Identification of non-linear autoregressive models with exogenous inputs for room air temperature modelling

Thilker, Christian Ankerstjerne; Bacher, Peder; Cali, Davide; Madsen, Henrik

Published in:
Energy and AI

Link to article, DOI:
10.1016/j.egyai.2022.100165

Publication date:
2022

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

Energy and AI 9 (2022) 100165

Contents lists available at ScienceDirect

Energy and AI

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

Identification of non-linear autoregressive models with exogenous inputs for room air temperature modelling

Christian Ankerstjerne Thilker *, Peder Bacher, Davide Cali, Henrik Madsen

Technical University of Denmark, Department of Applied Mathematics and Computer Science, Asmussens Allé, Building 303B, DK-2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
Time series analysis
Non-linear models
District heating
Smart energy systems

ABSTRACT

This paper proposes non-linear autoregressive models with exogenous inputs to model the air temperature in each room of a Danish school building connected to the local district heating network. To obtain satisfactory models, the authors find it necessary to estimate the solar radiation effect as a function of the time of the day using a B-spline basis expansion. Furthermore, this paper proposes a method for estimating the valve position of the radiator thermostats in each room using modified Hermite polynomials to ensure monotonicity of the estimated curve. The non-linearities require a modification in the estimation procedure: Some parameters are estimated in an outer optimisation, while the usual regression parameters are estimated in an inner optimisation. The models are able to simulate the temperature 24 h ahead with a root-mean-square-error of the predictions between 0.25 °C and 0.6 °C. The models seem to capture the solar radiation gain in a way aligned with expectations. The estimated thermostatic valve functions also seem to capture the important variations of the individual room heat inputs.

1. Introduction

In Denmark, more than 65% of households are heated by district heating [1]. It is standard practice to measure the heat consumption for an individual household — to be able to bill each household for its consumption. But for an individual room in a (large) building, the heat each radiator emits is not known. Hence, room control relies only on temperature measurements. It is an interesting and relevant task to control a single room. Firstly, because rooms have different dynamics due to differences in size and heating capacity and thus require different treatment and control in order to keep them comfortably regulated. Secondly, because occupants perceive the indoor climate individually and therefore want individual settings in the rooms they use [2]. For these reasons, it is desirable to control buildings on room level.

1.1. Literature review

Predictive room-level control obviously requires temperature-models of the individual rooms. However, the popular Resistor–Capacitor-based models [3] are not possible to employ due to the missing knowledge of the heat load on room-level. We are left, then, to use models that relate less to physics. AutoRegressive with eXogenous input (ARX) models are a popular class of models for time series modelling [4,5]. ARX models are a variant of AutoRegressive and Moving Average (ARMA) models where the MA-part is left out and input-terms are added [6]. Examples of applications are solar radiation forecasting [7], wind power forecasting [8], and glucose level predictions [9]. For thermal prediction of buildings, popular black-box models include ARX and neural network models. The latter has received much attention recently [10–16]. Standard Artificial Neural Networks (ANNs) are the simplest kind of neural network model and Root Mean Square Error (RMSE) between 0.77 and 0.9 are reported [10,11]. Long short-term memory models, that are suitable neural networks for time series predictions, are also studied intensively with various variations (e.g. in combination with an error correction model or a convolutional neural network) [13,14]. A combination of grey-box models with a neural network to make correct prediction errors have also been done [12]. RMSE between 0.6 and 0.75 was reported. Conclusions regarding the best kind of model is not unanimous, with some studies finding ARX models performing [15] better and vice versa [16]. ARX models are in general much simpler compared to neural networks making them more robust toward overfitting and faster to fit [17,18]. Given time series data of the system, the optimal set of parameters in linear ARX models has a closed-form solution, equivalent to linear regression, which is fast and robust to compute [6]. Sometimes, however, non-linear models are necessary to sufficiently describe input effects due to their richer solution structures compared to linear models [19].

* Corresponding author.
E-mail addresses: chant@dtu.dk (C.A. Thilker), pbac@dtu.dk (P. Bacher), dcal@dtu.dk (D. Cali), hmad@dtu.dk (H. Madsen).

https://doi.org/10.1016/j.egyai.2022.100165
Received 19 February 2022; Received in revised form 23 April 2022; Accepted 7 May 2022
Available online 23 May 2022

Neural network models can be thought of as (very high-dimensional) non-linear variants of linear ARX models readily able to model non-linearities. The present approach, however, use the low-dimensional ARX models while including and isolating the non-linear contributions in the model. The heat dynamics between the indoor and outdoor air are well known to be linear [20,21], and therefore there is no need to use non-linear models to capture this effect. However, the solar gain can be very non-linear [22,23]. The proposed method isolates the non-linearities to the solar gain using relatively few parameters compared to neural network models. Thus, the presented approach keeps the robustness and simplicity of the ARX model, while still modelling non-linear effects. The few number of parameters and simple model structure of the proposed ARX models impose and act as regularisation in the model, which ensures robustness of the prediction capabilities. ARX models are also readily used for control purposes.

In the present paper, it is described how important non-linearities that affect the room air temperature can be modelled. The gain from solar radiation is a significant source of heat in rooms. Solar radiation forecasts are typically given as an average effect on a horizontal surface during a time interval, i.e. in the units [W/m^2]. However, the solar gain in rooms vary non-linearly throughout the day and is hence a non-linear function of time [24,25]. Another non-linearity arises from the radiator thermostats. A thermostat controls the valve (and mass flow rate in the radiators) as a PI-controller. However, it is well-known that the map from a measured temperature and set point to a valve state is a non-linear function [26,27]. Even chaos in the valve dynamics has been reported [28]. To model the entering non-linearities from the solar gain and the thermostatic valve, functions are fitted to data using B-splines and Hermite polynomials as basis functions.

1.2. Main contributions of the paper

This paper describes a method for estimating ARX-models where the estimation problem is non-linear in the parameters. Conventional methods cannot estimate parameters in such models, thus, the proposed method generalises the identification procedure for ARX models. Applying the method for indoor air temperature models enables us to estimate nonlinear effects such as varying solar radiation gain and heat inlet from the heating system in each room. The room air temperature models are able to predict the temperature over a long period with different conditions and may enable smart predictive control on room level. Results indicate that the models perform on par or better compared to the literature in terms of RMSE.

1.3. Structure and outline of the paper

Section 2 introduces the building and experimental setup carried out to obtain the data. In Section 3, ARX models are introduced and how to carry out the variable transformation and parameter estimation. Next, Section 4 introduces B-splines and Hermite polynomials for data fitting and explains how to use them for modelling inputs in the ARX models. Section 5 showcases the results from fitting ARX models to the individual room air temperatures of a Danish school building. Lastly, Section 6 concludes and sums up the findings of the paper.

2. Case study: A Danish school building

This section presents the building and the experimental setup. Also, it introduces and explains the problems related to operating the indoor climate of the individual rooms and sets the stage for the rest of the paper. It is identified how the following key heat gains of the room air temperature to include in the models:

- **The outdoor air temperature** constantly affects the indoor air temperature through walls and windows. The outer surfaces act as a low-pass filter between the indoor and outdoor air temperature.

Fig. 1. A photograph of the building site.

Fig. 2. Floor plan of the school building.

- **The solar gain** may deliver significant amounts of energy to the indoor air in short time periods by entering through windows and heating floors, walls, furniture and other materials.
- **The room radiators** are the rooms' main heating sources and are controlled by a thermostatic valve that opens according to a given set-point.

2.1. The building and rooms

The building, a school with three floors and a basement, is located in Høje Taastrup, Denmark. The uppermost floor is a partly-refurbished roof attic. A photo of the building is given in Fig. 1. Fig. 2 shows a floor plan of the building. It has 10 classrooms while fewer rooms are hallways, storage rooms, toilets, and teacher's rooms. The following paragraph gives an overview of the basic properties of the building here. For more technical details, the reader is referred to Bruun [29], Lex et al. [30].

The building was built in 1929 and thus is not insulated according to modern standards. The facade and internal walls consist of solid bricks. The windows have wooden frames and double-paned low-E glazings. Floors are made from wood joists and the roof is partly uninsulated and partly insulated slate roof. The building is connected to the local district heating (DH) system. The building uses district heating for domestic hot water and space heating. For this building, the space heating is a separate water-based circuit with dedicated pumps. Radiators of different types (cast-iron, panel convectors, plane conductors) with individual thermostats deliver the heat in the individual rooms. Each radiator has an individual thermostatic valve that automatically and individually

regulates the water flow into the radiator unit to maintain a certain air temperature, and can be remotely controlled by temperature set-points. To measure the room air temperature, each room is equipped with an air temperature sensor, placed somewhere on the inner wall in around 2–3 m height.

2.2. Some room modelling problems

Since radiators do not heat up the room in a spacial uniform manner, the measured room air temperature is only representative for the air close to the sensor. There is thus a temperature difference between the air temperature close to the thermostat and measured temperature by the sensor on the wall. Thus, the measured room temperature may react very differently to set-point increases based on room geometry, room size, air circulation, number and placement of radiators and thermostats etc. Such factors support the need for individual room air temperature models. Due to the many complicated physical factors related to modelling the thermostatic behaviour, a data-driven method is proposed using interpolating polynomials to describe the state of the radiator valves as a function of the room air temperature and the room set-point.

Another significant challenge on room level arises from the solar gain. The sun radiates large amounts of energy to rooms in buildings, which leads to a fast increase in the indoor air temperature [31]. It is thus important to model the solar gain in thermal models of buildings [32]. The solar gain pattern is individual for each room due to the individual room and window sizes and different navigational orientations of the windows. Furthermore, the solar gain changes throughout the day as the sun traverses the sky (due to Earth's rotation around itself), which is difficult to describe using regular ARX models. To describe the time-varying solar gain in each room, this paper propose a data-driven approach using B-splines as a basis expansion [33].

2.3. The experiment and data

Each room has been equipped with a sensor, which measures the indoor air temperature. The accuracy of the measurements are ±0.2 °C and are taken in resolution of 0.1 °C. All radiators in the rooms are also equipped with smart thermostats, where the operators can read and write set-points in °C. Sensors located in buildings central heating system measure the total heat load of the building and the forward temperature. See Table 1 for an overivew of the experimental data.

The experiment was designed to produce dynamical responses from the rooms to easier learn the thermal dynamics. The experiment lasted from the 1st of March, 2021, through the 27th of March 27, 2021, during which the school was only partly occupied (due to covid-19) but most rooms were used. During the experiment, individual and independent set-points were sent to each room. However, different schedules were used during day- and night time, such that the temperature was more comfortable during occupation times. The measured signals for each room are: the indoor air temperature, $T_{\mathrm{i},t}$, and the thermostatic set-point, $T_{\mathrm{set},t}$. The forward temperature of the space heating water in the building was set constant at 60 °C. The ventilation system was operating at a constant rate with inlet temperature set-point between 20 and 24 °C. The weather data is from a local weather station and is provided by the Danish Meteorological Institute.

3. ARX-models for dynamical systems

This section introduces ARX models for modelling dynamical systems. In linear ARX models, the optimal parameters are given by a closed-form solution to a linear regression problem. However, the closed-form solution exists only for models that are linear in the parameters. This section addresses this problem and proposes a method for estimating parameters that appear non-linearly in ARX models.

Table 1
Data interpretation.

Name	Quantity	Unit
$T_{\mathrm{i},t}^{j}$	Indoor air temperature in room j	[°C]
T_{t}^{for}	Building supply temperature	[°C]
$T_{j,t}^{\mathrm{set}}$	Temperature set-point in room j	[°C]
I_t	Global solar radiation	[W/m^2]
T_{t}^{a}	Outdoor air temperature	[°C]

3.1. Introduction to ARX models

Let $\{Y_t; t \in \mathbb{N}\}$, $Y_t \in \mathbb{Y} \subseteq \mathbb{R}$, be a stochastic process and let the time series $\boldsymbol{y}_N = \left(y_1, y_2, \dots, y_N\right)^\top$ be a realisation of N consecutive observations of Y_t. Let $\boldsymbol{X}_N = \left\{\boldsymbol{x}_i\right\}_{i=1}^{N}$, $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{N_x,t})^\top \in \mathbb{X}^{N_x} \subseteq \mathbb{R}^{N_x}$ be a vector time series containing the inputs to the system associated with the realisation. An ARX model of order M has the form

$$\varphi(\mathrm{B})Y_t = \boldsymbol{\beta}(\mathrm{B})^\top \boldsymbol{x}_t + \varepsilon_t\,, \tag{1}$$

where $\varphi(\mathrm{B}) = 1 + \varphi_1\mathrm{B} + \varphi_2\mathrm{B}^2 + \cdots + \varphi_M\mathrm{B}^M$ is a polynomial in the back shift operator B, $\mathrm{B}^k Y_t = Y_{t-k}$. $\boldsymbol{\beta}(\mathrm{B})^\top = (\beta_1(\mathrm{B}), \beta_2(\mathrm{B}), \dots, \beta_{N_x}(\mathrm{B}))$ is a vector with back-shift polynomials where the i'th polynomial

$$\beta_i(\mathrm{B}) = \beta_{i,1}\mathrm{B} + \beta_{i,2}\mathrm{B}^2 + \cdots + \beta_{i,M}\mathrm{B}^M, \quad i = 1, \dots, N_x \tag{2}$$

is associated with the i'th input. $\{\varepsilon_t; t \in \mathbb{N}\}$ is a white noise process where $\varepsilon_t \sim N(0, \sigma^2)$ is i.i.d. Now, let $\boldsymbol{\beta}_i = \{\beta_{i,m}\}_{m=1}^{M}$ be the set of coefficients in the i'th back-shift polynomial. Then $\{\varphi_i\}_{i=1}^{M}$ and $\{\boldsymbol{\beta}_i\}_{i=1}^{N_x}$ are regression coefficients in the ARX model in (1). Isolating Y_t,

$$Y_t = \sum_{m=1}^{M} -\varphi_m \mathrm{B}^m Y_t + \sum_{i=1}^{N_x} \beta_i(\mathrm{B}) x_{i,t} + \varepsilon_t \tag{3}$$

we obtain the system on a regression form. Then the optimal set of parameters in Eq. (3) when minimising the sum of squared errors is given by the closed form solution [6]

$$\begin{aligned}(\{\hat{\varphi}_i\}_{i=1}^{M}, \{\hat{\boldsymbol{\beta}}_i\}_{i=1}^{N_x}) = \underset{\varphi_1,\dots,\varphi_M,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_{N_x}}{\arg\min} \sum_{i=1}^{N} \varepsilon_i^2 = \\ (\boldsymbol{X}(\boldsymbol{y}_N, \boldsymbol{X}_N)^\top \boldsymbol{X}(\boldsymbol{y}_N, \boldsymbol{X}_N))^{-1} (\boldsymbol{X}(\boldsymbol{y}_N, \boldsymbol{X}_N)^\top \boldsymbol{y}_N)\,,\end{aligned} \tag{4}$$

where $\boldsymbol{X} : \mathbb{Y}^N \times \mathbb{X}^{N \times N_x} \mapsto \mathbb{R}^{N \times M(N_x+1)}$ is the so-called design matrix of Eq. (1), where each row constitute a time instance of Eq. (3).

3.2. Parameter estimation in ARX models

The overall goal of the parameter estimation procedure is to find the set of parameters that minimises the sum of squared residuals (as in Eq. (4)),

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \hat{\varepsilon}_i^2\,. \tag{5}$$

Therefore, the central objective in order to enable us to minimise (5) is to evaluate the squared residuals. Then, a numerical optimisation routine may minimise the squared residuals w.r.t. the model parameters. Since our ARX model includes parameters that are non-linearly coupled (e.g. some regression coefficients may depend non-linearly on some parameters $\boldsymbol{\theta}_{\mathrm{tr}}$, i.e. $\beta_{i,m}(\boldsymbol{\theta}_{\mathrm{tr}})$), the closed-form solution (4) cannot be evaluated directly. Instead, to evaluate the sum of squared residuals (5), the paper propose the following two-stage procedure:

- Transformation stage: Here, transform the regressors such that they appear linearly in the ARX model (given the set of transformation variables).
- Regression stage: Here, insert the transformed regressors into the regression form in (3) to compute the regression coefficients using (4), which afterwards allows us to evaluate the sum of squared residuals.

Performing the two above steps returns the sum of squared residuals, which may be minimised by a numerical optimisation routine to find the optimal set of parameters. This reduces the size of the numerical optimisation to a subset of the model parameters compared to letting all parameters be optimised by numerical means.

3.2.1. Step one: Transformation stage

The regular ARX-model in (1) is linear in the regressor coefficients and does not directly allow terms that are non-linear in the parameters. In order to include non-linear terms in the ARX model, the input variables are *transformed* using a desired non-linear transformation needed in the model. Let $\boldsymbol{U}_t = (\boldsymbol{u}_{1,t}, \dots, \boldsymbol{u}_{N_x,t})^\top \in \mathbb{U}^{N\times N_x}$ be the "raw" input to the model, e.g. horizontal global solar radiation. Then, define the *transformed input variables* by

$$x_{i,t} = h_i(u_{i,t}, y_t, t, \boldsymbol{\theta}_{\text{tr}}), \quad i = 1, \dots, N_x, \tag{6}$$

where $\boldsymbol{\theta}_{\text{tr}} \in \mathbb{P}_{\text{tr}} \subseteq \mathbb{R}^{N_{\text{tr}}}$ are input-specific parameters that may enter into e.g. a basis expansion such as a spline basis, a polynomial basis, a Fourier basis etc., $h_i : \mathbb{U}\times\mathbb{Y}\times\mathbb{R}_+\times\mathbb{P}_{\text{tr}} \mapsto \mathbb{X}_i \subseteq \mathbb{R}$ is the (in general) non-linear transformation for the i'th input. Note that the variable $x_{i,t}$ defines a new regressor as a function of the input $u_{i,t}$. But now, a regular ARX model, which is linear in the new regressors, $x_{i,t}$, can be written

$$\begin{aligned} &\varphi(\mathrm{B})Y_t = \boldsymbol{\beta}(\mathrm{B})^\top \boldsymbol{x}_t + \varepsilon_t\,, \\ \implies \quad & Y_t = \sum_{m=1}^{M} -\varphi_m \mathrm{B}^m Y_t + \sum_{i=1}^{N_x} \beta_i(\mathrm{B}) x_{i,t} + \varepsilon_t\,, \\ \implies \quad & Y_t = \sum_{m=1}^{M} -\varphi_m \mathrm{B}^m Y_t + \sum_{i=1}^{N_x}\sum_{m=1}^{M} \beta_{i,m} \mathrm{B}^m x_{i,t} + \varepsilon_t\,. \end{aligned} \tag{7}$$

where $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{N_x,t})^\top$. Note that if h_i does not depend on any parameters, i.e. if $\boldsymbol{\theta}_{\text{tr}} = \emptyset$, the model simplifies to a regular ARX model that is linear in the regression coefficients.

3.2.2. Step two: Parameter estimation in the transformed regression variables

If one writes (7) for all observations, y_t, it may be write on the following matrix–vector form

$$\boldsymbol{Y}_N = \boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}}; \boldsymbol{X}_N, \boldsymbol{y}_N)\boldsymbol{\theta}_{\text{reg}} + \boldsymbol{\varepsilon}_N\,, \tag{8}$$

where $\boldsymbol{X} : \mathbb{P}_{\text{tr}}\times\mathbb{X}^{N\times N_x}\times\mathbb{Y}^N \mapsto \mathbb{X}_m^{N\times M(N_x+1)}$, $\boldsymbol{X}_N = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)^\top$, is a function that returns the design matrix of the linear regression problem in (7) as a function of $\boldsymbol{\theta}_{\text{tr}}$. Eq. (8) is now linear in the coefficients $\{\varphi_m\}_{m=1}^M$ and $\{\boldsymbol{\beta}_i\}_{i=1}^{N_x}$. It can then be interpreted as a linear regression problem with $\boldsymbol{\theta}_{\text{reg}} = (-\varphi_1, \dots, -\varphi_M, \beta_{1,1}, \dots, \beta_{N_x,M})^\top \in \mathbb{P}_{\text{reg}} \subseteq \mathbb{R}^{M(N_x+1)}$ as the regression coefficients and y_i and $\boldsymbol{x}_i$ as regressors.

Eq. (8) appears as a linear regression problem w.r.t. $\boldsymbol{\theta}_{\text{reg}}$. It is immediately evident that the design matrix depends on $\boldsymbol{\theta}_{\text{tr}}$. Therefore, consider the case where $\boldsymbol{\theta}_{\text{tr}}$ is given and fixed, then the optimal estimator of $\boldsymbol{\theta}_{\text{reg}}$ when minimising the sum of squared residuals in (4) is

$$\hat{\boldsymbol{\theta}}_{\text{reg}}(\boldsymbol{\theta}_{\text{tr}}) = \left(\boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}})^\top \boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}})\right)^{-1} \left(\boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}})^\top \boldsymbol{Y}\right) \tag{9}$$

where $\boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}}) = \boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}}; \boldsymbol{X}_N, \boldsymbol{y}_N)$ is simply a short-hand notation.

3.2.3. Optimisation of transformation parameters

The above implies that for each given value of $\boldsymbol{\theta}_{\text{tr}}$, the optimal set of $\boldsymbol{\theta}_{\text{reg}}$ is given in a closed form. With $\boldsymbol{\theta}_{\text{tr}}$ and $\boldsymbol{\theta}_{\text{reg}}$ at hand, the sum of squared residuals can be evaluated. Thus, one can write a function alone in $\boldsymbol{\theta}_{\text{tr}}$, which 1) transforms the regressors, 2) solves the arising linear regression problem, and 3) evaluates the sum of squared residuals. The optimisation problem in the transformation variables become:

$$\hat{\boldsymbol{\theta}}_{\text{tr}} = \arg\min_{\boldsymbol{\theta}_{\text{tr}}} \quad \sum_{i=1}^{N} \hat{\varepsilon}(\boldsymbol{\theta}_{\text{tr}})_i^2 \tag{10}$$

where $\varepsilon(\boldsymbol{\theta}_{\text{tr}})_i$ is the i'th prediction error as a function of $\boldsymbol{\theta}_{\text{tr}}$. Any numerical optimiser may be used to solve this problem. Practically, is done by defining a function that takes as input $\boldsymbol{\theta}_{\text{tr}}$, computes $\boldsymbol{\theta}_{\text{reg}}$, and in turn computes and returns the sum of squared errors. Algorithm 1 outlines the framework for evaluating (5). The function is fed into a numerical optimiser. The `ipopt` solver through `Python` was used.

Algorithm 1 Evaluation of sum of squared residuals

require: $\boldsymbol{\theta}_{\text{tr}}$, $\boldsymbol{y}_N$, $\boldsymbol{U}_N$

given $\boldsymbol{\theta}_{\text{tr}}$, $\boldsymbol{y}_N$, and $\boldsymbol{U}_N$, compute transformed input variables, $x_{i,t}$, and construct design matrix $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\theta}_{\text{tr}}, \boldsymbol{X}_N, \boldsymbol{y}_N)$

compute regression variables $\hat{\boldsymbol{\theta}}_{\text{reg}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{y}_N)$

compute sum of squared residuals, $\text{SSE} = \sum_i \varepsilon_i^2 = (\boldsymbol{y}_N - \boldsymbol{X}\hat{\boldsymbol{\theta}}_{\text{reg}})^\top(\boldsymbol{y}_N - \boldsymbol{X}\hat{\boldsymbol{\theta}}_{\text{reg}})$

return SSE

4. Curve estimation and interpolation using B-splines and Hermite polynomials

This section introduces the modelling techniques applied for modelling non-linear effects to expand inputs and parameters into basis functions. Curve fitting using splines or polynomials is usually a problem of estimating coefficients to *basis functions*. The basis functions typically form a linear function space. The goal of the estimation is to find the element from this space, $\boldsymbol{x}$, that minimises some distance between $\boldsymbol{x}$ and the data points. This section shows how to model the solar radiation and the thermostatic valve function using B-splines and Hermite-polynomials expansion basis functions, respectively [34].

4.1. de Boor B-splines

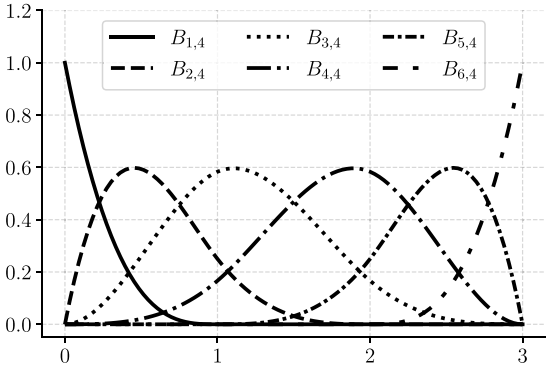
Due to the properties of B-splines, they have proved useful in many data fitting applications [35–37]. de Boor B-splines [38] are defined on a finite interval $[a, b] \subseteq \mathbb{R}$ by a recursion formula via a set of control points $\Delta : a = t_0 \le t_1 \le \cdots \le t_{N_c} = b$ and a polynomial degree k. This uniquely defines a set of non-decreasing knot placements $a = z_0 \le z_1 \le \cdots \le z_{N_c+k+1} = b$, where the i'th B-spline of order k is given by

$$B_{i,k} = \frac{z - z_i}{z_{i+k} - z_i} B_{i,k-1} + \frac{z_{i+k+1} - z}{z_{i+k+1} - z_{i+1}} B_{i+1,k-1}\,, \tag{11}$$

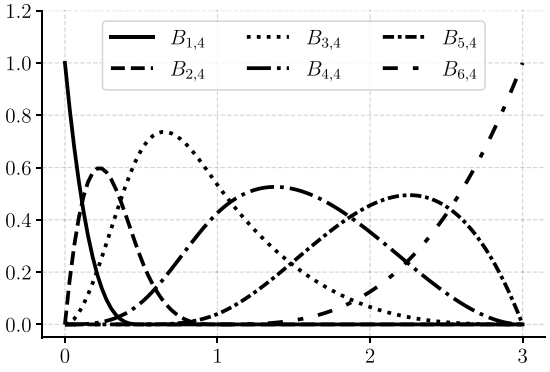
where $B_{i,0}(z) = \chi_{[z_i, z_{i+1}[}(z)$. Fig. 3(a) shows a set of six fourth order de Boor B-splines with equidistant control points $\Delta = (0, 1, 2, 3)$ on the domain $[0, 3]$. The estimation problem then comes down to estimating the scaling coefficients in a linear combination of basis B-splines to form the estimated function.

4.1.1. Estimation of B-spline knot placements

In addition to the scaling coefficients, it is possible to optimise over the placements of the control points. One strategy to place the control points is to put them according to the quantiles of the data. This way, the control points are put according to the amount of data in the state space. This strategy, however, does not account for the amount of curvature of the true function. By freely optimising the control point placements, splines are able to "move closer" in parts of their domain where more fluctuating and fast-changing dynamics occur. Define the parameters $\boldsymbol{\tau} = \{\tau_k\}_{k=0}^{N_c-2}$ as the distances between interior knots t_i and t_{i+1} for $k = 0, \dots, N_c - 2$ (minus two comes from the last distance being uniquely given by the first $N_c - 1$ distances). Fig. 3(b) shows the de Boor B-splines as in Fig. 3(a), where now the control points are $\Delta = (0, 1/2, 1, 3)$. By moving the two central knots to the left in the domain, the "density" of the spline variation has also moved. This can be useful to capture large curvatures of the data.



(a) Fourth order de Boor B-splines (that are piece-wise cubic polynomials) with control points $\Delta = (0, 1, 2, 3)$.



(b) Fourth order de Boor B-splines (that are piece-wise cubic polynomials) where the control points are shifted to $\Delta = (0, 1/2, 1, 3)$. This shifts the variations of the splines to the left.

Fig. 3. de Boor's B-splines with uniform and non-uniform knot placements.

4.2. Estimation of the solar gain using B-splines

The solar radiation is very fluctuating and is governed by seasonal, diurnal, and hourly variations [39]. It becomes even more complex when estimating the solar gain of a single room of a building since it depends much on the size and orientation of the window(s) of the room. It is very common to estimate models with a constant solar gain parameter e.g. on the form

$$\beta_1(\mathrm{B})I_t = \beta_{1,1}I_{t-1} + \beta_{1,2}I_{t-2} + \cdots + \beta_{1,M}I_{t-M}\,, \tag{12}$$

where I_t is the global solar radiation (typically given by a third party source) at time t and $\{\beta_{1,m}\}_{m=1}^{M}$ are time-independent parameters. As an example, imagine a room with a single window pointing toward east. In that case, solar radiation enters the room in the morning and disappears as the sun travels around south on the sky. The above model formulation (12) is not able to catch such variable gains. Instead, one can expand the coefficients in (12) in a basis formed by a linear combination of B-splines and define a transformed input variable by the expansion

$$x_{I,t} = h_I(I_t, t, \boldsymbol{\alpha}, \boldsymbol{\tau}) = (\alpha_1 B_1^{\boldsymbol{\tau}}(t) + \alpha_2 B_2^{\boldsymbol{\tau}}(t) + \cdots + \alpha_{N_c+k} B_{N_c+k}^{\boldsymbol{\tau}}(t))I_t\,, \tag{13}$$

where $B_n^{\boldsymbol{\tau}}(t)$ is the n'th B-spline basis function evaluated at time t with given knot displacements $\boldsymbol{\tau}$. $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^{N_c+k}$ is the associated basis spline coefficients. Here, N_c is a tuning paratemeter set by the user. The model in (12) can thus be expanded to

$$\beta_1(\mathrm{B})x_{I,t} = \sum_{m=1}^{M} \beta_{1,m} h_I(I_{t-m}, t-m, \boldsymbol{\alpha}, \boldsymbol{\tau})\,. \tag{14}$$

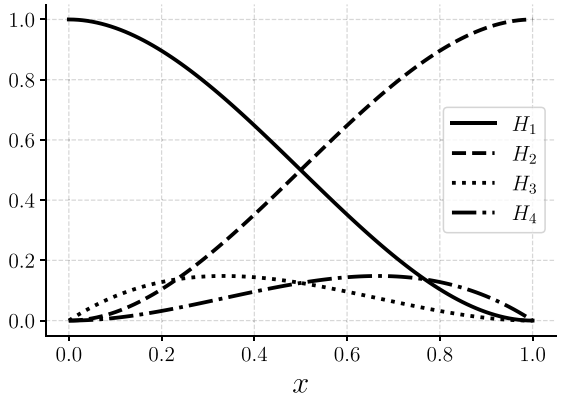


Fig. 4. The Hermite basis functions as given in Fritsch and Carlson [40] on the interval $[a, b] = [0, 1]$.

Given $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$, Eq. (14) is linear in the transformed regressors. Thus, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ belong to the parameters related to the transformed regressors $\boldsymbol{\theta}_{\mathrm{tr}}$, and $\{\beta_{1,m}\}_{m=1}^{M}$ are the regression coefficients.

4.3. Estimation of the thermostatic control function using Hermite polynomials

Thilker et al. [41] propose the following function to describe the valve state of a radiator as a function of the measured room temperature and the thermostat set-point T_t^{set},

$$f_t^{\mathrm{valve}} = \frac{1}{1 + \exp\left(-\alpha(T_t^{\mathrm{set}} + T_{\mathrm{offset}} - T_t^{\mathrm{i}})\right)}\,. \tag{15}$$

This simple model is characterised by two parameters $(\alpha, T_{\mathrm{offset}}) \in]0, \infty[\times\mathbb{R}$ and has a relatively fixed shape that mimics the intuitive behaviour of thermostats: It acts as a PI-controller, which opens when it is too cold and closes when it is too warm (in a continuous way). The advantage of the model in Eq. (15) is its simplicity and few number of parameters. The disadvantage, however, is that the function's shape might be too restrictive to capture the actual room-specific behaviour. The thermostatic behaviour depends on many parameters such as room size, number of radiators, radiator placements etc. To simplify the complicated thermostatic modelling while keeping the function reasonably constrained, it is proposed to use a Hermite polynomial basis to fit a curve to the valve states as a function of indoor air temperature and the set-point.

In these settings, the estimation problem of the thermostatic valve function comes down to estimating coefficients in a basis expansion. But monotonicity in the solution is also required since it is natural to think that the valve opens monotonically as the relative temperature difference increases.

4.3.1. Cubic interpolation with Hermite polynomials

Let the following be given: an interval $[a, b] \subset \mathbb{R}$, function values at the end points $(a, f(a))$ and $(b, f(b))$, and derivatives at the end points $f'(a)$ and $f'(b)$. Consider then the problem of finding a polynomial, P, of degree three with $P(a) = f(a), P'(a) = f'(a), P(b) = f(b), P'(b) = f'(b)$. A solution to this problem has the form [42]

$$P(x) = f(a)H_1(x) + f(b)H_2(x) + f'(a)H_3(x) + f'(b)H_4(x)\,, \tag{16}$$

where H_i are the usual cubic Hermite basis functions (given in e.g. Fritsch and Carlson [40]; see Fig. 4).

Consider now the partition $\Delta : a = x_1 < x_2 < \cdots < x_{N_v} = b$ over the interval $[a, b]$ with associated function values and derivatives for all partition points. This forms an interpolating polynomial between each sub-interval $[x_i, x_{i+1}]$ by using the solution in (16). If one further constrain the function values $\{f_k\}_{k=1}^{N_v}$ to be monotonic increasing, $0 \leq$

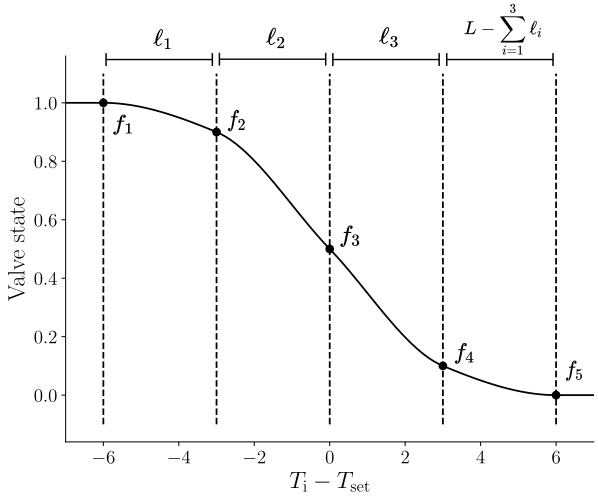


Fig. 5. An illustration of the expanded basis formed by piece-wise Hermite interpolations of the valve state function. Each interval of length ℓ_i, $i = 1, \dots, N_v$, consists of cubic Hermite polynomials that interpolates the interval between function values. The Hermite polynomials have been modified to be monotonic between the function values, which makes the overall function monotonic due to the constraint $f_i \leq f_{i+1}$.

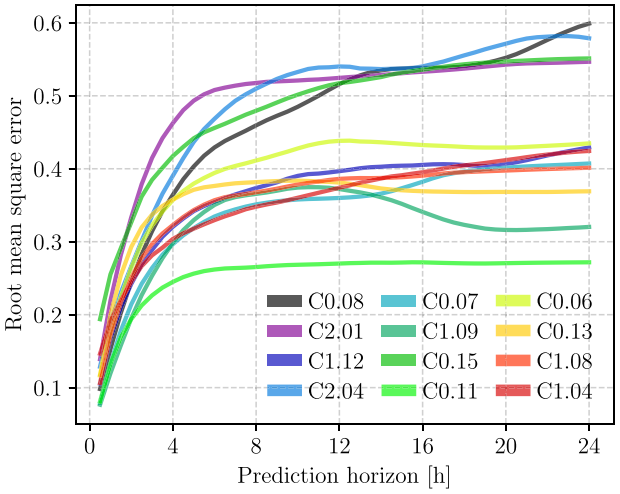


Fig. 6. The RMSE of the temperature predictions as a function of the prediction horizon for the rooms. Most rooms reaches a point where the RMSE becomes rather flat after 6–8 h. The performance of the individual model varies a bit: a cluster of rooms whose RMSE is significantly higher compared to the rest that constitute its own cluster.

$f_1 \leq f_2 \leq \cdots \leq f_{N_v} \leq 1$, a monotonic interpolation over the entire interval $[a, b]$ is guaranteed. [40] describes an algorithm for computing interpolating polynomials that are monotonic in each sub-interval. However, data is usually given simply by $(x_i, f(x_i))$, thus a procedure to compute the derivatives at each partition point is also needed. Algorithms for this also exists, see e.g. [40]. We can now optimise over the partition placements and the associated function values to estimate a monotonic increasing curve given by

$$\begin{aligned}\hat{P}(x) =& \hat{f}_i H_1(x) + \hat{f}_{i+1} H_2(x) + \hat{f}'_i H_3(x) + \hat{f}'_{i+1} H_4(x), \\ & x \in [x_i, x_{i+1}[\,.\end{aligned} \tag{17}$$

To perform the piece-wise cubic interpolation, i.e. compute $\hat{P}$ in (17), the authors use `PchipInterpolator` from the `SciPy` library in `Python`. To optimise over the partition points, the lengths of the sub-intervals by $\ell_i = x_{i+1} - x_i$ are defined as parameters. Fig. 5 depicts and illustrates the ideas of this estimation scheme. The parameters related to estimate the thermostatic valve function are $(\ell_1, \dots, \ell_{N_v-2}, f_2, \dots, f_{N_v-1})$ since the end points are fixed to $f_1 = 1$ and $f_{N_v-1} = 0$. The transformed regressors are

$$x_{\text{valve},t} = h_{\text{valve}}(T_{\text{set},t}, y_t, \boldsymbol{\ell}, \boldsymbol{f}) = P_{\boldsymbol{\ell},\boldsymbol{f}}(y_t - T_{\text{set},t})(T_{\text{for}} - y_t), \tag{18}$$

where $P_{\boldsymbol{\ell},\boldsymbol{f}}$ is given by (17), and the subscripts indicates the polynomial's dependence on the parameters. The term $(T_{\text{for}} - y_t)$ is the temperature difference between the supply (forward) water in the heating system of the building and the room air temperature. This term is then multiplied by "how open the valve in the radiator is". The regression model related to the heat input becomes

$$\beta_2(\mathrm{B}) x_{\text{valve},t} = \sum_{m=1}^{M} \beta_{2,m} h_{\text{valve}}(T_{\text{set},t-m}, y_{t-m}, \boldsymbol{\ell}, \boldsymbol{f}). \tag{19}$$

Again, given $\boldsymbol{\ell}$ and $\boldsymbol{f}$, Eq. (19) is linear in the transformed regressors. $\boldsymbol{\ell}$ and $\boldsymbol{f}$ then belongs to $\boldsymbol{\theta}_{\text{tr}}$ and $\{\beta_{2,m}\}_{m=1}^{M}$ are the regression coefficients.

5. Modelling results

This section presents and discusses the modelling results and quantifies their performance.

5.1. The final non-linear ARX model

Based on the previous section, the following ARX model describes the indoor air temperature, denoted y_t, based on inputs from the outdoor air temperature, solar radiation, and radiators

$$\begin{aligned}&\varphi(\mathrm{B}) Y_t = \beta_1(\mathrm{B}) x_{I,t} + \beta_2(\mathrm{B}) x_{\text{valve},t} + \beta_3(\mathrm{B})(T_{a,t} - Y_t) + \varepsilon_t, \\ \implies\ & Y_t = \sum_{m=1}^{M} -\varphi_m \mathrm{B}^m y_t + \sum_{m=1}^{M} \beta_{1,m} \mathrm{B}^m x_{I,t} \\ & \quad + \sum_{m=1}^{M} \beta_{2,m} \mathrm{B}^m x_{\text{valve},t} + \sum_{m=1}^{M} \beta_{3,m} \mathrm{B}^m (T_{a,t} - Y_t) + \varepsilon_t.\end{aligned} \tag{20}$$

The models are estimated using data sampled every 30 min. The authors found that the order $M = 3$ was optimal for almost all rooms based on information criterions such as AIC or BIC (one room model was optimal for $M = 2$).

5.2. The estimation problem

We write up the optimisation problem as outlined in Section 3.

$$\min_{\boldsymbol{\theta}_{\text{tr}}} \quad \sum_{i=1}^{N} \hat{\varepsilon}(\boldsymbol{\theta}_{\text{tr}})_i^2 \tag{21a}$$

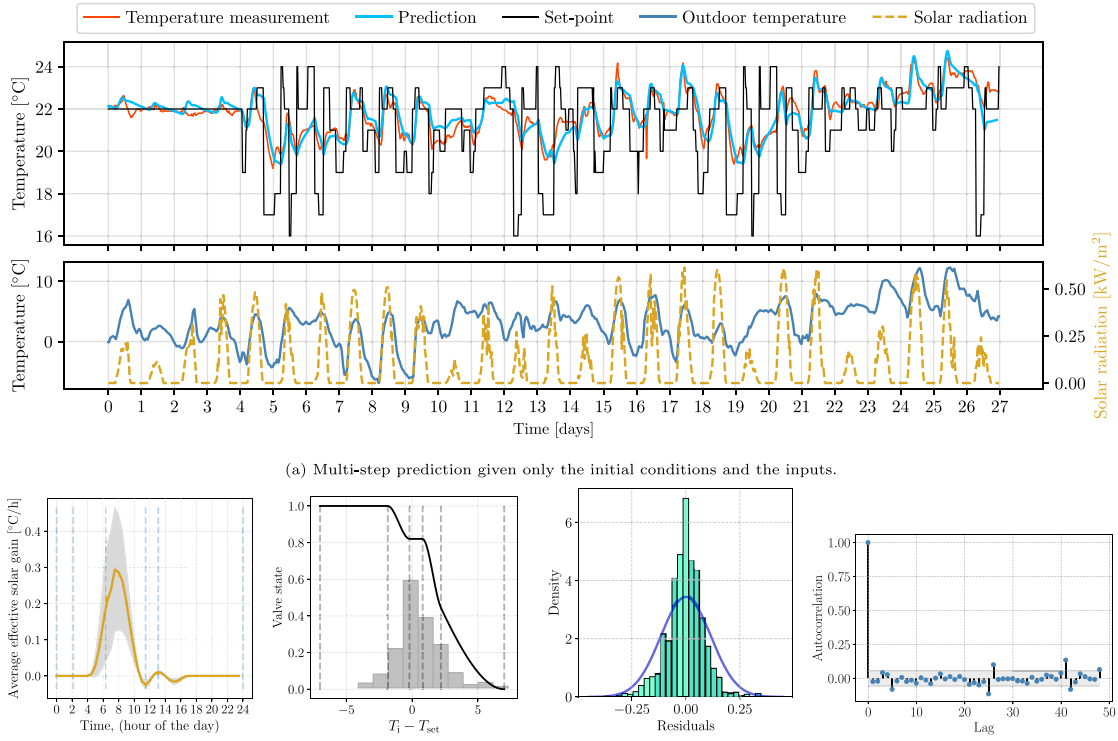
$$\text{s.t.} \quad \boldsymbol{\theta}_{\text{tr}} = \left[\boldsymbol{\alpha}^\top, \boldsymbol{\tau}^\top, \boldsymbol{f}^\top, \boldsymbol{\ell}^\top\right]^\top \tag{21b}$$

$$\boldsymbol{A}\boldsymbol{\theta}_{\text{tr}} \leq \boldsymbol{b}, \tag{21c}$$

where the sum in (21a) is computed by Algorithm 1. The in-equality constraints $\boldsymbol{A}\boldsymbol{\theta}_{\text{tr}} \leq \boldsymbol{b}$ specifies potential constraints that might be needed on the parameters. For instance, the valve function values are required to be monotonically decreasing, $f_i \geq f_{i+1} \implies \begin{bmatrix}-1 & 1\end{bmatrix} \begin{bmatrix} f_i \\ f_{i+1} \end{bmatrix} \leq 0$. Small minimal distances between knot points in the B-spline and Hermite representations are also formulated as constraints. Keep in mind that the optimisation problem is not necessarily convex and thus a numerical solver may find local optimal solutions. However, the authors did not find that the solution was sensitive to variations in the initial condition.

5.3. Room air temperature model validation

Fig. 6 shows the RMSE of the individual room air models for each room as a function of the prediction horizon. For day-ahead predictions,



(a) Multi-step prediction given only the initial conditions and the inputs.

(b) Solar gain throughout the day. The vertical dashed lines indicates the placement of the knots.

(c) Thermostatic valve function. The vertical dashed lines indicates the placement of the knots. The histogram shows the distribution of the data points.

(d) Histogram of residuals and the associated Gaussian distribution.

(e) Autocorrelation function of the one-step ahead residuals.

Fig. 7. Modelling results of room C0.06.

the rooms deliver an average RMSE of around 0.4° C with the highest being 0.6° C and the lowest being 0.25° C. Most rooms lie in the range of 0.25° C-0.45° C. However, four rooms seem to form a cluster that performs worse compared to the rest. This section investigates the resulting models of three of the rooms. The rest is omitted due to an otherwise large space use.

5.3.1. Room C0.06

Fig. 7 shows the results related to the model of room C0.06. To present an example of data and a model simulation to get an intuition of the forecasting abilities, Fig. 7(a) shows a multi-step simulation of the model for the room. Based on the first four days, where the set-point is constant, the model seems to catch the offset of the measured room temperature and the thermostat set-point well. Also, the model does a proper job at catching the exponential decay and increase, when the set-point is lowered or raised. However, there seems to be some more random dynamics where the model struggles; around day 10, the temperature drops much faster than predicted and around day 15/16, the temperature suddenly increases. These random dynamics may be due to occupancy. Lastly, it is also worth noting that the model seems to be "stable" in time; it does not drift away from the measured trajectory.

Fig. 7(b) displays the estimated average solar gain during a day based on the estimated B-spline representation. The shape of the estimated curve is large during the morning hours and close to zero otherwise. This indicates that the sun has no influence in the afternoon, which is in line with the fact that the windows point toward east and only receives sun during the morning (see Fig. 2).

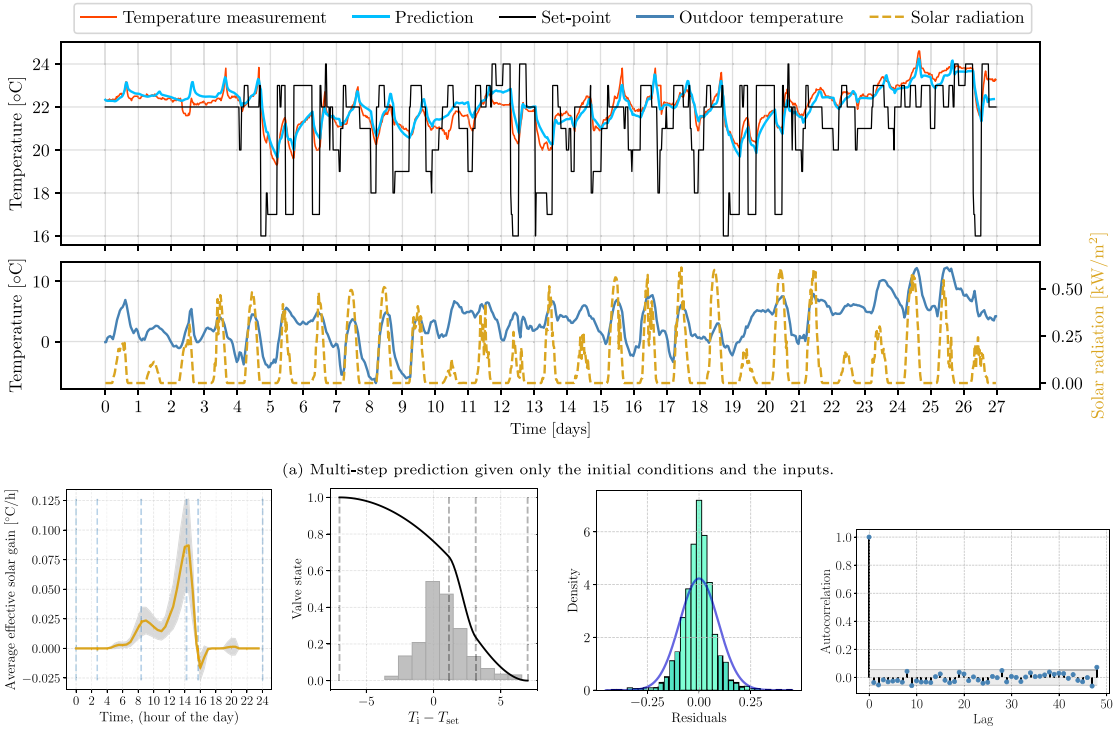
The estimated thermostatic valve function is displayed in Fig. 7(c). It suggests that the valve is almost fully open at 1°C above the set-point and closes fast below that point. Notice that the knots are centred around the data and where the estimated curvature is high.

Figs. 7(d) and 7(e) show the histogram and autocorrelation of the 1-step prediction residuals. The shape of the histogram is close to a Gaussian distribution, however its tails seem to be too wide. This could be due to a non-constant variance in either time or space. To model spacial variation in the error variance, a transformation might be suitable to mitigate this. For time-varying variance, using e.g. a variant of the generalised autoregressive conditional heteroskedasticity (GARCH) model may be used [43]. The autocorrelation function shows only a few small significant lags. These may be due to e.g. periodic occupant behaviour or some other dynamics that are not captured properly. However, the small magnitude of the significant lags indicates that the model captures the dynamics well overall.

5.3.2. Room C0.08

Fig. 8 shows the results related to the model of room C0.08. Fig. 8(a) shows a multi-step prediction of the model for the room. The dynamics for this room have spikes in the measured temperature. Looking at 8(b), it is immediately evident, that the estimated solar gain peaks in the afternoon. This also supports that the room windows point toward west and sees the sun in the evening. After proving the exact placement of the sensor in the room, it is found that the temperature sensor is placed such that it is hit by direct solar radiation. Hence, the measured air temperature does not fully reflect the actual room air temperature during the evening. However, the model still does a good job at catching these spikes. Looking at Fig. 6, C0.08 has a large RMSE which may arise from the difficulty of predicting the spikes. Nevertheless, the model overall seems to catch the temperature offset well together with the exponential decays/increases.

The estimated thermostatic valve function is displayed in Fig. 8(c). It suggests a more linear shape of the valve function compared to room C0.06 with a small flat step around 0. This shape may correspond more to the expected shape the valve state in an ideal thermostat [26]. Again the control points are placed around the largest parts of the density and where the largest estimated curvature is.



(a) Multi-step prediction given only the initial conditions and the inputs.

(b) Solar gain throughout the day. The vertical dashed lines indicates the placement of the knots.

(c) Thermostatic valve function. The vertical dashed lines indicates the placement of the knots. The histogram shows the distribution of the data points.

(d) Histogram of residuals and the associated Gaussian distribution.

(e) Autocorrelation function of the one-step ahead residuals.

Fig. 8. Modelling results of room C0.08.

Figs. 8(d) and 8(e) show the histogram and autocorrelation of the 1-step prediction residuals. Again, the shape of the histogram is close to a Gaussian distribution, but again the tails seem to widen the distribution too much. The autocorrelation function shows essentially no significant lags. The spikes in the temperature due to the solar radiation hitting seems to be captured quite well, since it could have shown in the autocorrelation if it was not the case. This implies that the model captures the dynamics well overall.

5.3.3. Room C1.09

Fig. 9 presents the results from the model of room C1.09. Fig. 9(a) shows a multi-step prediction for the room. Again, based on the first four days, where the set-point is constant, the model seems to catch the offset of the measured room temperature and the thermostat set-point well. Also, the model does a good job at catching the exponential decay and increase, when the set-point is lowered or raised. C1.09 seems to be governed by less random dynamics compared to the other rooms, which makes the fit better (in terms of RMSE, see Fig. 6). The reason hereof is, first, that the room seems not to be occupied. Second, the magnitude of the estimated average solar gain in Fig. 9(b) is small. This small contribution of the solar radiation matches the fact that the room points toward north and less to none solar radiation enters during the day.

The estimated thermostatic valve function is displayed in Fig. 9(c). Its shape is simple and suggests that the valves close completely at 1°C above the set-point and opens fast below that point. It could indicate that the room temperature reaction to the set-point is consistent.

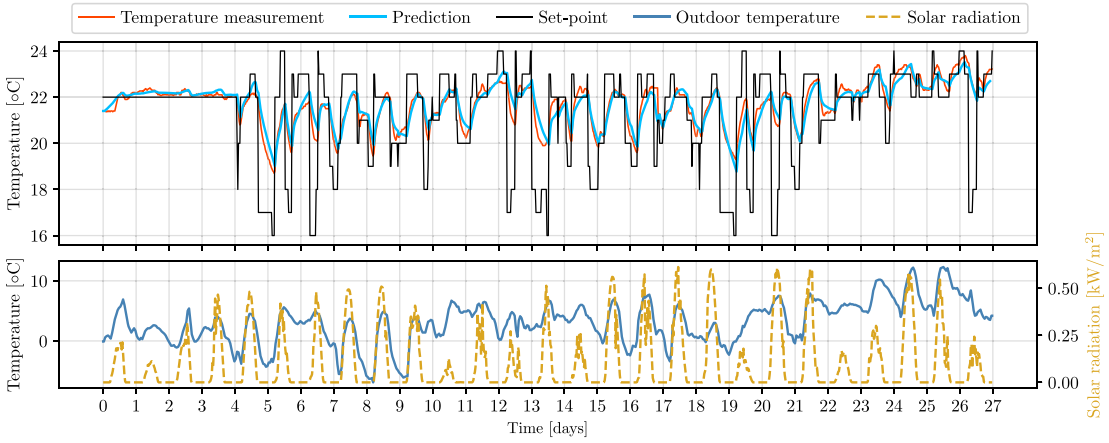
Figs. 9(d) and 9(e) show the histogram and autocorrelation of the 1-step prediction residuals respectively. The shape of the histogram is more narrow compared to the other rooms (and again too large tails), and the autocorrelation function is insignificant. The good fit may again be due to the few disturbances and few random fluctuations in the room temperature.

5.4. Discussion and summary

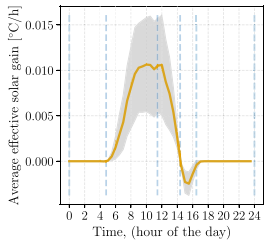
The overall picture from the results is, that the models are very suitable for temperature predictions for many hours ahead. The RMSE indicates that the performance is on par or better compared to the state-of-the-art in the literature. And at the same time, their consistency in terms of RMSE emphasises the robustness of the identification. The indoor air temperature has significant diurnal variations, which are captured well. The solar radiation effect is successfully isolated by the model and is easily interpreted and is in alignment with physics. The estimated valve functions turned out being significantly different for each room, which confirms the need for room-specific air temperature models in general. A possibly significant effect neglected in the models presented in this paper, is the heat transmission between rooms. However, to model this effect requires all models to be estimated simultaneously, and to identify all neighbouring rooms. This is left for further studies. The flexibility of the valve function is visible from the results and captures what seems to be sudden changes in the dynamics. E.g. for room C1.09 where the valve closes fast when lowering the set-point below the measured temperature.

5.4.1. Potential energy savings in optimal control

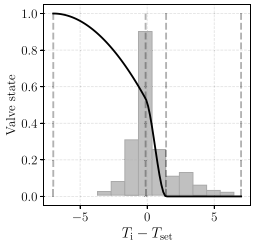
The presented ARX models do not have knowledge of the heat load of the individual rooms. This makes direct control of the heat usage infeasible. However, the second regression term in (20) describes the increase in air temperature caused by the radiators, $\sum_m \beta_{2,m} \mathrm{B}^m x_{\mathrm{valve},t}$. Therefore, instead of using the heat load in an objective function in an economic MPC, the air temperature increase as in (20) may be used instead. In such a setup, the room temperatures respond to a price signal instead of the heat load. But this produces the same result under the assumption that the air temperature increase and the heat load are significantly correlated. A possible optimal control problem for each



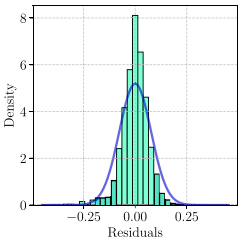
(a) Multi-step prediction given only the initial conditions and the inputs.



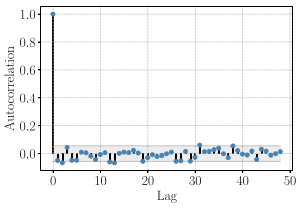
(b) Solar gain throughout the day. The vertical dashed lines indicates the placement of the knots.



(c) Thermostatic valve function. The vertical dashed lines indicates the placement of the knots. The histogram shows the distribution of the data points.



(d) Histogram of residuals and the associated Gaussian distribution.



(e) Autocorrelation function of the one-step ahead residuals.

Fig. 9. Modelling results of room C1.09.

room could look like the following

$$\min_{T_{\text{set},t+k}} \quad \sum_{k=0}^{N-1} c_k \cdot x_{\text{valve},t+k} \tag{22a}$$

$$\text{s.t.} \quad \text{Eq. (20)} \tag{22b}$$

$$T_{\min,t+k+1} \leq \hat{Y}_{t+k+1} \leq T_{\max,t+k+1} \tag{22c}$$

where c_k is a time-varying price signal, and $T_{\min,t+k+1}$ and $T_{\max,t+k+1}$ are upper and lower temperature bounds. In such a control setup, the price in the objective function, c_k, is related to the temperature increase in the room at time t_k, and not the heat usage. However, due to the presumably high correlation between the two variables, the optimal control sequences of the two problems may be highly correlated. Further work on this control strategy is needed to clarify its potential.

6. Conclusion and future work

This paper presented room air temperature models for individual rooms of a Danish school building, and identified non-linearities in the system arising from the time-varying solar gain and the heat input from the radiators. It is proposed to model the varying solar gain by a B-spline basis expansion and the thermostatic valve state using Hermite-polynomials that guarantees monotonicity of the function. The proposed ARX model for each room was consequently non-linear in the parameters, which required us to perform a two-stage identification procedure to estimate these parameters.

However, the individual room temperature models have no knowledge of the their individual heat usage, which is necessary for e.g. peak shaving or load shifting optimisation. To estimate and control the heat consumption, one needs a model of the entire building (since the heat consumption on building level is known). An interesting future work direction could be toward using the individual room models to predict the entire building heat load. With such a model, it is possible to enable flexible control of the building while considering each room's needs.

The individual room temperature models may also be used for fault detection and live diagnostics. For instance, to identify if a room responds slowly or not at all to a set-point increase. They can also be used to identify outliers in operations, to identify e.g. bad occupant behaviour or if a valve breaks etc. In an online monitoring and reporting setup, this may help alert building operators and identify what rooms to pay attention to in order to improve the indoor climate in the building and optimise operations.

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.

Acknowledgements

The authors received funding from the following projects; *Sustainable plus energy neighbourhoods (syn.ikia)* (H2020 No. 869918), FME-ZEN (Research Council of Norway - 257660), Top-up (Innovation Fund Denmark 9045-00017B), SCA+ (Interreg Öresund-Kattegat-Skagerrak) and *Flexibile Energy Denmark (FED)* (IFD 8090-00069B).

References

[1] Fjernvarme D. Fakta om fjernvarme. 2020, URL https://www.danskfjernvarme.dk/presse/fakta-om-fjernvarme.

[2] Cali D, Kindler E, Ebrahimy R, Bacher P, Hu K, Østrup M, et al. climify.org: an online solution for easy control and monitoring of the indoor environment. Vol. 111, E D P Sciences; 2019, p. 05006. http://dx.doi.org/10.1051/e3sconf/201911105006, Clima 2019: 13th REHVA World Congress ; Conference date: 26-05-2019 Through 29-05-2019.

[3] 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, URL https://www.sciencedirect.com/science/article/pii/S0378778811000491.

[4] Bacher P, Madsen H, Nielsen HA. Online short-term solar power forecasting. Sol Energy 2009;83(10):1772–83. http://dx.doi.org/10.1016/j.solener.2009.05.016, URL https://www.sciencedirect.com/science/article/pii/S0038092X09001364.

[5] Ríos-Moreno G, Trejo-Perea M, Castañeda-Miranda R, Hernández-Guzmán V, Herrera-Ruiz G. Modelling temperature in intelligent buildings by means of autoregressive models. Autom Constr 2007;16(5):713–22. http://dx.doi.org/10.1016/j.autcon.2006.11.003, URL https://www.sciencedirect.com/science/article/pii/S0926580506001208.

[6] Madsen H. Time series analysis. Chapman & Hall; 2007, http://dx.doi.org/10.1201/9781420059687.

[7] Bacher P, Madsen H, Nielsen H. Online short-term solar power forecasting. Sol Energy 2009;83(10):1772–83. http://dx.doi.org/10.1016/j.solener.2009.05.016.

[8] Durán MJ, Cros D, Riquelme J. Short-term wind power forecast based on ARX models. J Energy Eng 2007;133(3):172–80. http://dx.doi.org/10.1061/(ASCE)0733-9402(2007)133:3(172).

[9] Romero-Ugalde HM, Garnotel M, Doron M, Jallon P, Charpentier G, Franc S, et al. ARX model for interstitial glucose prediction during and after physical activities. Control Eng Pract 2019;90:321–30. http://dx.doi.org/10.1016/j.conengprac.2019.07.013, URL https://www.sciencedirect.com/science/article/pii/S0967066119301121.

[10] Mba L, Meukam P, Kemajou A. Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Energy Build 2016;121:32–42. http://dx.doi.org/10.1016/j.enbuild.2016.03.046, URL https://www.sciencedirect.com/science/article/pii/S0378778816302006.

[11] Thomas B, Soleimani-Mohseni M. Artificial neural network models for indoor temperature prediction: Investigations in two buildings. Neural Comput Appl 2007;16:81–9. http://dx.doi.org/10.1007/s00521-006-0047-9.

[12] Cui B, Fan C, Munk J, Mao N, Xiao F, Dong J, et al. A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses. Appl Energy 2019;236:101–16. http://dx.doi.org/10.1016/j.apenergy.2018.11.077, URL https://www.sciencedirect.com/science/article/pii/S0306261918317938.

[13] Xu C, Chen H, Wang J, Guo Y, Yuan Y. Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. Build Environ 2019;148:128–35. http://dx.doi.org/10.1016/j.buildenv.2018.10.062, URL https://www.sciencedirect.com/science/article/pii/S0360132318306875.

[14] Elmaz F, Eyckerman R, Casteels W, Latré S, Hellinckx P. CNN-LSTM architecture for predictive indoor temperature modeling. Build Environ 2021;206:108327. http://dx.doi.org/10.1016/j.buildenv.2021.108327, URL https://www.sciencedirect.com/science/article/pii/S0360132321007241.

[15] Ferracuti F, Fonti A, Ciabattoni L, Pizzuti S, Arteconi A, Helsen L, et al. Data-driven models for short-term thermal behaviour prediction in real buildings. Appl Energy 2017;204:1375–87. http://dx.doi.org/10.1016/j.apenergy.2017.05.015, URL https://www.sciencedirect.com/science/article/pii/S0306261917305032.

[16] Delcroix B, Le Ny J, Bernier M, Azam M, Qu B, Venne J-S. Autoregressive neural networks with exogenous variables for indoor temperature prediction in buildings. Build Simul 2020;14. http://dx.doi.org/10.1007/s12273-019-0597-2.

[17] Sarwar R, Cho H, Cox SJ, Mago PJ, Luck R. Field validation study of a time and temperature indexed autoregressive with exogenous (ARX) model for building thermal load prediction. Energy 2017;119:483–96. http://dx.doi.org/10.1016/j.energy.2016.12.083, URL https://www.sciencedirect.com/science/article/pii/S0360544216318898.

[18] Yun K, Luck R, Mago PJ, Cho H. Building hourly thermal load prediction using an indexed ARX model. Energy Build 2012;54:225–33. http://dx.doi.org/10.1016/j.enbuild.2012.08.007, URL https://www.sciencedirect.com/science/article/pii/S0378778812003933.

[19] Scott A. The nonlinear universe. Berlin, Heidelberg: Springer; 2007.

[20] Andersen KK, Madsen H, Hansen LH. Modelling the heat dynamics of a building using stochastic differential equations. Energy Build 2000;31(1):13–24. http://dx.doi.org/10.1016/S0378-7788(98)00069-3, URL https://www.sciencedirect.com/science/article/pii/S0378778898000693.

[21] Andersen P, Jiménez M, Madsen H, Rode C. Characterization of heat dynamics of an arctic low-energy house with floor heating. Build Simul 2014;7(6):595–614. http://dx.doi.org/10.1007/s12273-014-0185-4.

[22] Zhang X, Ritosa K, Saelens D, Roels S. Comparing statistical modeling techniques for heat loss coefficient estimation using in-situ data. J Phys Conf Ser 2021;2069(1):012101. http://dx.doi.org/10.1088/1742-6596/2069/1/012101.

[23] Hollick FP, Gori V, Elwell CA. Thermal performance of occupied homes: A dynamic grey-box method accounting for solar gains. Energy Build 2020;208:109669. http://dx.doi.org/10.1016/j.enbuild.2019.109669, URL https://www.sciencedirect.com/science/article/pii/S0378778819314999.

[24] Csáky I, Kalmár F. Effects of solar radiation asymmetry on buildings' cooling energy needs. J Build Phys 2016;40(1):35–54. http://dx.doi.org/10.1177/1744259115597444.

[25] Chwieduk D, Bogdanska B. Some recommendations for inclinations and orientations of building elements under solar radiation in Polish conditions. Renew Energy 2004;29(9):1569–81. http://dx.doi.org/10.1016/j.renene.2003.12.018, URL https://www.sciencedirect.com/science/article/pii/S0960148104000096.

[26] Hansen LH. Stochastic modelling of central heating systems. (Ph.D. thesis), Technical University of Denmark; 1997, URL http://www2.imm.dtu.dk/pubdb/pubs/2460-full.html.

[27] Zou X, Jordan J, Shillor M. A dynamic model for a thermostat. J Eng Math 1999;291–310. http://dx.doi.org/10.1023/A:1004587425961.

[28] Hayashi S, Hayase T, Kurahashi T. Chaos in a hydraulic control valve. J Fluids Struct 1997;11(6):693–716. http://dx.doi.org/10.1006/jfls.1997.0096, URL https://www.sciencedirect.com/science/article/pii/S0889974697900967.

[29] Bruun CG. Optimization of building operation using high-resolution sensor data. (Master's thesis), Department of Civil Engineering: Technical University of Denmark; 2019, URL https://findit.dtu.dk/en/catalog/2448777767.

[30] Lex SW, Calí D, Koed Rasmussen M, Bacher P, Bachalarz M, Madsen H. A cross-disciplinary path to healthy and energy efficient buildings. Technol Forecast Soc Change 2019;142:273–84. http://dx.doi.org/10.1016/j.techfore.2018.07.023, URL https://www.sciencedirect.com/science/article/pii/S0040162518301999, Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges.

[31] Thilker CA, Madsen H, Jørgensen JB. Advanced forecasting and disturbance modelling for model predictive control of smart energy systems. Appl Energy 2021;292:116889. http://dx.doi.org/10.1016/j.apenergy.2021.116889, URL https://www.sciencedirect.com/science/article/pii/S0306261921003755.

[32] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014;37:517–37. http://dx.doi.org/10.1016/j.rser.2014.05.056, URL https://www.sciencedirect.com/science/article/pii/S1364032114003815.

[33] Rasmussen C, Frölke L, Bacher P, Madsen H, Rode C. Semi-parametric modelling of sun position dependent solar gain using B-splines in grey-box models. Sol Energy 2020;195:249–58. http://dx.doi.org/10.1016/j.solener.2019.11.023, URL https://www.sciencedirect.com/science/article/pii/S0038092X19311235.

[34] Christensen O. Functions, spaces, and expansions. Birkhauser Basel; 2010, http://dx.doi.org/10.1007/978-0-8176-4980-7.

[35] Bruce PD, Kellett MG. Modelling and identification of non-linear aerodynamic functions using B-splines. Proc Inst Mech Eng G 2000;214(1):27–40. http://dx.doi.org/10.1243/0954410001531890.

[36] Meyer K. Random regression analyses using B-splines to model growth of Australian angus cattle. Genet Select Evol 2005;37. http://dx.doi.org/10.1186/1297-9686-37-6-473.

[37] Sekhar Roy S, Roy R, Balas VE. Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. Renew Sustain Energy Rev 2018;82:4256–68. http://dx.doi.org/10.1016/j.rser.2017.05.249, URL https://www.sciencedirect.com/science/article/pii/S1364032117308961.

[38] de Boor C. On calculating with B-splines. J Approx Theory 1972;6(1):50–62. http://dx.doi.org/10.1016/0021-9045(72)90080-9, URL https://www.sciencedirect.com/science/article/pii/0021904572900809.

[39] Thilker CA, Junker RG, Bacher P, Jørgensen JB, Madsen H. Model predictive control based on stochastic grey-box models. In: Ploix S, Amayri M, Bouguila N, editors. Towards energy smart homes: algorithms, technologies, and applications. Cham: Springer International Publishing; 2021, p. 329–80. http://dx.doi.org/10.1007/978-3-030-76477-7_11.

[40] Fritsch FN, Carlson RE. Monotone piecewise cubic interpolation. SIAM J Numer Anal 1980;17(2):238–46, URL http://www.jstor.org/stable/2156610.

[41] 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, URL https://www.sciencedirect.com/science/article/pii/S0378778821007416.

[42] Burden RL, Faires JD. Numerical analysis. Boston, USA: Cengage Learning; 1989.

[43] Lamoureux CG, Lastrapes WD. Persistence in variance, structural change, and the GARCH model. J Bus Econom Statist 1990;8(2):225–34. http://dx.doi.org/10.1080/07350015.1990.10509794, URL https://www.tandfonline.com/doi/abs/10.1080/07350015.1990.10509794.