d_o	Outer diameter [m]
L	Length of pipe [m]
R	Thermal resistance per meter [(K m)/W]
x_a	Insulation thickness [m]

Subscripts

t	Time
g	Ground

Superscripts

s	Street
(i)	House index

Hence, the plant's objective is to generate as much power as possible during periods with high electricity prices while satisfying the heating demand. This is achieved by keeping the supply temperature as low as possible. Consequently, lowering the temperature will also reduce heat loss in the network and production cost [5]. Hence, it is crucial to lower the supply temperature in order to optimise the operation of the entire district heating system.

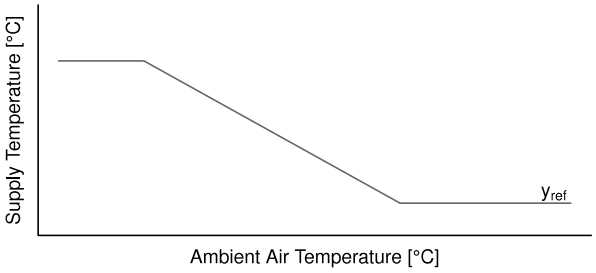


Fig. 1. Reference curve, y_{ref} shows the desired supply temperature of hot water for given ambient air temperature.

A lower temperature in the network will translate to lower production costs as the operation of most power producing devices can be more efficient. For example, in a CHP plant, lowering the temperature results in increase in ratio of power to heat output, and electricity is more valuable than heat [6]. Also, if the utility uses several energy sources then lowering the temperature will also increase the flexibility of utilising the optimal energy sources at given time-point due to the different limitation of each energy sources. Thus, this makes it possible to utilise heat from new sources like excess heat from comfort cooling. Without a lowering of the temperature these sources which would otherwise have been disregarded due to too low temperatures for entering the network. Likewise, it will be more efficient to operate heat pumps in the network. This increased efficiency will result in better investment feasibility for heat pumps. Heat pumps are likely to play a bigger role in supplying heat to the network as they can utilise heat sources with a low-temperature range, e.g. wastewater, ambient water, industrial heat waste, and solar heat storage [7]. Therefore, for optimal operation of a district heating network with multiple heat sources (e.g. heat pumps and solar heat), temperature optimisation is needed in order to keep the supply temperature as low as possible while increasing both the efficiency and feasibility of heat sources. The complexity of the dynamics of district heating networks implies that data-driven models for temperature optimisation are needed to lower the supply temperature in the network.

Traditionally, the optimisation of supply temperature has been controlled using a reference curve based on the ambient air temperature, as indicated in Fig. 1 [8]. The reference curve dictates the minimum allowed supply temperature as a function of the ambient air temperature. The maximum at the left endpoint of the curve is related to the physical limitation of the system, and the minimum temperature to the right ensures a high enough temperature for domestic hot water usage without risking the formation of bacteria, e.g. legionella. However, this control scheme is a conservative estimate since it only considers one variable to ensure that the heat supply is sufficient at any time, and this often leads to an unnecessarily high supply temperature. Obviously, heat demand is not only dependent on the ambient air temperature but also by other climate variables, e.g. solar radiation and wind. The weather effect on heat demand is also not instantaneous, as buildings are known to have slow thermal reactions. The slow thermal reaction in a single building is demonstrated in Madsen and Holst [9]. In Nielsen and Madsen [10] it is shown how the physical knowledge about heat

consumption in buildings can be used to establish a heat consumption model for the district heating network. Preferably, the controller should include heat demand forecasts and the network characteristics of the system (transportation time, heat loss).

Madsen et al. [5] propose a control strategy for the supply temperature at the plant by utilising heat demand forecasts and network characteristics. The strategy estimates future set-points of the supply temperature at the production plant by using a model for the network characteristics determined from receiving feedback of the network temperatures, and the approach allows the system to adapt to changes in the system. This stochastic time-varying system is both nonlinear and non-stationary, and consequently, a nonlinear and time-varying transfer function was proposed in order to model the relationship between the supply temperature and flow at the plant and the network temperature from a critical point in the network. This strategy has been demonstrated to lower the supply temperature at the district heating plant and thereby reduce production costs and the heat losses in the system [11].

This control strategy requires feedback from the system. Usually, this information is measured at a selected number of so-called critical points in the network at street level using sensors in measurement wells. A critical point in the network is typically located close to the end-users showing the largest temperature losses in transportation from the plant. A single critical point is sub-optimal, since the location of the critical points in the network can change over time as a consequence of the diurnal pattern and consequently several (e.g. five) critical points are normally used. However, networks change, e.g. pipes get older, pipes are replaced, new areas are connected to the network, buildings are refurbished and so forth, and hence it would be advantageous if system feedback could be received at any particular location in the network. This is possible by using data from smart meters installed at consumers, to establish temperature feedback. This approach furthermore eliminates the need to install temperature sensors in measurement wells. For a number of years, smart meters have been installed, allowing consumers to link their consumption to bills from the district heating utility. There is a requirement from the European Union that all buildings have individual energy meters where feasible, including heat meters for houses connected to district heating networks [12].

Smart meters create new opportunities to develop data-intelligent methods for district heating operations. This digital transformation has fostered research related to district heating at consumers. Data from smart meters can be used to give valuable insight into network performance and building energy efficiency through investigations of possible leakage in the network or insufficient cooling at consumers. For example, Kristensen and Petersen [13] use smart meter data to derive three heating efficiency indicators of buildings to compare the energy performance of the buildings, and also give insights on district heating and smart meters in Denmark. However, the feedback of the network temperature needs to be robust, and using measurements from smart meters from single-family houses without a quality check could give a wrong representation of the network characteristics. Bergsteinsson et al. [14] propose a simple method to estimate the network temperature by resampling and aggregating data from a group of smart meters. Based on this data, an artificial network temperature is estimated using time-wise quantile estimation at each time step. However, the method is quite naive and is not robust when the measurements are either of bad quality or they lack extended periods of time. A more advanced model is therefore required to give more accurate and reliable continuous feedback of the network. Hence, a model derived from physical knowledge of the system is needed which uses meter measurements to estimate the parameters of the network, i.e. grey-box modelling.

Hence, by establishing a robust method to estimate the network temperature at multiple points in the network by using smart meter data will enable feedback options which again makes it possible for utilities to implement controllers for temperature optimisation. Data-driven temperature optimisation lowers supply temperature in the network, thus decreasing the production cost and reducing heat losses in the network.

1.1. Pipe dynamics and grey-box modelling in district heating networks

Thermodynamics modelling of hot water pipes in district heating has been studied extensively, both to provide a deeper understanding of the dynamics and to obtain information that can be used to minimise cost: e.g. by selecting the optimal size of pipes or by reducing heat losses in the network. For planned new district heating systems, models of the network can be used to simulate scenarios for the design of new pipelines and to analyse the hydraulic and thermal behaviour for the purpose of minimising the costs of establishing the new networks and to decide on their operation. For already established systems, the methods can be used to simulate different scenarios, e.g. peak loads or to identify locations where maintenance is needed. Hence, adequate physical models of district heating networks are important for efficient operation. There are studies that have proposed different methods for modelling the thermodynamics in hot water pipes for operational purposes. The most widespread approach is a finite element method in which the pipe is divided into infinitesimal segments in order to solve a governing partial differential equation for computations of the temperature difference and heat loss in the pipes.

This can be solved using finite volume schemes, e.g. in Vandermeulen [15] the first-order upwind finite volume model is used to study district heating network flexibility by storing heat in the network by altering the supply temperature. Other studies have proposed using a finite element method to simulate the thermodynamics of having the temperature distribution from plant to end-user for operation during changes in the system [16]. Benonysson et al. [17] propose a node method and derive a mathematical model of the pipes using heat transfer equations to compute the outlet temperature of the pipe when the inlet temperature is measured. Søgaard [18] propose modelling the network as a dynamic input–output system, describing the network response characteristics between measurements from the plant and a point in the network. Hence, different possibilities have been proposed in the literature to describe the network dynamics, from white-box methods entirely based on physics to black-box models based solely on measurements. There is no globally optimal method to describe the dynamics, as each method has its pros and cons. White-box modelling is suitable for network design and flexibility/peak-shaving simulations. White-box methods are usually computationally heavy and require human maintenance to validate and select appropriated values, while black-box methods are fast and usually need no maintenance. Therefore, black-box methods are better suited for control applications where the computation time needs to be low such that the control can respond quickly to changes, as proposed in Madsen et al. [19]. An extensive summary of hot water pipe dynamics is given in van der Heijde [20] and Vandermeulen [15].

The method proposed in this paper, will take advantage of this physical knowledge of pipe thermodynamics to estimate street-level network temperature (critical point) using smart-meter measurements. This will be done using statistical methods to estimate the parameters of equations that are derived from physics, i.e. by using a grey-box modelling approach. Grey-box models bridge the gap between physical and statistical modelling and are frequently studied in literature and have shown promising results for parameter estimation and control purposes. Madsen and Holst [9] demonstrates how grey-box models can be used to describe the dynamics of the indoor air temperature in buildings and its dependency on weather and heat input. The forecasting ability of the method has proved to have high accuracy. Bacher and Madsen [21] describe an approach for using data for optimal model selection of grey-box models and for estimating model parameters for a particular building. They also discuss grey-box model applications for validating the energy performance of buildings, energy consumption forecasting and indoor climate control. Thilker et al. [22] propose a grey-box model to describe the heating dynamics of an old school building, and Thilker et al. [23] demonstrates the potential of using the grey-box model to control the return temperature to lower the operational cost of the building.

1.2. Contribution of the paper and overview

The main contribution of this work is the formulation and application of a set of partially observed stochastic differential equations to be used for inference on the temperature of the distribution network at street level based on smart-meter readings from individual consumers. The system of stochastic differential equations SDEs is derived from partial differential equation PDE that describe the heat transfer dynamics from the distribution pipe into single-family houses over a service pipe. The estimated network temperature can be used to gain information about the network response characteristics; i.e. how the network reacts to changes in temperature or flow at the plant. This information can be used as feedback for temperature control, for example. Thus, this result makes physical measurement wells redundant, as a group of houses with smart meters can be used to estimate the network temperature.

The paper is organised as follows. Section 2 presents the reasoning behind the chosen SDE formulation and explains the chosen estimation method used for both parameter and state estimation. Section 3 presents the result from applying the methods to the presented data, including parameter interpretation. Finally, Sections 4 and 5 discuss the presented results and draw some general conclusions.

2. Methods

In this section, a model will be established which describes the heat transfer between the distribution pipe and houses through a service pipe. The model is derived from physics and takes the form of a partial differential equation which is subsequently approximated by a stochastic differential equation for continuous temperature estimates in order to incorporate the information from smart meters. The heat dynamics is affected by the thermodynamic properties of the pipes and their relationship to the surroundings.

2.1. Stochastic differential equations

The models consider in this paper, will be on the following continuous–discrete time stochastic state space form [24],

$$\mathrm{d}x_t = f(x_t, u_t, \theta)\mathrm{d}t + g(\theta)\mathrm{d}\omega_t \tag{1}$$

$$y_{tk} = h(x_{tk}) + e_{tk} \tag{2}$$

where the system state x_t evolves in continuous time as determined by the drift f and the diffusion g. The observation function h relates measurements to system states. The drift function depends on the state itself, external inputs u_t and the system parameters θ. The drift term, accounts for most of the known phenomena of the system and draws on physical knowledge, while the diffusion relates to unaccounted for and unknown system drivers as well as noise. The aim is to describe these effects by the random perturbations imposed by the Wiener process ω_t whose non-overlapping increments are independent and Gaussian-distributed i.e. $\omega_t - \omega_s \sim N(0, t-s)$. The (hidden) state x_t is observed (indirectly) through the measurements y_{tk} which become available at certain discrete times $t = t_k$. In this paper, it will be assumed that a subset of the states (the individual smart meters) are directly observed, although contaminated by Gaussian noise e_{tk}.

2.2. Pipe modelling

The thermodynamics modelling of the processes inside a pipe is rather complicated, and consequently a number of assumptions are introduced to approach the problem and make it feasible. The water in the pipes will be assumed to be incompressible, the pipe is a grounded, insulated single pipe (which implies no influence from the return pipe), and that the system is in a steady-state. It will also be assumed that the ground temperature is constant throughout each month using the information from Grunnet Wang et al. [25].

van der Heijde et al. [26] uses a dynamic thermo-hydraulic pipe model for district energy systems for the purpose of creating dynamic simulations of the temperature in district heating networks and cooling pipe systems. The pipe model is a partial differential equation describing the temporal evolution (t) of the energy across the axial dimension (x) in the pipe. The PDE describes the heat transfer and the associated heat losses to the surroundings through a combination of the energy and the continuity equation. The equation is

$$\underbrace{\frac{\partial(\rho c_v T A)}{\partial t}}_{\substack{\text{time}\\ \text{derivative}}} + \underbrace{\frac{\partial(\rho v(c_v T + p/\rho)A)}{\partial x}}_{\substack{\text{spatial}\\ \text{derivative}}} = \underbrace{vA\frac{\partial p}{\partial x}}_{\substack{\text{pressure}\\ \text{difference}\\ \text{energy}}} + \underbrace{\frac{1}{2}\rho v^2|v| f_D S}_{\substack{\text{wall friction}\\ \text{dissipation}}} + \underbrace{\frac{\partial}{\partial x}(kA\frac{\partial T}{\partial x})}_{\substack{\text{axial heat}\\ \text{diffusion}}} - \underbrace{\dot{q}_e}_{\substack{\text{heat}\\ \text{loss}}}, \tag{3}$$

where ρ [kg/m^3] is the mass density of the fluid in the pipe, c_v [kJ/(kg K)] is the specific heat of the fluid in the pipe, A [m^2] is the cross sectional area of the pipe, v [m/s] is the flow velocity, p [kg/m^3] is the absolute pressure, f_D [−] is the Darcy friction coefficient, S [m] is the pipe circumference, k [W/(m K)] is the thermal conductivity, T [°C] is the temperature inside the pipe, and $\dot{q}_e$ [W/m] is the heat loss per unit length.

van der Heijde et al. [27] and Vandermeulen [15] argue that most of the terms in Eq. (3) can be assumed to be negligible, and with the additional assumption that the water is incompressible the equation simplifies to

$$\frac{\partial(\rho c_v A T)}{\partial t} + \frac{\partial(\rho c_v A v T)}{\partial x} = -\dot{q}_e. \tag{4}$$

The equation remains a PDE which describes the heat transfer through a pipe in the form of an advection equation with the loss term $-\dot{q}_e$. The equation can be solved analytically when assuming steady-state operation, but it is a challenging task. Hence, it is usually solved instead by using finite volume methods, where the pipe is split into multiple smaller sections. The solution is then obtained by integrating across each of these sections. For further details, see van der Heijde et al. [26], Dénarié et al. [28], and Grosswindhager et al. [29].

Eq. (4) will be used to estimate the supply temperature in the distribution pipe in the street before entering the service pipe into the house. This is illustrated in Fig. 2, where the street hot water $T_t^{(s)}$ enters the service pipe and the temperature, $T_t^{(i),obs}$ is measured by the smart meters after travelling over the service pipe. The temperature loss in the system is assumed to be caused by heat loss to the surroundings. The flow is assumed to be constant through the service pipe. The only information that is known is the measurements from the smart meters and the assumed constant ground temperature for each month using the average temperature given by Grunnet Wang et al. [25]. The goal is to use only the smart meter data to estimate the street temperature. The data-driven model is a reformulation of Eq. (4) using a stochastic differential equation as shown in Section 2.1.

The first step is to transform Eq. (4) into a standard Resistance–Capacitance (RC) form. First, the heat capacity over the service pipe per unit length of the water is defined as $C = Ac_v\rho$ [J/m K]. Next, it will be assumed that the mass flow is constant through the service pipe, $Q = \rho A v$ [kg/s]. The final assumption for this model derivation is that the heat loss is proportional to the temperature difference between the water and ground with the proportionality constant being the inverse thermal resistance (between pipe and ground) as shown in Wallentén [30]. Hence, Eq. (4) becomes,

$$C\frac{\partial T}{\partial t} + \frac{\partial(c_v Q T)}{\partial x} = \frac{T_g - T^{(i)}}{R}. \tag{5}$$

The next step is to discretise the equation in space

$$\frac{\partial(c_{\mathrm{v}} Q T)}{\partial x} \underbrace{\longrightarrow}_{\text{discretisation}} \frac{c_{\mathrm{v}} Q \Delta T}{\Delta x}. \tag{6}$$

Thus, the temperature difference is that between the street and house $\Delta T = T^{(\mathrm{s})} - T^{(\mathrm{i})}$ over the service pipe length Δx. The length of the pipe is then multiplied through the equation, which redefines the heat capacity as $C = C\Delta x$ and the thermal resistance as $R = R/\Delta x$. The system now becomes

$$\frac{\partial T_{\mathrm{t}}}{\partial t} = C^{-1}\left(c_{\mathrm{v}} Q_t (T_t^{(\mathrm{s})} - T_t) - R^{-1}(T_t - T_{\mathrm{g}})\right). \tag{7}$$

This differential equation describes how the house temperatures change with time. It is now straight-forward to formulate the proposed stochastic differential equation (for one house)

$$\mathrm{d}T_{\mathrm{t}} = C^{-1}\left(c_{\mathrm{v}} Q_t (T_t^{(\mathrm{s})} - T_t) - R^{-1}(T_t - T_{\mathrm{g}})\right)\mathrm{d}t + \sigma \mathrm{d}\omega_t, \tag{8}$$

where σ is the diffusion coefficient and ω_t is a standard Wiener process, i.e. the source of the noise in the system. Hence, we have established a time-dependent stochastic state space model to describe the energy exchange between the street and individual houses. Jointly with the observation equation, this type of model is also referred to as a grey-box model in the literature, see for instance [9].

As already mentioned, the only information available is the smart-meter data, and the objective of this study is then to estimate the supply temperature at the distribution pipe in the street before the water enters the houses. To clarify; the information available is from multiple smart-meter measurements installed in houses connected to the same street, but there is no information on the street temperature. The street temperature must in principle be greater or equal to the highest observed house temperature, although estimated values of it can be lower when accounting for measurement uncertainty. The actual behaviour of the street temperature is determined by the plant output and the network characteristics. However, in order to keep the specifications simple, these detailed descriptions are considered to be outside the scope of the current model, and, in the absence of any "real" drivers for the street temperature, its dynamics are simply modelled by a random walk. The model for the forwarded water temperature at the individual houses $T_t^{(i)}$ is that presented in Eq. (4). These equations are combined to obtain the considered system of equations

$$\mathrm{d}T_t^{(\mathrm{i})} = C_i^{-1}\left(c_{\mathrm{v}} Q_t^{(\mathrm{i})}(T_t^{(\mathrm{s})} - T_t^{(\mathrm{i})}) - R_i^{-1}(T_t^{(\mathrm{i})} - T_t^{(\mathrm{g})})\right)\mathrm{d}t + \sigma_{\mathrm{i}}\mathrm{d}\omega_t^{(\mathrm{i})}, \tag{9a}$$

$$\mathrm{d}T_t^{(\mathrm{s})} = \sigma_{\mathrm{s}}\mathrm{d}\omega_t^{(\mathrm{s})}, \tag{9b}$$

which is illustrated in Fig. 2. The model takes as inputs the mass flow rate $Q_t^{(\mathrm{i})}$ [kg/s], and the ground temperature T_{g} [°C]. The constant $c_{\mathrm{v}} \approx 4.186$ [kJ/(kg K)] is the specific heat capacity of water. In this formulation, three parameters are associated with each house — namely the thermal capacity C_{i} [J/K], the thermal resistance R_{i} [K/W], and the diffusion scaling σ_{i} [°C/$\sqrt{\mathrm{s}}$]. The individual house temperatures $T_t^{(\mathrm{i})}$ are observed directly by the measurement device $T_t^{(\mathrm{i,obs})}$, but it is assumed that the uncertainty of these measurements can be approximated by Gaussian noise, which gives rise to the following observation equation

$$T_t^{(\mathrm{i,obs})} = T_t^{(\mathrm{i})} + e_t^{(\mathrm{i})}, \quad e_t^{(\mathrm{i})} \sim N(0, V_t^i), \tag{10}$$

In order to establish a flow-dependent variance construction, the logistics function S is used

$$V_t^i = \mathbb{V}\left[e_t^{(\mathrm{i})}\right] = \sigma_{\mathrm{obs}}^2 + S(-Q_{\mathrm{i}}(t) + b) = \sigma_{\mathrm{obs}}^2 + \frac{K}{1 + \mathrm{e}^{a(Q_{\mathrm{i}}(t) - b)}}, \tag{11}$$

which acts as a smooth approximation to the Heaviside step function $H(Q) = \mathbf{1}_{Q<b}$, as illustrated in Fig. 3. The purpose of this variance construction is to decrease the weight of observations that are gathered under low flow conditions, where the observations obviously contain very limited information about the street temperature — as is illustrated in Figs. 6 and 7. This construction was seen to be necessary

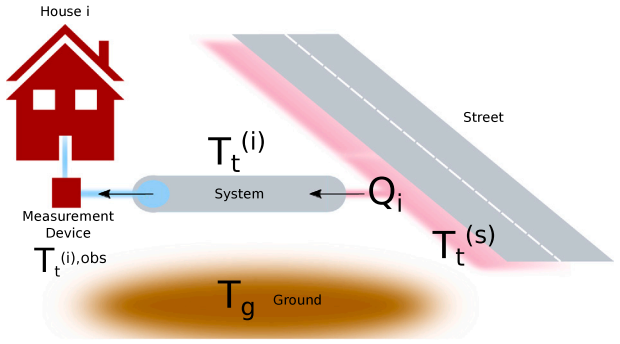


Fig. 2. A drawing that demonstrates the interaction between street, house, and ground with labels for states, observations, and inputs. The modelled system is that of the proposed model and it is centred around the pipe. The temperature of the water travelling through the pipe is the result of mixing the already present water of temperature $T_t^{(i)}$ with that of the street of temperature $T_t^{(s)}$, while heat is transferred through the pipe to the surrounding colder ground proportional to the temperature difference $T_t^{(i)} - T_g$.

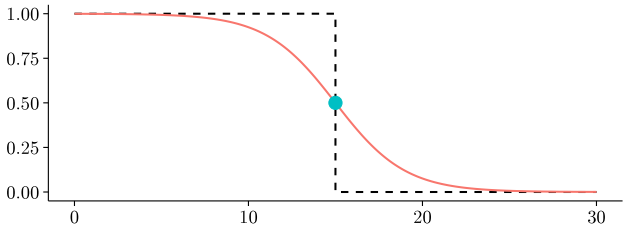


Fig. 3. A plot of the logistics function in (11) with parameters $K = 1$, $a = 0.5$ and $b = 15$ (as used in this work) which approximates to a step-function at b. The variance increases towards K for $Q \to 0$ and vanishes for $Q \to \infty$.

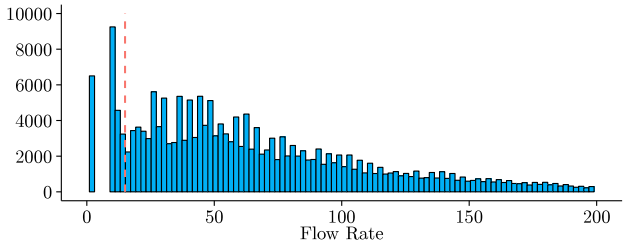


Fig. 4. A histogram of the flow rates for a specific house which illustrates the location of the cut-off level as determined by the value of b in (11).

in order to prevent the street temperature estimates from decreasing according to the decreasing house temperatures during such periods of low flow, i.e. no energy consumption. The parameter b in Eq. (11) was chosen based on inspection of flow rate histograms for the buildings in that neighbourhood. An individual neighbourhood assessment must be performed, since the lower flow rate threshold at which rapid temperature decreases occur varies. A typical bi-modal flow histogram from one particular house is shown in Fig. 4. The red dashed line marks the threshold value of $b = 15$ below which a given observation whose flow satisfies $Q \leq b$ will have its variance increased. The magnitude of this increase lies in the range $k \in [K/2, K)$. Fig. 5 demonstrates the correlation between temperature and flow rate, by marking with black dots the observations with flow rates below the set threshold. In this study, the parameters of the logistics function will be set to; $K = 10\,000$ (maximum value), $a = 0.25$ (steepness of the curve), and $b = 15$ (midpoint of the function).

The house-specific thermal parameters C and R summarise the thermal properties of the transportation system between street and house. The *system* here should be interpreted as an idealisation of the

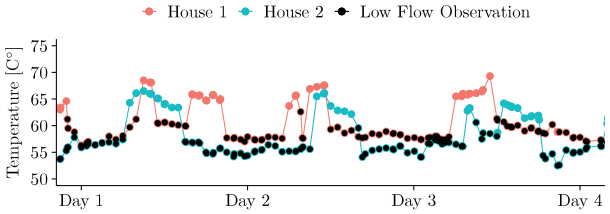


Fig. 5. Observations of temperatures for two houses with indication (black dots) of the observations where the flow rate is less than $Q \leq 20L/h$ in correspondence with the cut-off level seen in Fig. 4 and the associated curve in Fig. 3.

entire pipe installations and the different environments that those are embedded in. The underlying assumption that the potentially huge variety of such installations can be described by the simple model in Eqs. (9a)–(9b) enables a crude way of analysing houses for potential issues, like detecting poorly insulated pipes or potential leakages.

2.3. Parameter estimation

The parameter estimation is carried out using the maximum likelihood approach. In this paper, two slightly different computational methods are applied, both of which rely on the R-package Template Model Builder (TMB) [31]. The first method applies the intended methodology of TMB which is that of a generalised mixed-effects model, while the latter method employs the discrete Kalman filter. The filter is also implemented with TMB simply to draw on the package's employment of automatic differentiation which yields the gradient (and also the hessian with respect to the parameters) of the likelihood function which drastically reduces computation times. A comparison between these methods will be presented here in terms of speed, parameter estimates and uncertainty.

In the mixed effects formulation, TMB employs the Laplace approximation to integrate out the states Eqs. (9a)–(9b) which are considered random effects in this framework. The TMB procedure; the marginal likelihood $L(\theta)$ of the fixed effects θ by integrating out the random effects U of the joint likelihood $L(U,\theta)$ using the Laplace approximation. The marginal likelihood is

$$L(\theta) = \int L(U,\theta)\mathrm{d}U, \tag{12}$$

and the log-likelihood $\mathcal{L}$ is therefore

$$\mathcal{L}(\theta) = \log \int \exp \mathcal{L}(U,\theta)\mathrm{d}U. \tag{13}$$

A second-order Taylor expansion around the random effects maximum $\hat{U}$

$$\hat{U} = \arg\max_{U} \mathcal{L}(U,\theta), \tag{14}$$

of the joint log-likelihood $\mathcal{L}$ is

$$\mathcal{L}(U,\theta) \approx \mathcal{L}(\hat{U},\theta) + \frac{1}{2}(U-\hat{U})H(\hat{U},\theta)(U-\hat{U})^T. \tag{15}$$

where $H(\hat{U},\theta)$ is the hessian. The integral in Eq. (13) can then be approximated by inserting Eq. (15), which evaluates to unity (after having corrected for the missing factor of the normal distribution), since the integrand is then a multivariate normal density with mean $\hat{U}$ and covariance matrix H. The approximation is only exact if the joint distribution is also Gaussian. The marginal likelihood can now be computed as

$$\mathcal{L}(\theta) \approx \mathcal{L}(\theta,\hat{U}) + \frac{(M+1)N}{2}\log 2\pi - \frac{1}{2}\log\det H(\hat{U},\theta), \tag{16}$$

where M is the number of smart meters and N is the number of observations. For additional information about the Laplace approximation and its properties, see e.g. Madsen and Thyregod [32].

In practice, when employing the TMB framework, the user writes the negative log-likelihood as a C++ file which is then compiled and imported as a function into R. In the present case, the likelihood contributions from a stochastic differential equation system come from (1) the (hidden) state transitions and (2) the state observations. The former contribution is given by

$$U_\mathrm{t} \sim \mathcal{N}(\mathcal{M}_\mathrm{t}, P_\mathrm{t}), \tag{17}$$

where $U_\mathrm{t} = \left[T_t^{(1)} \quad T_t^{(2)} \quad \dots \quad T_t^{(\mathrm{M})} \quad T_t^{(\mathrm{s})}\right]^T$ is the random effects state vector with mean $\mathcal{M}_\mathrm{t}$ and covariance P_t (see Eq. (30)) which directly depend on $U_{\mathrm{t}-1}$ through the one-step transition density. The contribution from the observations is given by

$$Y_\mathrm{t} \sim \mathcal{N}(U_\mathrm{t}, V_t), \tag{18}$$

where V_t is given by Eq. (11).

In the case of the Kalman filter there is only one contribution to the likelihood, due to the state updating scheme. The posterior state $X_{\mathrm{t}|\mathrm{t}}$ and covariance $P_{\mathrm{t}|\mathrm{t}}$ estimates are obtained from updating the prior estimates $X_{\mathrm{t}|\mathrm{t}-1}$ and $P_{\mathrm{t}|\mathrm{t}-1}$ once new information Y_t becomes available. The updating scheme is

$$X_{\mathrm{t}|\mathrm{t}} = X_{\mathrm{t}|\mathrm{t}-1} + K_\mathrm{t}e_\mathrm{t}, \tag{19}$$

$$P_{\mathrm{t}|\mathrm{t}} = (I - K_\mathrm{t}H)P_{\mathrm{t}|\mathrm{t}-1}(I - K_\mathrm{t}H)^T + K_\mathrm{t}V_tK_t^\mathrm{T}, \tag{20}$$

where K_t is the associated Kalman gain given by

$$K_\mathrm{t} = P_{\mathrm{t}|\mathrm{t}-1}H^T{\Sigma_{\mathrm{Y,t}}}^{-1}, \tag{21}$$

$$\Sigma_{\mathrm{Y,t}} = HP_{\mathrm{t}|\mathrm{t}-1}H^T + V_t, \tag{22}$$

with innovation

$$e_\mathrm{t} = Y_\mathrm{k} - h(X_{\mathrm{k}|\mathrm{k}-1}), \tag{23}$$

and with $H = \dfrac{\mathrm{d}h(u)}{\mathrm{d}u}$. The likelihood contribution arises from the innovation e_t and the covariance matrix $\Sigma_{\mathrm{Y,t}}$ as

$$\mathcal{L}(\theta)_\mathrm{t} = \frac{1}{2}\left[\log\det\Sigma_{\mathrm{Y,t}} + d_\mathrm{t}\log 2\pi + e_\mathrm{t}\Sigma_{\mathrm{Y,t}}e_\mathrm{t}^T\right], \tag{24}$$

where d_t is the number of available observations at time t.

2.4. Theoretical Resistance–Capacitance parameters in pipe

The theoretical values of the Resistance–Capacitance (RC) model of a pipe can be computed based on physical knowledge of the pipe. The thermal capacitance of water in the service pipe is computed as

$$C = Ac_\mathrm{v}\rho. \tag{25}$$

The pipes from the distribution pipe to the substation are usually DN25 pipes. If it is assumed that the inner diameter of the pipe is 0.0273 [m] then $C = 2.45$ [kJ/m K].

The resistance of the pipe is

$$R = \frac{1}{2\pi\lambda_\mathrm{p}}\ln\frac{d_\mathrm{o}+2x_\mathrm{a}}{d_\mathrm{i}}, \tag{26}$$

with the thermal conductivity of insulation of the pipe as, $\lambda_\mathrm{p} = 0.028$ [W/m K], outer diameter $d_\mathrm{o} = 0.0337$ [m], and insulation thickness $x_\mathrm{a} = 0.0182$ [m], which gives a resistance of $R = 4.16$ [m K/W].

The theoretical estimates of the RC parameters will be used to compare them to the estimated RC parameters.

2.5. First and second-order moments

The likelihood contributions of the stochastic differential equation require computing the expectation and variance of the one-step predictions regardless of the use of the two methods. In the case of TMB, this amounts to computing $\mathcal{M}_\mathrm{t}$ and P_t and similarly for the Kalman filter $U_{\mathrm{t}|\mathrm{t}-1}$ and $P_{\mathrm{t}|\mathrm{t}-1}$. This will generally require integrating the first and

second-order moments of the SDE forward in time, and while that is possible using standard ordinary differential equation (ODE) solvers, such an approach introduces both the choice of integral method and a time-step Δt. The accuracy of the integration will depend on both, but since integration methods are standard (e.g. a 4th order Runge–Kutta method), the time-step will introduce a trade-off between computing time and integration accuracy. However, this can be avoided altogether if the moment equations can be solved analytically, and that is possible for the presented model. In particular, the system in Eqs. (9a)–(9b) is linear but parameters are time-dependent due to the flow rate, and an analytical solution is therefore not tractable. A common solution is to impose a zero-order hold (ZOH) condition on the time-dependence (i.e. $Q_i(t) = Q_{i,k}$ for $t \in [t_k, t_{k+1}]$) such that the system becomes piece-wise linear and time invariant. Imposing this on the system in Eqs. (9a)–(9b) for the M houses yields the matrix–vector form

$$\mathrm{d}\begin{bmatrix} T_t \\ T_t^{(s)} \end{bmatrix} = \left(\underbrace{\begin{bmatrix} A_1 & S \\ 0 & 0 \end{bmatrix}}_{A} \begin{bmatrix} T_t \\ T_t^{(s)} \end{bmatrix} + \underbrace{\begin{bmatrix} B_1 \\ 0 T_g \end{bmatrix}}_{B} \right) \mathrm{d}t + \underbrace{\begin{bmatrix} \sigma & 0 \\ 0 & \sigma_s \end{bmatrix}}_{G} \begin{bmatrix} \mathrm{d}\omega_t \\ \mathrm{d}\omega_t^{(s)} \end{bmatrix} \tag{27}$$

$$= \left(A \begin{bmatrix} T_t \\ T_t^{(s)} \end{bmatrix} + B \right) \mathrm{d}t + G \begin{bmatrix} \mathrm{d}\omega_t \\ \mathrm{d}\omega_t^{(s)} \end{bmatrix}, \tag{28}$$

where the auxiliary matrices and vectors have been introduced,

$$A_1 = \begin{bmatrix} a_1 & 0 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & a_{M-1} & 0 \\ 0 & 0 & 0 & 0 & a_M \end{bmatrix}, \qquad S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{M-1} \\ s_M \end{bmatrix},$$

$$B_1 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{M-1} \\ b_M \end{bmatrix}, \qquad \sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma_{M-1} & 0 \\ 0 & 0 & 0 & 0 & \sigma_M \end{bmatrix},$$

with the lowercase variables given as

$$a_\mathrm{i} = -(s_\mathrm{i} + b_\mathrm{i}), \qquad s_\mathrm{i} = C_\mathrm{i}^{-1} Q_{\mathrm{i,k}} c_\mathrm{p}, \qquad b_\mathrm{i} = \left(C_\mathrm{i} R_\mathrm{i}\right)^{-1},$$

and where

$$T_t = \left[T_t^{(1)}, T_t^{(2)}, \ldots, T_t^{(M)}\right]^T, \qquad \mathrm{d}\omega_t = \left[\mathrm{d}\omega_t^{(1)}, \mathrm{d}\omega_t^{(2)}, \ldots, \omega_t^{(M)}\right].$$

The solution to this $(M+1)$-dimensional system of stochastic differential equations from time $t = t_\mathrm{k}$ to $t = t_{\mathrm{k}+1}$ can be written using the Itô interpretation of the stochastic differential equation as

$$x_{\mathrm{k}+1} = \mathrm{e}^{A\Delta t} x_\mathrm{k} + \int_{t_\mathrm{k}}^{t_{\mathrm{k}+1}} \mathrm{e}^{A\Delta t_\mathrm{s}} B \mathrm{d}s + \int_{t_\mathrm{k}}^{t_{\mathrm{k}+1}} \mathrm{e}^{A\Delta t_\mathrm{s}} G \mathrm{d}\omega_\mathrm{s}, \tag{29}$$

using the time differences $\Delta t = t_{\mathrm{k}+1} - t_\mathrm{k}$ and $\Delta t_\mathrm{s} = t_{\mathrm{k}+1} - s$ and using the zero-order hold assumption across these time intervals. Only the transitional mean and covariance must be computed and these remain Gaussian by the linearity of the system. They are given by

$$\mathbb{E}\left[x_{\mathrm{k}+1}\right] = \hat{A}\,\mathbb{E}\left[x_\mathrm{k}\right] + \hat{B}, \tag{30a}$$

$$\mathbb{V}\left[x_{\mathrm{k}+1}\right] = \hat{A}\,\mathbb{V}\left[x_\mathrm{k}\right]\hat{A}^T + \hat{Q}, \tag{30b}$$

where the involved matrices are

$$\hat{A} = \mathrm{e}^{A\Delta t}, \tag{31}$$

$$\hat{B} = \int_{t_k}^{t_{k+1}} \mathrm{e}^{A\Delta t_s} B \mathrm{d}s, \tag{32}$$

$$\hat{Q} = \int_{t_k}^{t_{k+1}} \mathrm{e}^{A\Delta t_s} G G^T \mathrm{e}^{A^T \Delta t_s} \mathrm{d}s, \tag{33}$$

which can be calculated by computing the matrix exponential of the augmented matrices

$$\exp\left(\begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \Delta t\right) = \begin{bmatrix} \hat{A} & \hat{B} \\ 0 & I \end{bmatrix}, \tag{34}$$

$$\exp\left(\begin{bmatrix} -A & GG^T \\ 0 & A^T \end{bmatrix} \Delta t\right) = \begin{bmatrix} V_{11} & V_{12} \\ 0 & V_{22} \end{bmatrix}, \tag{35}$$

and subsequently extracting the variance as $\hat{Q} = V_{22}^T V_{12}$ [33]. The elements of these matrices are

$$\hat{Q}_{i,i} = \frac{s_i^2}{2a_i^3}\sigma_s^2 \left[\mathrm{e}^{2a_i\Delta t} - 4\mathrm{e}^{a_i\Delta t} + 2a_i\Delta t + 3\right] + \frac{1}{2a_i}\sigma_i^2\left[\mathrm{e}^{2a_i\Delta t} - 1\right], \tag{36}$$

$$\hat{Q}_{i,j} = \frac{s_i s_j}{(a_i + a_j)a_i^2 a_j^2}\sigma_s^2 \cdot \Big(\mathrm{e}^{(a_i+a_j)\Delta t} a_i a_j - \mathrm{e}^{a_i\Delta t} a_j (a_i + a_j) - \mathrm{e}^{a_j\Delta t} a_i (a_i + a_j) + a_i^2 (a_j\Delta t + 1) + (a_j^2\Delta t + a_j) a_i + a_j^2 \Big), \tag{37}$$

$$\hat{Q}_{M+1,i} = \frac{s_i}{a_i^2}\sigma_s^2\left[\mathrm{e}^{a_i\Delta t} - a_i\Delta t - 1\right], \tag{38}$$

for $i = 1, 2, \ldots, M$ and $j = i+1, i+2, \ldots, M$. The last diagonal element is $\hat{Q}_{M+1,M+1} = \sigma_s^2 \Delta t$. The non-zero elements of $\hat{A}$ lie on the diagonal and the last column i.e.

$$\hat{A}_{i,i} = \mathrm{e}^{a_i\Delta t}, \tag{39}$$

$$\hat{A}_{M+1,i} = \frac{s_i}{a_i}(\mathrm{e}^{a_i\Delta t} - 1), \tag{40}$$

for $i = 1, 2, \ldots, M$. The last diagonal element is $\hat{A}_{M+1,M+1} = 1$. The first M entries of $\hat{B}$ are

$$\hat{B}_i = \frac{b_i}{a_i}(\mathrm{e}^{a_i\Delta t} - 1)T_g, \tag{41}$$

for $i = 1, 2, \ldots, M$, and $B_{M+1} = 0$.

The likelihood computations can therefore be carried out without having to invoke costly integration techniques, simply by computing the one-step moments in Eq. (30) by calculating these elements necessary to directly construct $\hat{A}$, $\hat{B}$ and $\hat{Q}$. The source code for the implemented likelihood functions for the two methods is available through a GitHub repository.[1]

3. Results

This section presents the results in terms of parameter estimation using the two proposed approaches, the TMB and Kalman filter to estimate the street temperature from smart meter data. Computation time and difference between the parameter estimations will be compared between the two approaches to investigate if one method is more advantageous. The results between the two different areas are also compared to validate the model's generalisability as the dynamics between the areas are different. The performance of the model in different seasons is also investigated. Finally, the challenges of the model are discussed. The data used in this paper is presented before the results from the proposed approaches are discussed.

3.1. Data

The data used in this study was provided by the district heating utility in Brønderslev, *Brønderslev Forsyning*. The data consists of measurements from smart meters in individual buildings and measurement wells from two critical areas inside the Brønderslev district heating network. Only a subset of the smart meters is needed for the proposed method. The smart-meter measurements used in this study are from 30 different single-family houses: 15 from each area. They consist of time series of *supply temperature* [°C], *return temperature* [°C], *flow* [L/h], *energy* [kWh], *volume* [m^3] with associated *timestamps*. The measurement wells are located in the distribution network at street level before the houses and they measure the *forward temperature* [°C] and the

[1] https://lab.compute.dtu.dk/hgbe/smartmeters-kalmanfilter-tmb.

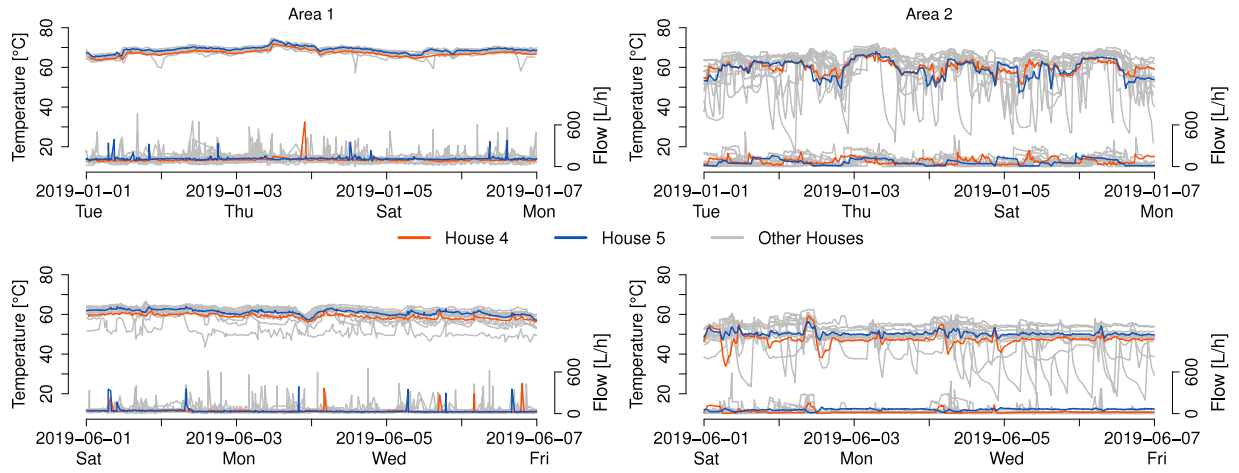


Fig. 6. The plot demonstrates for an interval of one hour, a different number of readings at different timestamps when comparing smart meter data from each house. Two houses are shown with colours while the other houses are illustrated as grey. The two houses are highlighted with colour to illustrate the difference in the dynamics between the two areas.

return temperature [°C]. In order not to violate privacy and comply with the General Data Protection Regulation (GDPR), Brønderslev Forsyning made the smart-meter data anonymous before making it available for the study by not disclosing the location of the houses. Hence, the only information given in the data is which of the two areas the house belongs to, and that houses are grouped closely together for each area. Only metering data has been used to estimate the distribution network temperature at street level, whereas the temperature measurement from the wells has only been used to validate the results.

3.2. Measurements from smart meters and measurement wells

In this paper, only temperature and flow measurements from the smart meters are used. The objective of this study is to estimate the supply temperature in the distribution pipe in the street that feeds into the houses where the smart meters are installed, without using the nearby measurement well. The temperature measurements from the well will only be used for model validation. The measured variables are denoted as follows:

$$\text{Network Temp. (Street)} : T_t^{\text{s}}, \quad t = 1, \dots, N, \tag{42}$$

$$\text{Temp. at houses} : T_t^{(\text{i})}, \ t = 1, \dots, N, \ i = 1, \dots, M, \tag{43}$$

$$\text{Flow at house} : Q_t^{(\text{i})}, \ t = 1, \dots, N, \ i = 1, \dots, M, \tag{44}$$

where the subscript t is the time index (N number of observations), and the superscript i is a label of the smart meter (or house) number and M is the number of smart meters. The measurements from the smart meters in Area 1 and Area 2 were obtained from 1 July 2018 to 1 July 2019, and 1 January 2018 to 25 September 2020, respectively. The resolution of the smart-meter data is not fixed and the number of readings each day changes over time. The measurements from the wells that represents the temperatures in the streets from Area 1 and Area 2 are on two minutes resolution, and the periods are from 4 December 2018 to 8 January 2020 and 12 September 2019 to 12 September 2020.

Smart meters are usually located close to the substation inside the houses where they measure district-heating information. The hot water is delivered to the substation by a service pipe that is connected to the district heating distribution network at street level as shown in Fig. 2. Measurement wells are located at the critical areas in the distribution network and are usually placed before the hot water enters the first house in the area at street level. Unfortunately, the measurements from the two wells do not overlap during a cold period (January) or warm period (summer) as Fig. A.15 shows. Also, there are some errors with the measurements over a longer period for both areas.

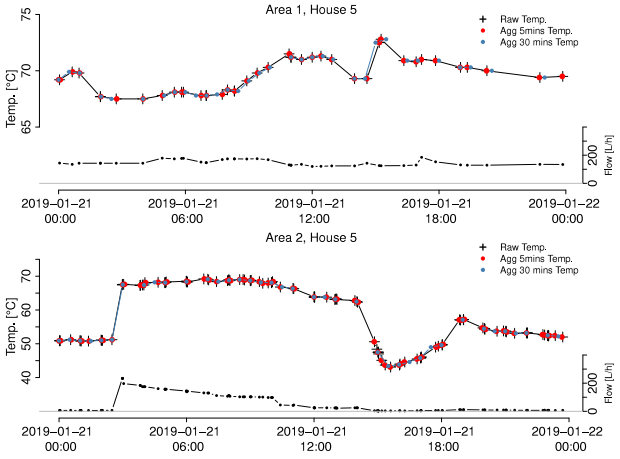


Fig. 7. The plots visualise how the resolution of the aggregation can result in less information on the temperature dynamics of the pipe. The upper and lower plots show data from Area 1 and Area 2. The raw supply temperature and flow with two different aggregation resolutions, 5 min and 30 min are visualised together. The flow demonstrates the heating dynamics in more detail, i.e. how the temperature increases and decreases with the flow behaviour.

Time series plots of the supply temperature and flow from the smart meters at the houses in both areas for a subset of the data are shown in Fig. 6. Two random houses from both areas are visualised with coloured lines, while the rest are presented as grey. These two houses are selected to highlight the different dynamics between the areas and seasons. The dynamics are very different between the areas, while houses in the same area tend to show similar behaviour, i.e. similar houses with similar control strategies. There are also seasonal patterns, e.g. usually there is no need for space heating during summer so the temperature is lower during summer as the heat demand is significantly lower compared to cold periods, e.g. January. This can also be seen in Table 1, which lists the quantiles of the temperature and flow measurements in January and June. As expected, the temperatures are higher during January. Due to the significant difference between the areas, each area will be discussed separately in the following.

Area 1: demonstrates rather constant temperature, especially during cold periods, as shown in both Fig. 6 and Table 1. Hence, there is always some consumption of heat, as the flow is almost never zero (see Q_5 for flow in Table 1). However, during a warmer period, less heat consumption is needed and therefore there is frequently no heat consumption.

Table 1
The table presents flow and temperature measurement quantiles from both areas for the periods January and June in 2019. The differences in both flow rate and temperature between June (summer) and January (winter) are very noticeable for all shown quantiles.

		Q_5	Q_{25}	Q_{50}	Q_{75}	Q_{95}
	January					
Area 1	Flow [L/h]	53.36	87.67	107.5	133.09	182.5
	Temp [°C]	66.10	68.30	69.7	71.30	72.9
Area 2	Flow [L/h]	3.59	12.73	40.75	94.0	172.30
	Temp [°C]	43.34	55.70	61.10	65.1	68.70
	June					
Area 1	Flow [L/h]	13.00	21.00	30.0	46.09	83.0
	Temp [°C]	52.50	56.80	59.5	61.20	63.3
Area 2	Flow [L/h]	7.00	13.00	17.33	22.0	56.37
	Temp [°C]	42.80	48.90	50.30	51.8	56.60

Area 2: demonstrates a more dynamic operation where the flow is frequently turned on and off (see Fig. 7 for an example). The hot water in the service pipe frequently loses heat to the surroundings, as it becomes still in the pipe for long periods when the flow is low. Hence, there is frequent temperature drops in Fig. 6 and a large gap in the quantiles in Table 1, most notably during January.

Area 1 has a more constant behaviour, as both the temperature and flow do not change rapidly. Area 2 shows more of an on/off behaviour, i.e. turning on and off the heating over time. There is a significantly larger variability in Area 2 in the temperature readings. Area 2 might therefore be low-energy buildings where the substation is controlled intelligently compared to the traditional operation of substations where only the outside temperature is used. It can be assumed that Area 1 uses traditional operations and the buildings are likely to be older. Therefore, they usually have lower energy efficiency and need more heat to keep the indoor climate comfortable. The on/off control strategy impairs the quality of the temperature measurement from the smart meters. The water in the service pipe becomes still and therefore it does not give an accurate representation of the temperature in the distribution pipe. Consequently, this needs to be considered when estimating the temperature, to avoid including these periods where the signal becomes unreliable. The seasonal behaviour demonstrated in both areas occurs when the ambient air temperature has increased above a certain cutoff temperature where no space heating is required to feel comfortable inside, as space heating is highly correlated with the ambient air temperature. Hence, the quality of the temperature signal is also seasonal-dependent, as the amount of flow is affected by the desired space heating.

Bergsteinsson et al. [14] describe that each smart meter is unique, as the quality of the signal depends on the smart meter and the quality of the temperature signal is flow-dependent. Smart meters send instantaneous values with different resolutions, as is highlighted in Fig. 8. The plot visualises the temperature readings from all houses in Area 1 over a one-hour period. Note that one house did not send any readings during this period. Some meters only send one value, while others send multiple. Therefore, the data from the smart meters needs to be resampled, as required by the method used to estimate the supply temperature.

In this study, the discrete-time Kalman filter is used to estimate the house and street temperatures. It is therefore convenient to have the smart-meter data resampled at the same time points. As demonstrated in Fig. 8, the data does not have a fixed resolution. The purpose of this method is to estimate the temperature in the street distribution pipe from historical data, but it is not used to predict future temperature values. The smart-meter data is therefore aggregated with the desired resolution by rounding to the nearest time point. An appropriate resolution needs to be used such that as little as possible of the information vital for the analysis is lost. Fig. 7 shows the time series of the raw

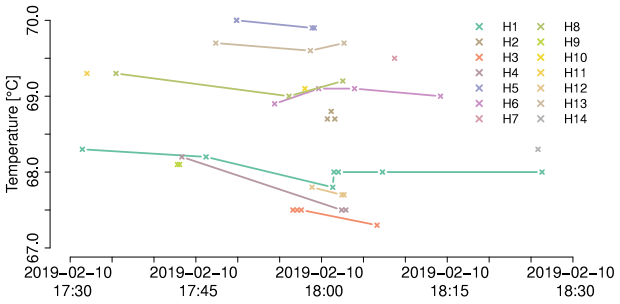


Fig. 8. The plot demonstrates an interval of one hour. There is a different number of readings with different timestamps when comparing smart-meter data from each house. There is also one house missing in this period.

supply temperature data and aggregates by rounding to 5 min resolution and 30 min. The flow is also plotted to highlight the heating dynamics of the houses and between the areas. Because of the on/off dynamics of the heating for House 5 in Area 2, the importance of the resolution can be seen more clearly than in Area 1. When the flow is shut off, the temperature starts to decrease and the rapid drop happens quite fast. Therefore, having all information on how the temperature decreases is important. Hence, using the correct aggregation resolution becomes essential to capture the heat dynamics of the service pipe. Investigating the difference between 5 min and 30 min in the drop, the 30 min resolution results in less information, as expected.

The smart-meter data in this study will be resampled to the same 5 min resolution and using one month from cold and warm periods to demonstrate the performance of the proposed method of estimating the distribution network temperature at street level. In Area 1, measurements from January 2019 and June 2019 will be used to estimate the parameters of the model. The estimated temperature will be validated by comparing it to the measured temperature from the well. The validation will be done in the same period as the parameters are estimated, as the well temperatures are not used in the model estimation and this method's purpose will only be used to estimate the temperature from the previous day when the smart-meter data arrives. Area 2 will use January 2020 and June 2020 to estimate the model parameters. However, during January it is not possible to validate the estimated temperature, as the measured temperature at the well for this period is wrong, as seen in Fig. A.15.

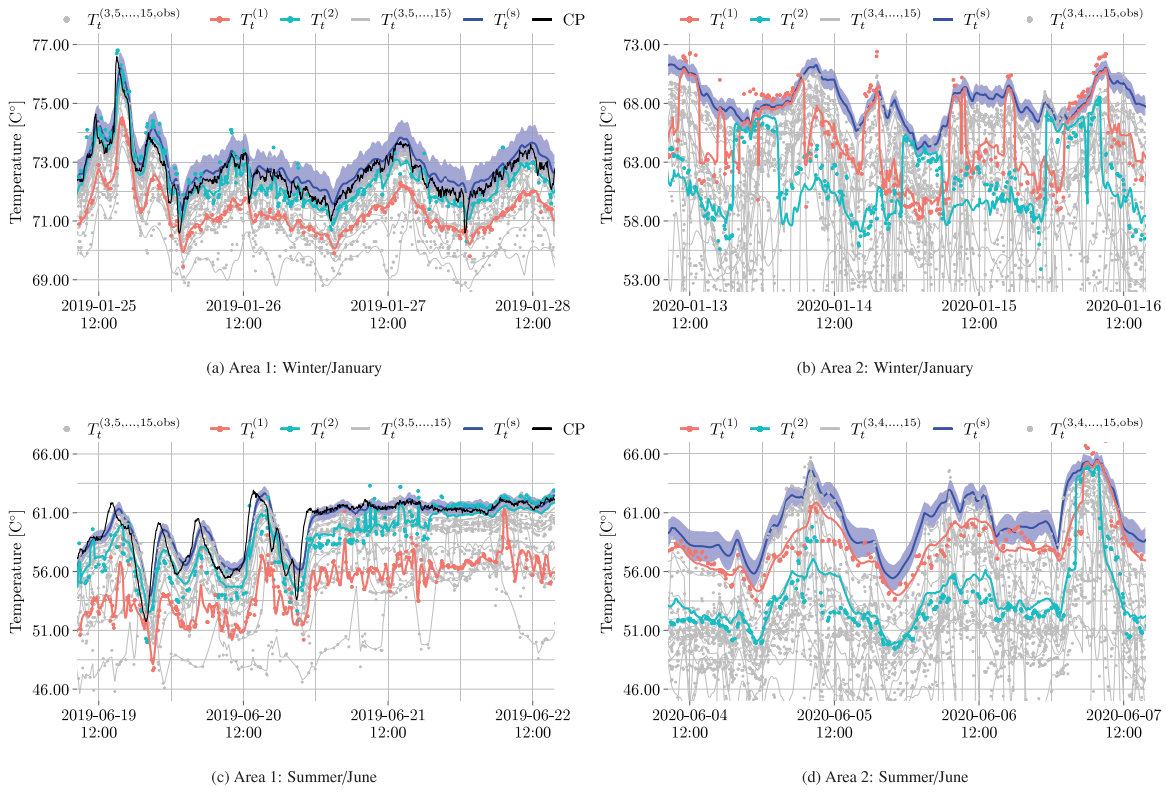
3.3. Settings and procedure

The results presented in the following section were obtained using the Kalman Filter method, which proved the fastest and most robust. The parameter estimation is carried out using the Kalman Filter and the smoothed estimate of the states and uncertainties are then found with the mixed-effects method by supplying the estimated parameters. The parameter estimation is roughly ten times faster using the Kalman filter method, as shown in Table A.2. A linear fit to this data in the log–log domain shows that the computation time for the Kalman filter and mixed-effects method grows by $M^{2.6}$ and $M^{2.2}$ respectively. The faster computation time using the Kalman Filter is explained by a much lower intercept. There was no drawback to the faster parameter estimation, since results were consistent regardless of the method used (Table A.3). A total of fifteen houses were used, and data from a one-month period containing 15 × 8353 observations of the house temperatures (≈80% are NA-values) and the same amount of house flow rate observations.

The following upper and lower parameter bounds were used

$$\theta^{(i)}_{\text{upper}} = \left[C^{(i)}_{\text{upper}}, R^{(i)}_{\text{upper}}, \sigma^{(i)}_{\text{upper}}\right] = [500, 1500, 2], \tag{45}$$

$$\theta^{(i)}_{\text{lower}} = \left[C^{(i)}_{\text{lower}}, R^{(i)}_{\text{lower}}, \sigma^{(i)}_{\text{lower}}\right] = \left[1, 1, 10^{-5}\right], \tag{46}$$



(a) Area 1: Winter/January

(b) Area 2: Winter/January

(c) Area 1: Summer/June

(d) Area 2: Summer/June

Fig. 9. Plots showing the estimated street temperature and the measured house temperatures. The top and bottom plots show the estimation from winter and summer respectively, and plots on the left and right belong to Area 1 and Area 2 respectively. The measured street temperature that is the critical point (CP) temperature is plotted for Area 1 to demonstrate the performance of the proposed method. These measurements were not available for Area 2.

for each house to ensure realistic estimates (see Eqs. (25) and (26)). The parameter bounds were imposed naturally by introducing the inverse logit domain transformation

$$\theta = \left(\theta_{\text{upper}} - \theta_{\text{lower}}\right) S(\theta_{\text{logit}}) + \theta_{\text{lower}}, \tag{47}$$

where $S(x) = (\exp(-x) + 1)^{-1}$. The diffusion parameters σ_i were generally very uncertain, thus difficult to estimate. It was necessary to introduce a lower bound constraint to avoid numerical instability as a consequence of diffusion parameters tending towards zero. The observation variance behaved in a similar fashion and was fixed at $\sigma_{obs}^2 = 1$ (which amounts to allowing temperature fluctuations on the order of ± 2 degrees). The uncertainty is a rough approximation of the uncertainties informed by a specific smart meter manufacturer [34], and is reported to be 3%–5% (depending on the flow), which amounts to 1.5–2.5 °C assuming a temperature of 50 °C.

3.4. Empirical results

The smoothed-state reconstructions for the two areas are presented along the rows, with an estimate for a winter month (January) and a summer month (June) along the columns, in Fig. 9. A pronounced difference can be seen between the dynamics in the two areas; Area 1 displays rather stable house temperatures relative to Area 2 where they surge up and down. This surging is driven by similar behaviour in the flow, as discussed in Section 3.1. Hence, the necessity of having the flow-dependent observation noise as shown in Eq. (11) is clear. One advantageous effect of the flow-dependent implementation is that the street temperature estimate is less prone to rapid changes during periods where all flows simultaneously decrease below the set threshold. This occurs because the estimated street temperature attempts to follow the fast dynamics of the houses, but such behaviour is contrary to its slower temperature dynamics. The implementation was therefore necessary in order to achieve more realistic and accurate results. It can be argued that, even though increasing the number of houses would decrease the probability that all flows decrease below the set threshold simultaneously, the implementation remains valuable in order to appropriately penalise the impact of low-flow observations.

The model performance can be validated by comparing the estimated street temperature with the critical point measurements in Area 1. The street temperature predictions seem to have captured the overall trends and oscillations quite well, although a systematic bias is evident during summer, and in winter a slight temporal delay is seen. Comparing the two areas, the effect of having accurate information (due to the high flow) is evidently a smoother temperature curve for all houses and the street. The estimation in Area 2 displays faster variations as a result of the rapid variance changes and differences in the number of houses that provide information. In particular, inspecting the two highlighted temperatures ($T_t^{(1)}$ and $T_t^{(2)}$) emphasises that whenever information becomes available (and the flow is high) the individual houses surge upwards to some limit from where they determine the street temperature estimate. A primary challenge with the estimation in Area 2 is to prevent the slower dynamics of the street from being controlled by the much faster dynamics of the houses. This is a trade-off between measurement variance and obtaining what appears to be adequately slow changes in the street temperature estimates. The effect of the variance increase cannot be assessed due to the lack of critical point measurements in this area.

It was discovered that the estimation accuracy is sensitive to temperature observations that are much higher than expected, e.g. from poorly calibrated sensors. The challenge of detecting such problematic houses is difficult to solve in general. These houses are identified by inspection of the thermal resistances R_{i} because they tend to hit the upper bound. The modeller should be aware of the possibility that such houses have a controlling influence on the street temperature. In this particular case, the thermal resistances of both House 8 and House 4 were converging to the boundary, although only the former had observations that were

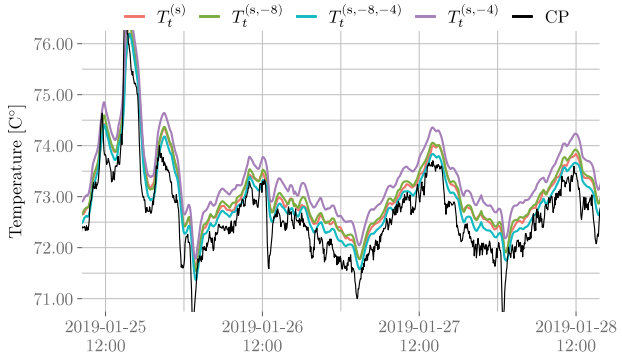


Fig. 10. The estimated street temperature using all houses and after removing the two houses whose thermal resistance estimates hit the boundary $R_i = 1500$. The estimate is not significantly improved until the second house is removed.

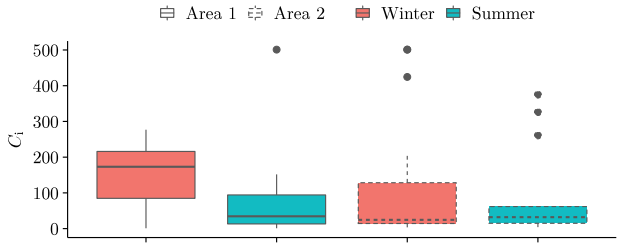


Fig. 11. Boxplots of estimated thermal capacities C for the two areas and two months. The median values are clearly largest in Area 1 during the winter month but a few outliers are seen in Area 2 regardless of season.

much above the critical point measurements. The effect of omitting these two houses from the estimation was investigated after omitting none, one or both. The estimated street temperatures are compared to the critical point temperatures in Fig. 10, and the associated mean average errors (MAEs) are

$$\begin{aligned} \text{None}: &\quad \text{MAE} = 0.46 \\ \text{House 4}: &\quad \text{MAE} = 0.76 \\ \text{House 8}: &\quad \text{MAE} = 0.50 \\ \text{House 4 \& 8}: &\quad \text{MAE} = 0.32 \end{aligned}$$

This shows that the MAE increases (0.3/65%) after removing House 4, increases slightly (0.04/8%) after removing House 8 and decreases (0.14/30%) after removing both House 4 and House 8. The surprising outcome in the former case was further investigated and apparently it occurs because only a single house will act as a controller of the street temperature. In this case, House 4 dominates the street temperature estimation, so if removed House 8 will start to dominate the temperature estimation which has significantly higher observations and thus pulls the estimated street temperature up towards itself. In contrast, there is no effect from omitting House 8, because House 4 remains in control.

3.5. Thermal parameters and house dynamics

A statistical overview of the obtained values of C_i and R_i is provided in Figs. 11 and 12, respectively. The former figure shows that the estimated winter capacities from Area 1 stand out from the others by having a significantly higher median value, and the parameters are generally more dispersed in Area 1 regardless of period (as evident from the interquartile ranges of 131, 81, 114 and 47, respectively). Comparing areas shows that the dispersion is high during winter and low during summer. This conclusion is reversed when inspecting the estimated resistances, but could not be explained by strong correlations between

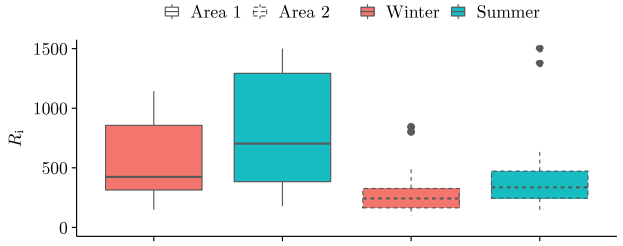


Fig. 12. Boxplots of the logarithmic value of the thermal resistance R. There is a clear difference between the two areas, and generally speaking the interquartile range is larger during the summer.

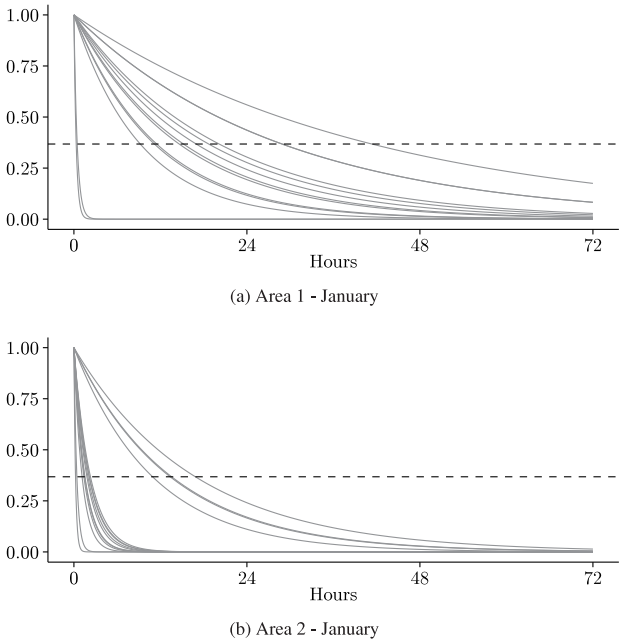


Fig. 13. The decaying exponential functions $\exp(-t/\tau)$ which depict how fast the temperature decreases from some initial set-point temperature down to the ground temperature T_g in the individual house systems. The dashed line shows when it has decayed to 37% of its initial value.

R_i and C_i. The estimates and the estimated confidence intervals for winter and summer estimation are shown in Table A.4.

The stochastic differential equations describing each house give rise to *time constants* (τ) which can be interpreted as the duration after which the temperature has decreased by roughly 36%. This yields a straight-forward way to compare and characterise the thermal properties of the service pipe at each house in a single quantity, and this enables quick identification of e.g. bad insulation, leakage or out-dated installations. The time constants are the eigenvalues of the system matrix A in Eq. (27)

$$\tau_i = \left(c_p \frac{Q_i}{C_i} + \frac{1}{R_i C_i} \right)^{-1} \approx R_i C_i. \tag{48}$$

The approximation is only valid when the flow is low (approaching zero), since the former term then becomes very small. The time constants are illustrated as decaying exponential functions in Fig. 13 for the two areas in the winter month. Evidently, certain houses can be identified as cooling down much more rapidly than the others; the houses where temperature converges to zero instantly. In Area 2, there seems to be two groups of houses: one which slowly converge to zero and another which cool down faster. There also seem to be two houses that cool down faster than other houses in the same area. This information could be used to investigate whether there is a problem with the service pipe.

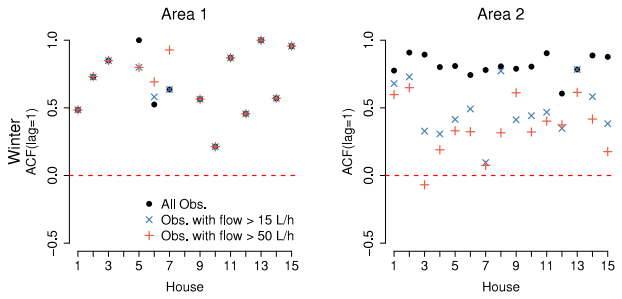


Fig. 14. The figure shows the one lag value of the ACF from the prediction errors for each house for the two areas in separate plots. It highlights that, when the flow is low, the model does not capture the dynamics of the temperature adequately.

3.6. Model validation

Model validation is always important, but in the presented case there are a few challenges of applying e.g. the autocorrelation function (ACF). One is that the data used in this study does not have a fixed resolution (Fig. 8), and the more difficult one is that the model does not try to capture the behaviour at all times. When the flow is low, the dynamics of the temperature is not captured well by the model, as the dynamics will be more influenced by e.g. the placement of the smart meter inside the house and higher heat loss in the service pipe due to the slow transportation time. In the model, this is handled by increasing the observation variance. However, the 1-step residuals will show a very large autocorrelation in these situations (Figs. A.16 and A.17) leading to a very large lag-one autocorrelation for the overall data. Disregarding data when the flow is high results in smaller autocorrelation (Fig. 14). Even with these reservations, it seems clear that, at least for some houses, the lag-one autocorrelation is high, indicating that some model deficiencies could be addressed in future work.

Also, as an approximation, the temperature distribution over the service pipe is is assumed to be uniform; i.e. the discretisation which results in Eq. (7). This assumption is only valid when flow is high. Hence, the temperature distribution is not uniform during situations when flow is low. Also, the transportation time of the water moving will therefore take longer. It can be calculated from

$$Q = \rho V = \frac{\rho A L}{t} \rightarrow t = \frac{\rho A L}{Q}, \tag{49}$$

where Q [kg/s] is the mass flow rate, ρ [kg/m^3] is the density of the water, V [m^3/s] is the volumetric flow rate, A [m^2] is the cross sectional area of the pipe, L [L] is the length of the pipe, and t [s] is the time. Some of the pipe properties needs to be assumed, for instance that the pipe is a DN25 with inner diameter of 0.0273 [m], the length of the pipe is 10 [m] and the flow is 50 [kg/h], then the transportation time can be computed,

$$t = \frac{997\ [\text{kg/m3}](\pi(0.0273\ [\text{m}])^2/4)10\ [\text{m}]}{50\ [\text{kg/h}]/60\ [\text{h/min}]} \approx 7\ \text{min}. \tag{50}$$

In this example, the transportation time is longer than the resolution time of the data used in this study, which indicates that the uniform temperature distribution is not valid and could lead to higher autocorrelation in the errors.

4. Discussion

This paper modelled the thermodynamics of a service pipe that delivers heat from a distribution pipe in a district heating network to consumers' substations. A partial differential equation of the thermodynamics in the service pipe was presented, and then approximated by a stochastic differential equation. Using this set of equations, an approach was established to obtain temperature feedback at arbitrary points in the district heating network by using measurements from a group of smart meters located inside single-family houses close to the chosen point. The network temperature is the temperature of the hot water inside the distribution pipe before it delivers hot water to the consumers' service pipe. The street temperature was modelled as a random walk. A combined SDE model composed of the street temperature connected to all of the houses' service pipes through identical Resistance-Capacity SDEs was constructed. This automatically restricts the street pipe temperature to at least as high as the temperature measurements from the smart meters. It was shown that the presented methodology produced accurate results when compared to the measured network temperatures in a single area with fifteen houses. In another area of fifteen houses, it was shown how to deal with rapidly decreasing temperatures when the flow rates become too low, by increasing the uncertainty of these observations.

The parameter estimation was carried out using two different methods, namely a discrete Kalman filter and a mixed-effects method, both implemented as `C++` files and used with the `R` package `TMB`. The latter was the intended formulation using `TMB` while the former only took advantage of the automatic differentiation provided by `TMB`. The speed-up gained using `C++` together with automatic differentiation is substantial, and results showed that the computation time for the Kalman filter method was approximately ten times faster than that of the mixed-effect formulation. For this reason, parameter estimation was carried out using the Kalman filter, while smoothed-state estimates and variances were calculated using the mixed-effects formulation.

The smoothed-state estimates are presented here since these are more appropriate than one-step prediction estimates due to the in-sample use case. Thus, because the model is intended to assess the feedback from the network for computing a transfer function for the network characteristics, e.g. time delay and temperature loss in the system, the posterior state estimates that use all available information are appropriate. In a concrete scenario, one could imagine that the previous 24 h of data are sent from the smart meters once per day. This information can then be used for temperature optimisation of the supply temperature at the plant with the aim of reducing the heating cost and heat losses in the system by lowering the supply temperature at the plant. The ability to establish temperature feedback in district heating networks opens many possibilities for utility companies to improve their operation. Most importantly, it gives the ability to use controllers for data-driven temperature optimisation of the network. Temperature optimisation reduces the heat loss and lower the needed network supply temperature, and thereby the operational cost of the heat production is reduced. It furthermore makes physical measurement wells in the network redundant, reducing costs, planning time, installations and maintenance. The feedback becomes flexible since any group of houses can be selected to establish a new network temperature. This is highly beneficial for the utility company, since the location where the highest temperature loss in the network will vary across time, due to deterioration of pipes, replacement of older pipes with newer ones, leakage and so forth. Finally, using smart meters can make multi-temperature zones inside the network more feasible, where more local heat sources can be added to the network. Thus, lowering the operating temperature in the grid using temperature optimisation and having more detailed information from the grid will give rise to more decentralised heat sources. For instance, heat sources such as heat pumps, waste heat from industries, and solar thermal collectors (with thermal storage systems) can be included to provide heat to consumers. This will also increase the efficiency of sector coupling with the electricity sector, where the renewable energy systems are better utilised.

The authors argue that the presented model is advantageous due to (1) its relative simplicity, (2) its ability to handle scenarios with a lot of information, and (3) its ability to assess thermal properties

and the outlier detection possibilities that this enables. In particular, it was proposed that houses with resistance lying on the upper boundary should be omitted from the analysis. A few issues are, however, worth mentioning: Firstly, the choice of a random walk to estimate the street temperature generally creates poor conditions for long-term forecasting because its variance increases linearly with time. While this is of no concern here due to the in-sample model purposes, should one use the model for predictions it is crucial that the uncertainty assessment is corrected. Therefore, in order to improve the forecastability of the proposed method, it would be necessary to replace the random walk using some model. For instance, the Ornstein–Uhlenbeck process could be used, although that would require some assumptions or knowledge about the parameters that will have to be derived from the forwarded temperature at the district heating facility. A second issue is the assumption that the temperature distribution over the service pipe is uniform. however, this assumption is not valid when the flow is low as the flow would also not be constant through the service pipe. It also depends on the properties of the pipe, e.g. the diameter and length. One suggestion for model extension would be to increase the model order by dividing the pipe into multiple segments to model the temperature distribution over the pipe. This extension could lower the autocorrelation of the prediction errors.

5. Conclusion

This paper has demonstrated how smart-meter data can be used to improve the operation of a district heating network by establishing temperature feedback of the network. Temperature feedback is highly valuable for temperature control of the network. Simplified descriptions using stochastic differential equations are formulated from considering partial differential equations that describe the thermodynamics of the hot water in the service pipe from the distribution pipe to consumer substations. A random walk is used to model the temperature variations in the distribution pipe that is connected to all houses. The street temperature and thermal parameters of the model are estimated by minimising the negative log-likelihood function using a discrete Kalman filter. Smoothed state estimates are subsequently computed using the mixed-effects formulation in `TMB`. The results show that the proposed method can mimic the measured street temperature accurately when compared to observations. It is important that houses whose thermal parameters hit the upper boundary are removed to reduce the estimation bias in the street temperature. The estimation procedure entails that smart-meter data arrives daily and contains measurements with same resolution from the past 24 h. The procedure uses this information to estimate the network temperature for the past 24 h. The proposed method further allows for potential identification of houses with e.g., bad insulation or leakages. They can be identified by inspecting whether the estimated thermal resistance, capacity and resulting time constant of a particular house, lie outside of the expected range.

The possibilities for future work for using smart meters to enhance the operation of district heating networks are endless. Hence, the utilities have become very data-rich and can use this information to learn how the network is performing. Utilising this opportunity will help district heating to become more energy efficient in the transformation to RES and increase its flexibility. The extension of the proposed method in this paper would be to improve the forecasting ability and validate its potential for on-line control of supply temperature at the plant. In future work, the authors also aim to investigate model extensions e.g., by dividing the service pipe into multiple segments. Furthermore, a simulation study could be beneficial to investigate the estimation of R and C parameters of the service pipe.

CRediT authorship contribution statement

Hjörleifur G. Bergsteinsson: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Phillip B. Vetter:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Jan Kloppenborg Møller:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Henrik Madsen:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work is funded by Innovation Fund Denmark through the projects HEAT 4.0 (8090-00046B), CITIES (1305-00027B), Flexible Energy Denmark (8090-00069B), TOP-UP (9045-00017B), and Region H through the IDASC project (18012745). Finally, the project is partly funded under the Norwegian FME-ZEN project financed by the ZEN partners and the Research Council of Norway (257660).

We thank the district heating utility *Brønderslev Forsyning* for their support and for supplying data & information for this study. We would also like to thank Torben Skov Nielsen and ENFOR for providing their support and knowledge about the temperature optimisation of a district heating network.

Appendix

See Figs. A.15–A.17 and Tables A.2–A.4.

Table A.2
The approximate computation times (seconds) for parameter estimation using either the Kalman Filter or the mixed-effects method as a function of the number of houses M. The dimensions of the system are $M+1$ and the number of parameters are $3M+1$. The estimation is based on one month of data with a sampling time of 5 min, which corresponds to 8350 temperature observations for each house, although roughly ≈80% are missing values (NA-values).

Number of houses [M]	Kalman	Mixed-Effects	Ratio
2	1	29	1:22
3	4	59	1:15
4	6	101	1:18
5	11	140	1:13
6	16	206	1:13
7	26	282	1:11
8	33	416	1:13
9	54	518	1:10
10	69	685	1:10
11	86	829	1:10
12	118	1157	1:10
13	143	2112	1:15
14	168	2273	1:14
15	198	2266	1:11

Table A.3
The thermal parameter estimates for 10 houses using the Kalman or the mixed-effects model formulations. The estimates are seen to be identical for all practical purposes.

Parameter	Kalman	TMB	Difference
C_1	153	155	2
C_2	11	12	1
C_3	254	258	3
C_4	29	30	1
C_5	355	358	3
C_6	472	478	6
C_7	5	5	0
C_8	8	14	5
C_9	335	341	6
C_{10}	238	240	2
R_1	245	245	0
R_2	648	648	0
R_3	219	219	0
R_4	1085	1084	1
R_5	185	185	0
R_6	125	125	0
R_7	619	619	0
R_8	1501	1501	0
R_9	266	266	0
R_{10}	450	450	0

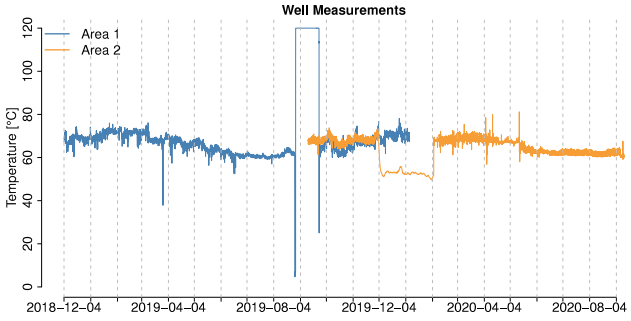


Fig. A.15. The figure shows the time series of the available temperature measured at the wells for both areas. The wells are located before the houses. Notice, that there are some problems with the measurements, e.g. during October 2019 for Area 1 and during December 2019 and January 2020 for Area 2.

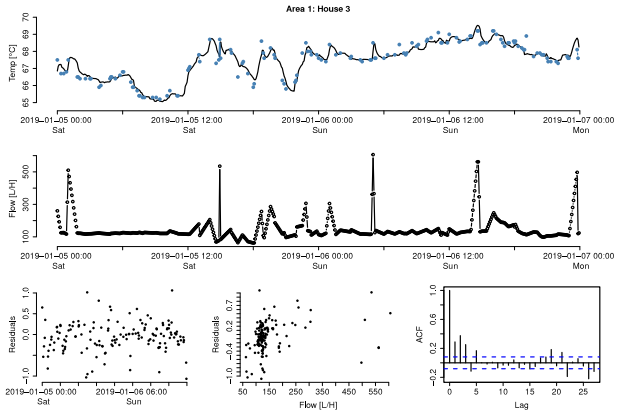


Fig. A.16. The figure displays difficulties in validating the model by doing residuals analysis on the one-step predictions in Area 1. The top plot shows the measured temperature (black solid line) and the one-step prediction (blue points). The flow for the same period is illustrated in the second top plot. In the bottom are three plots that show the residuals analysis of the one-step prediction errors. The first plot on the bottom shows the residuals over time, the next plot shows the residuals versus the flow, and the last plot shows the autocorrelation function (ACF) of the residuals.

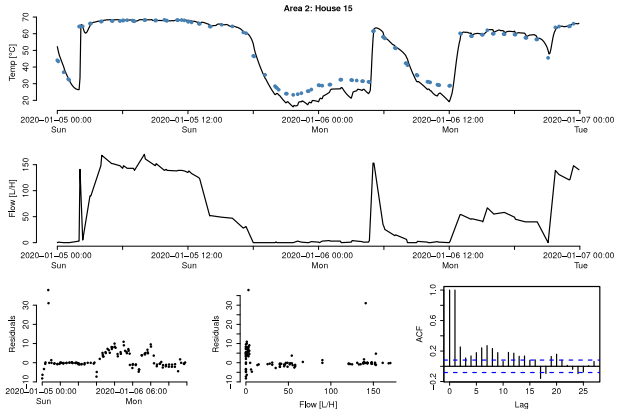


Fig. A.17. The figure displays difficulties in validating the model by doing residuals analysis on the one-step predictions in Area 2. The top plot shows the measured temperature (black solid line) and the one-step prediction (blue points). The flow for the same period is illustrated in the second top plot. In the bottom are three plots that show the residuals analysis of the one-step prediction errors. The first plot on the bottom shows the residuals over time, the next plot shows the residuals versus the flow, and the last plot shows the autocorrelation function (ACF) of the residuals.

Table A.4
Parameters and their confidence interval for the winter and summer estimation for both areas. For Area 1, there is no CI as the hessian could not be computed for the summer period.

Parameter	Area1		Area2	
	Winter Estimate (CI)	Summer Estimate (CI)	Winter Estimate (CI)	Summer Estimate (CI)
C_1	101.9 (95.56, 108.55)	23.18 (,)	1 (1, 501)	63.31 (36.46, 105.9)
C_2	1 (1, 501)	1 (,)	14.92 (13.23, 16.83)	28.31 (19.14, 41.72)
C_3	173.1 (165.92, 180.43)	34.48 (,)	33.3 (33.16, 33.45)	38.28 (37.87, 38.69)
C_4	(,)	(,)	209.22 (201.04, 217.5)	44.98 (44.24, 45.73)
C_5	245.16 (232.6, 257.76)	94.28 (,)	13.69 (13.56, 13.83)	3 (1.01, 215.3)
C_6	276.88 (256.16, 297.25)	501 (,)	6.9 (6.37, 7.47)	260.84 (238.7, 282.82)
C_7	1.91 (1, 482.67)	1 (,)	26.77 (26.65, 26.89)	32.02 (31.8, 32.25)
C_8	(,)	(,)	21.09 (20.58, 21.6)	11.68 (6.21, 22.63)
C_9	245.27 (234.49, 256.07)	92.26 (,)	6.53 (6.12, 6.97)	16.62 (7.25, 38.95)
C_{10}	188.13 (175.59, 201.01)	103.34 (,)	424.55 (386.15, 451.75)	325.95 (316.36, 335.29)
C_{11}	108.01 (103.37, 112.79)	61.95 (,)	501 (1, 501)	374.99 (356.71, 391.66)
C_{12}	215.99 (207.73, 224.32)	12.88 (,)	501 (1, 501)	60.27 (58.5, 62.09)
C_{13}	55.7 (47.37, 65.31)	13.16 (,)	23.83 (22.2, 25.58)	10.44 (8.32, 13.17)
C_{14}	84.73 (74.02, 96.66)	31.55 (,)	24.55 (24.32, 24.79)	25.76 (23.5, 28.24)
C_{15}	214.41 (210.22, 218.61)	151.78 (,)	47.72 (47.37, 48.06)	13.17 (9.27, 18.85)
R_1	325.13 (324.43, 325.82)	376.67 (,)	802.19 (796.28, 808.1)	1501 (1, 1501)
R_2	1144.12 (1101.73, 1183.31)	1501 (,)	334.75 (334.42, 335.08)	505.99 (504.8, 507.19)
R_3	316.06 (315.02, 317.1)	413.17 (,)	248.26 (248.06, 248.47)	369.16 (367.44, 370.89)
R_4	(,)	(,)	187.79 (187.53, 188.05)	295.92 (295.72, 296.12)
R_5	249.1 (248.48, 249.72)	383.63 (,)	511.81 (511.11, 512.52)	662.55 (659.75, 665.35)
R_6	149.37 (149.27, 149.46)	178.89 (,)	214.6 (214.54, 214.66)	336.28 (335.31, 337.24)
R_7	863.29 (858.63, 867.94)	1344.51 (,)	242.79 (242.52, 243.05)	315.6 (315.07, 316.13)
R_8	(,)	(,)	255.48 (255.43, 255.54)	359.45 (358.9, 359.99)
R_9	423.67 (420.63, 426.72)	702.91 (,)	843.58 (840.51, 846.64)	1376.71 (982.53, 1478.17)
R_{10}	790.97 (768.8, 813.06)	1314.43 (,)	113.28 (113.12, 113.43)	121.07 (121.03, 121.12)
R_{11}	962.96 (921.01, 1003.51)	911.85 (,)	121.09 (120.94, 121.25)	126.53 (126.49, 126.58)
R_{12}	243.45 (243.17, 243.73)	402.5 (,)	97.4 (97.26, 97.53)	228.06 (227.93, 228.19)
R_{13}	722.03 (711.75, 732.32)	917.09 (,)	318.39 (318.29, 318.49)	438.79 (437.89, 439.69)
R_{14}	856.07 (822.25, 889.46)	1292.8 (,)	166.48 (166.43, 166.52)	236.79 (236.63, 236.96)
R_{15}	313.96 (313.48, 314.44)	371.96 (,)	162.23 (162.13, 162.33)	253.68 (253.45, 253.91)
σ_1	1.005e−05 (1.000e−05, 2.000e+00)	2.908e−02 (,)	2.456e−01 (2.363e−01, 2.553e−01)	1.014e−05 (1.000e−05, 2.000e+00)
σ_2	1.123e−05 (1.000e−05, 2.000e+00)	1.000e−05 (,)	2.406e−02 (1.911e−02, 3.027e−02)	1.202e−05 (1.000e−05, 2.000e+00)
σ_3	1.001e−05 (1.000e−05, 2.000e+00)	4.019e−02 (,)	8.904e−02 (8.872e−02, 8.937e−02)	1.083e−01 (1.076e−01, 1.090e−01)
σ_4	(,)	(,)	4.869e−02 (4.851e−02, 4.886e−02)	1.151e−05 (1.000e−05, 2.000e+00)
σ_5	6.549e−03 (5.815e−03, 7.376e−03)	2.245e−02 (,)	4.023e−02 (3.938e−02, 4.109e−02)	8.395e−05 (1.000e−05, 2.000e+00)
σ_6	1.180e−05 (1.000e−05, 2.000e+00)	4.628e−02 (,)	7.117e−02 (6.957e−02, 7.279e−02)	1.981e−02 (1.950e−02, 2.012e−02)
σ_7	1.026e−05 (1.000e−05, 2.000e+00)	1.445e−05 (,)	1.117e−01 (1.113e−01, 1.120e−01)	6.177e−02 (6.088e−02, 6.267e−02)
σ_8	(,)	(,)	1.001e−05 (1.000e−05, 2.000e+00)	1.232e−05 (1.000e−05, 2.000e+00)
σ_9	1.002e−05 (1.000e−05, 2.000e+00)	1.013e−05 (,)	1.006e−05 (1.000e−05, 2.000e+00)	1.158e−05 (1.000e−05, 2.000e+00)
σ_{10}	1.006e−05 (1.000e−05, 2.000e+00)	1.014e−05 (,)	4.317e−02 (4.309e−02, 4.325e−02)	1.545e−02 (1.539e−02, 1.551e−02)
σ_{11}	1.002e−05 (1.000e−05, 2.000e+00)	1.010e−05 (,)	3.288e−02 (3.283e−02, 3.294e−02)	1.149e−02 (1.139e−02, 1.159e−02)
σ_{12}	1.002e−05 (1.000e−05, 2.000e+00)	2.857e−02 (,)	4.341e−02 (4.336e−02, 4.345e−02)	1.170e−05 (1.000e−05, 2.000e+00)
σ_{13}	1.001e−05 (1.000e−05, 2.000e+00)	1.025e−05 (,)	1.000e−05 (1.000e−05, 2.000e+00)	1.409e−05 (1.000e−05, 2.000e+00)
σ_{14}	1.004e−05 (1.000e−05, 2.000e+00)	1.050e−05 (,)	6.046e−02 (5.997e−02, 6.096e−02)	1.149e−05 (1.000e−05, 2.000e+00)
σ_{15}	1.000e−05 (1.000e−05, 2.000e+00)	1.690e−02 (,)	9.479e−02 (9.456e−02, 9.501e−02)	1.208e−02 (1.759e−05, 1.813e+00)
σ_{16}	0.01 (0.01, 0.01)	0.01 (,)	1.542e−02 (1.540e−02, 1.544e−02)	1.479e−02 (1.471e−02, 1.487e−02)

References

[1] Guelpa E, Verda V. Thermal energy storage in district heating and cooling systems: A review. Appl Energy 2019;252:113474. http://dx.doi.org/10.1016/j.apenergy.2019.113474.

[2] Mathiesen B, Lund H, Connolly D, Wenzel H, Ostergaard P, Möller B, Nielsen S, Ridjan I, KarnOe P, Sperling K, Hvelplund F. Smart energy systems for coherent 100% renewable energy and transport solutions. Appl Energy 2015;145:139–54. http://dx.doi.org/10.1016/j.apenergy.2015.01.075.

[3] Blanco I, Guericke D, Andersen AN, Madsen H. Operational planning and bidding for district heating systems with uncertain renewable energy production. Energies 2018;11(3310). http://dx.doi.org/10.3390/en11123310.

[4] Blanco I, Andersen AN, Guericke D, Madsen H. A novel bidding method for combined heat and power units in district heating systems. Energy Syst 2019;11:1137–56. http://dx.doi.org/10.1007/s12667-019-00352-0.

[5] Madsen H, Sejling K, Sgaard HT, Palsson OP. On flow and supply temperature control in district heating systems. Heat Recovery Syst CHP 1994;14(6):613–20. http://dx.doi.org/10.1016/0890-4332(94)90031-0.

[6] Arvastson L. Stochastic modeling and operational optimization in district heating systems (Ph.D. thesis), Centre for Mathematical Sciences, Lund University; 2001.

[7] David A, Mathiesen BV, Averfalk H, Werner S, Lund H. Heat roadmap europe: Large-scale electric heat pumps in district heating systems. Energies 2017;10(4). http://dx.doi.org/10.3390/en10040578.

[8] Oliker I. Steam turbines for cogeneration power plants. J Eng Power 1980;102(2):482–5. http://dx.doi.org/10.1115/1.3230281.

[9] Madsen H, Holst J. Estimation of continuous-time models for the heat dynamics of a building. Energy Build 1995;22(1):67–79. http://dx.doi.org/10.1016/0378-7788(94)00904-X.

[10] Nielsen HA, Madsen H. Modelling the heat consumption in district heating systems using a grey-box approach. Energy Build 2006;38(1):63–71. http://dx.doi.org/10.1016/j.enbuild.2005.05.002.

[11] Nielsen TS, Madsen H. Control of supply temperature in district heating systems. In: Proceedings of the 8th international symposium on district heating and cooling; 2002.

[12] Directive 2012/27/EU. Brussels: European Parliament and Council of the European Union; 2012, Accessed on 28 March 2021.

[13] Kristensen MH, Petersen S. District heating energy efficiency of Danish building typologies. Energy Build 2021;231:110602. http://dx.doi.org/10.1016/j.enbuild.2020.110602.

[14] Bergsteinsson HG, Nielsen TS, Mller JK, Amer SB, Dominković DF, Madsen H. Use of smart meters as feedback for district heating temperature control. Energy Rep 2021;7:213–21. http://dx.doi.org/10.1016/j.egyr.2021.08.153, The 17th International Symposium on District Heating and Cooling.

[15] Vandermeulen A. Quantification and optimal control of district heating network flexibility (Ph.D. thesis), KU Leuven; 2020.

[16] Stevanovic VD, Zivkovic B, Prica S, Maslovaric B, Karamarkovic V, Trkulja V. Prediction of thermal transients in district heating systems. Energy Convers Manage 2009;50(9):2167–73. http://dx.doi.org/10.1016/j.enconman.2009.04.034.

[17] Benonysson A, Bhm B, Ravn HF. Operational optimization in a district heating system. Energy Convers Manage 1995;36(5):297–314. http://dx.doi.org/10.1016/0196-8904(95)98895-T.

[18] Søgaard HT. Stochastic systems with embedded parameter variations - applications to district heating (Ph.D. thesis), Technical University of Denmark, Department of Applied Mathematics and Computer Science; 1993.
[19] Madsen H, Nielsen T, Søgaard H. Control of Supply Temperature: EFP 1323/93-07. Informatics and Mathematical Modelling, Technical University of Denmark; 1996.
[20] van der Heijde B. Optimal integration of thermal energy storage and conversion in fourth generation thermal networks (Ph.D. thesis), KU Leuven; 2019.
[21] Bacher P, Madsen H. Identifying suitable models for the heat dynamics of buildings. Energy Build 2011;43(7):1511–22. http://dx.doi.org/10.1016/j.enbuild.2011.02.005.
[22] Thilker CA, Bacher P, Bergsteinsson HG, Junker RG, Cali D, Madsen H. Non-linear grey-box modelling for heat dynamics of buildings. Energy Build 2021;252:111457. http://dx.doi.org/10.1016/j.enbuild.2021.111457.
[23] Thilker CA, Bergsteinsson HG, Bacher P, Madsen H, Calì D, Junker RG. Non-linear model predictive control for smart heating of buildings. In: E3S web of conferences, Vol. 246. EDP Sciences; 2021, p. 09005.
[24] Jazwinski A. Stochastic processes and filtering theory. San Diego, CA, USA: Academic Press; 1970.
[25] Grunnet Wang P, Scharling M, Pagh Nielsen K, Kern-Hansen C, Wittchen KB. 2001 – 2010 Danish design reference year: reference climate dataset for technical dimensioning in building, construction and other sectors. number 13-19 in DMI technical report, Danmarks Meteorologiske Institut; 2013.
[26] van der Heijde B, Fuchs M, Ribas Tugores C, Schweiger G, Sartor K, Basciotti D, Müller D, Nytsch-Geusen C, Wetter M, Helsen L. Dynamic equation-based thermo-hydraulic pipe model for district heating and cooling systems. Energy Convers Manage 2017;151:158–69. http://dx.doi.org/10.1016/j.enconman.2017.08.072.
[27] van der Heijde B, Aertgeerts A, Helsen L. Modelling steady-state thermal behaviour of double thermal network pipes. Int J Therm Sci 2017;117:316–27. http://dx.doi.org/10.1016/j.ijthermalsci.2017.03.026.
[28] Dénarié A, Aprile M, Motta M. Heat transmission over long pipes: New model for fast and accurate district heating simulations. Energy 2019;166:267–76. http://dx.doi.org/10.1016/j.energy.2018.09.186.
[29] Grosswindhager S, Voigt A, Kozek M. Linear finite-difference schemes for energy transport in district heating networks. In: Proceedings of the 2nd International Conference on Computer Modelling and Simulation; 2011, p. 5–7.
[30] Wallentén P. Steady-state heat loss from insulated pipes (Ph.D. thesis), Division of Building Physics, Lund University; 1991.
[31] Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. TMB: Automatic differentiation and Laplace approximation. J Stat Softw 2016;70(5):1–21. http://dx.doi.org/10.18637/jss.v070.i05.
[32] Madsen H, Thyregod P. Introduction to general and generalized linear models. Texts in statistical science series, CRC Press; 2010.
[33] Van Loan C. Computing integrals involving the matrix exponential. IEEE Trans Automat Control 1978;23(3):395–404. http://dx.doi.org/10.1109/TAC.1978.1101743.
[34] 5810827_Z2_GB_06.pdf. Kamstrup, 2021. URL: https://koce1-kamstrup.ocecdn.oraclecloud.com/content/published/api/v1.1/assets/CONT4AA1D9C46651406CA6AC17F2FAAFCC99/native/5810827_Z2_GB.pdf?channelToken=ed241bbb18f444908a8fc9ed97ca5d5b Accessed: 2021-12-20.