Low Order Grey-box Models for Short-term Thermal Behavior Prediction in Buildings

Article in Energy Procedia · May 2017
DOI: 10.1016/j.egypro.2017.03.592

CITATION
1

READS
44

5 authors, including:

Alessandro Fonti
Università Politecnica delle Marche
15 PUBLICATIONS 93 CITATIONS

Stefano Pizzuti
ENEA
67 PUBLICATIONS 330 CITATIONS

Alessia Arteconi
Università Telematica "E-Campus"
45 PUBLICATIONS 721 CITATIONS

Some of the authors of this publication are also working on these related projects:

- H2020-EE-04: Innova MicroSolar View project
- Computational Energy Management View project
Low order grey-box models for short-term thermal behavior prediction in buildings

Alessandro Fonti*, Gabriele Comodi*, Stefano Pizzutib, Alessia Arteconic, Lieve Helsen

Abstract

Low order grey-box models are suitable to be used in predictive controls. In real buildings in which the measured quantities are few the reliability of these models is crucial for the control performance. In this paper an identification procedure is analyzed to investigate the accuracy of different order grey-box models for short-term thermal behavior prediction in a real building, part of a living smart district. The building has a low number of zones and a single indoor temperature measuring point. The models are identified on the data acquired in 31 days during the winter 2015. The second order model shows the best performance with a root-mean-square error (RMSE) less than 0.5°C for a prediction horizon of 1-hour and a RMSE less than 1°C for a prediction horizon of 3-hours.

© 2016 The Authors. Published by Elsevier Ltd.

Keywords: Grey-box modeling; Reduced order models; Building energy modeling

1. Introduction

Model Predictive Control (MPC) has shown good results for achieving higher energy efficiency in buildings [1]. Such control strategy necessarily needs thermal building models [1] and the reliability of these models is crucial for the control performance [2]. Reduced order grey-box models are particularly suitable in predictive control as they combine building physics and model structure knowledge typical of the white-box approach with the parameter estimation through measured data of a black-box approach [3].

In this work an identification procedure is performed to investigate the accuracy of low order models for short-term thermal behavior prediction in a real building with a low number of zones and a single indoor temperature measuring point. The lumped element models analyzed are derived directly from Reynders et
al. [4], that proposed different reduced-order models, identified on simulation data obtained using Modelica IDEAS library [5]. Parameters of different order models are here identified using Matlab through measures of: indoor and outdoor air temperature, thermal and electrical load power and global horizontal irradiance. Indeed, the main contribution of the present study is the use in the identification process of data acquired in a real and living building rather than simulated data, as often done in literature. This helps to highlight practical issues, otherwise neglected, arising from the identification process in real applications.

2. Methods

A smart building that belongs to a living smart district has been considered as case study. Data obtained from sensors installed in the building are used to train and validate three grey-box models of increasing order. Different identification parameters have been analyzed to select the model that shows the best performance.

2.1. The Smart Village and the F70 building

The Smart Village is a small smart district located within the ENEA Research Centre in Rome and comprises 8 single floor buildings (progressively numbered F66÷F73) together with the thermal plant (F85). The buildings are intended for office use and are connected each other and to the common thermal plant by a heating-cooling thermal network.

Fig. 1a presents an overview of the Smart Village. For every building there is a remotely controlled three-port valve which allows to control the thermal fluid flow rate. The thermal plant outlet temperature and flow rate are fixed. Each office in the building is heated by a fancoil equipped with a hysteresis thermostat and there are no fancoils in the corridors. For every building measures of corridor temperature, global thermal power, difference between inlet and outlet supply water temperature, active and reactive electric power are available.

![Fig.1. The Smart Village (a) and the F70 plan (b). The F70 has a floor area of 363 m².](image)

The identification process carried out in this study focuses on the building F70. Fig.1b shows the F70 plan and the red asterisk indicates the position of the only temperature measuring point in the corridor. It has been assumed as hypothesis that the temperature detected along the corridor is closely related to the temperatures of the adjacent zones. This aspect will be further analyzed in the following sections.

2.2. Lumped element models

Resistance–capacitance (RC) lumped element models for representing the building are here analyzed. Fig. 2 shows the three considered models and reports the parameters and the input-output quantities. It reports also the equations of the 2nd order model which is the best identified model as shown in the results.
In the first order model (Fig. 2a) the entire thermal mass of the building is lumped to a single capacity and no distinction has been made between the structural mass and the indoor air mass. The 2nd order model (Fig. 2b) takes into account this difference by including a second capacity. The 3rd order model (Fig. 2c) instead, has 3 different capacities for the envelope ($C_{we}$), the internal walls ($C_{wi}$) and the indoor air ($C_i$) [4].

For these models the inputs are: the outdoor air temperature ($T_e$), solar gains ($Q_s$), internal gains ($Q_u$) and heating gains ($Q_h$). The observation variable is the indoor air temperature ($T_a$). The indoor air temperature is obtained by the sensor installed in the corridor, the heating gains are measured through the thermal power meter and the outdoor temperature measures are retrieved by the weather station.

As proposed in Reynders et. al. [4], the building electrical power is used as an alternative input for the effective user gains and the global horizontal irradiance from the weather station is used as an alternative to the effective solar gains which are difficult to obtain. For all model orders the solar gains, internal gains and heating gains are distributed over the capacities. The distribution coefficients are assumed to be constant and are identified as part of the parameter identification process.

The models in Fig. 2 are parametrized and so can be treated as grey-box models. A grey-box model consists of a set of continuous stochastic differential equations formulated in a state space form together with an output equation as follow [4,6]:

$$\begin{align*}
\dot{X}(t) &= A(\theta)X(t) + B(\theta)U(t) + \sigma(\theta)d\omega \\
Y(t) &= C(\theta)X(t) + D(\theta)U(t) + \epsilon
\end{align*}$$

$X(t)$ is the state vector of the dynamic system. $U(t)$ is a vector containing the measured inputs of the system and $\omega$ is a Wiener process. The measured output of the system $Y(t)$ is given as a function of the states $X(t)$ and the inputs $U(t)$. $\epsilon$ is the measurement error. The parameters $\theta$ are estimated using Matlab.

### 2.3. Identification process and data

The data used for the system identification process are retrieved from the Smart Village centralized database. Such database collects the sensor measurements coming from all the cluster buildings and the thermal plant with a fixed sample time of 900s. Measured data have been available since September 2014. Useful data for the identification are selected considering periods in which there were no data outliers and no missing data. From such selection two useful datasets are obtained: Train-set from 19 January 2015 to 7 February 2015 (20 days) and Test-set from 8 January 2015 to 18 January 2015 (11 days). Train-set has been used as training dataset while Test-set has been used as validation and test dataset in order to validate the identified models and to investigate a possible model overfitting.

The parameter identification procedure has been conducted using the greyest function in Matlab. This function leads to the maximum likelihood estimates and uses three different algorithms as search method.
for the iterative parameter estimation: the Gauss-Newton direction, the Levenberg-Marquardt and the steepest descent gradient search method. At each iteration greyest chooses the search method to obtain the highest reduction in error.

After identification, the models have been validated and tested by the following parameters:

- the root-mean-square errors (RMSE-values) at 1 step;
- the RMSE-values 1 hour-ahead and 3 hours-ahead;
- the final prediction errors (FPE);
- the level of fit (FIT);
- the auto-correlation of the residuals.

The 1-step prediction errors correspond to the residuals obtained by the method that is used to estimate the parameters and thus indicate the goodness of fit. In addition, the RMSE-values for 1-hour and 3-hours ahead predictions on the Train-set and Test-set are useful to evaluate possible model overfitting and to quantify the uncertainty that can be expected in MPC applications. The FPE-values are Akaike's final prediction errors and measure model quality as well. According to Akaike's theory, the most accurate model has the smallest FPE [7]. FIT-values are the percent normalized root mean square errors and thus these summarize in percentage the model goodness of fit (similarly to RMSE). Finally, the level of the auto-correlation in the residuals shows if the model explains well the dynamics contained in the dataset and if some dynamics of the real process cannot be taken into account by the low order analyzed model.

3. Results

In this section the results of the system identification process are discussed. Table 1 presents the identification parameters values calculated on both training and test datasets and for each model order.

The minimum value of the RMSE for the 1-step prediction on Train-set is obtained in the second and third order models. Thus increasing the model complexity beyond the second order does not produce a significant improvement in accuracy. Also analyzing the values of the same parameter calculated on the Test-set, the minimum value is reached with the second order model. Therefore the RMSE behavior indicates clearly overfitting for the third order model. Overfitting is also confirmed by the values of the FPE parameter which reaches the minimum for each dataset in the second order model. The second order model is therefore the best choice in terms of accuracy and complexity, although it can be noticed that even the simple first order model presents an accuracy not so much lower.

Table 1. Identification parameters calculated on both training and test datasets and for each model order. The values highlighted are the best values obtained. The second order model shows the best performance.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Dataset</th>
<th>Test Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>RMSE 1 step (°C)</td>
<td>0.154</td>
<td>0.148</td>
</tr>
<tr>
<td>RMSE 1 hour (°C)</td>
<td>0.491</td>
<td>0.442</td>
</tr>
<tr>
<td>RMSE 3 hours (°C)</td>
<td>1.116</td>
<td>0.896</td>
</tr>
<tr>
<td>FPE ($10^3$)</td>
<td>23.92</td>
<td>22.33</td>
</tr>
<tr>
<td>FIT 1 step (%)</td>
<td>94.24</td>
<td>94.46</td>
</tr>
</tbody>
</table>

In the following, the second order model prediction capabilities are analyzed. Fig. 3 presents a comparison between the measured indoor temperature and the same quantity predicted by the model with a prediction horizon of 1-hour (a) and 3-hours (b) on the Test-set.

The model performance is good for the 1-hour ahead prediction with a RMSE less than 0.5°C for both the datasets and pretty good in case of longer horizon with a RMSE close to 1°C. FIT-values are also
adequate since they are higher than 94% in all cases. The worst correspondence occurs during the weekends, when the thermal power is zero, and during working hours when there are temperature peaks.

Looking at Table 1, it is possible to observe a slightly better performance of all the models on the Test-set rather than on the Train-set. In a grey-box model this can happen when there are behaviors of the real process or approximations in the measurement system not taken into account in the model.

In the present identification process different reasons causing such effect can be recognized:
- Not predicted windows and doors opened manually by users (bad-behaviors)
- Sample time too wide for the thermal power measures
- Single point of temperature measurement may not be sufficiently representative of the adjacent zones temperatures.

In Fig. 3 indoor temperature rapid changes are indicated through black (solid) arrows and red (dotted) arrows. The rapid changes indicated by the black arrows occur at the early morning and are due to windows and doors opening by the cleaning service. The ones indicated through red arrows are due to the same bad-behavior by users. The sample time of 900s for the thermal power may be too wide in a building equipped with fancoils. Indeed, if fancoil switching times are too fast, related measures may be affected by a major aliasing. In this case part of the thermal power dynamic is lost and this cannot be taken into account by the identified models. Moreover, a single measuring point in the corridor for the indoor temperature may not be sufficient to describe the offices thermal behavior. Further proper measures would be needed to verify if this assumption was adequate for the performed analysis. However, the lack of a wide set of sensors in real applications can frequently occur and it represents a typical practical issue to be faced during process identification in reality. Finally, Fig. 4 shows the auto-correlation of residuals for all order models calculated with a lag of 25. Dotted lines indicate a 99% limit of confidence. The autocorrelation levels in the second and third order models are closer to the confidence limit, meaning that these models describe better the building dynamics in the dataset. The autocorrelation plots for the second and third order models are pretty good. The points out of the confidence limits in these cases are related to the behaviors not represented by the models, as discussed above.

4. Conclusion and future work

In this work three lumped element grey-box models of first, second and third order have been identified and validated on measured data coming from a living building, part of a smart district. The identification process of the unknown parameters has been conducted through Matlab. The analyzed error indices point out the best accuracy of the second order model. The RMSE values are less than 0.5°C for 1-hour prediction and close to 1°C for 3-hour prediction. The second order model can be used therefore in MPC applications in which it is sufficient to have a so short-term predictive horizon and accuracy. The residual autocorrelation analysis shows that the second order model describes pretty well the building dynamics but underlines also
that there are behaviors and approximations that the model cannot take into account. User bad-behaviors, measured signal aliasing and low correlation between the measured temperature and the temperatures of the represented thermal zones can make the identification difficult and even give a poor result in terms of accuracy. In order to obtain longer prediction horizon (i.e. 1-day ahead) with a good accuracy, future works will be targeted to deeply analyze and describe these sources of error. In particular, 1) it shall be evaluated the possibility of introducing new measuring points and utilize the average of these values as a reference temperature; 2) the measure sample time shall be reduced properly; 3) user bad-behaviors should be modelled and predicted. Finally the application of the studied grey-box model in an real MPC is foreseen.

Acknowledgements

The data used in this study were kindly provided by the Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Casaccia Research Centre, Rome.

References


Biography

Alessandro Fonti received the M.S. degree in electronic and automation engineering in 2010 and the Ph.D. degree in industrial engineering in 2016 from Università Politecnica delle Marche, Italy. Currently, he is a postdoctoral research fellow in the department of industrial engineering and mathematical sciences at Università Politecnica delle Marche.