Modeling and Identification of a Large Multi-Zone Office Building

Samuel Prívara, Zdeněk Váňa, Dimitrios Gyalistras, Jiří Cigler, Carina Sagerschnig,
Manfred Morari, Lukáš Ferkl

Abstract— **Predictive control in buildings has undergone an intensive research in the past years. Model identification plays a central role in a predictive control approach. This paper presents a comprehensive study of modeling of a large multi-zone office building. Many of the common methods used for modeling of the buildings, such as a detailed modeling of the physical properties, RC modeling, etc., appeared to be unfeasible because of the complexity of the problem. Moreover, most of the research papers dealing with this topic presents identification (and control) of either a single-zone building, or a single building sub-system. On contrary, we proposed a novel approach combining a detailed modeling by a building-design software with a black-box subspace identification. The uniqueness of the presented approach is not only in the size of the problem, but also in the way of getting the model and interconnecting several computational and simulation tools.**

I. Introduction

Climate changes, diminishing world supplies of the "traditional" fuels, ecological as well as economical aspects are only some of the many factors of a huge effort of today to save energy. Besides significant focus on renewable energy sources broaden, the goals can be reached only if the energy consumption is optimized. As the buildings account for about 40% of total final energy consumption (and its amount has been increasing at a rate 0.5–5 % *per annum* in developed countries [1]), the efficient building climate control can significantly contribute to reduction of the power consumption as well as the greenhouse gas emissions. Energy savings with minimal additional cost can be achieved by improvement of building automation system (BAS), which can nowadays control both heating, ventilation and air conditioning (HVAC) systems and the blind positioning and lighting systems [2], [3].

One of the control strategies suitable for building automation is the *Model Predictive Control* (MPC); unfortunately, the modeling and identification is rather difficult and time-consuming, not only in MPC. The special requirement for MPC is that the model should be reasonably simple and have good prediction properties on the control-relevant frequency range (see e.g. [4], [5], [6], [7]). One approach is to use the first-principle models (see [8], [9]), which are often used on systems such as TRNSYS, EnergyPlus (EP), ESP-r, etc., but these models are not explicit and cannot be used for control directly. The alternative is to use statistically-based, i.e. data-driven models [10]; in this approach, problems with sufficient excitation of the system modes arise.

In this work, we combined the benefits of both above mentioned approaches. A physical model in a building simulation software was created, such that it describes the real building as close as possible. Then identification signals were fed into the simulation software to obtain the high-quality identification data, and consequently these were used for obtaining a suitable control-oriented model.

The main contribution of this paper is twofold: Firstly, it presents in a detail the unique two step modeling procedure (real building $\rightarrow$ EnergyPlus model $\rightarrow$ linear-time invariant model for control), secondly it handles set-up of a large variety of tools used in different communities to deal with a problem of extraordinary size.

This paper is structured as follows: In the following section we will describe the problem and introduce the basic setup. Section III deals with identification and modeling procedures, describes the tools and algorithms used for obtaining the model. Section IV provides the results of the presented approach. Section V concludes the paper.

II. Problem description and setup

A. Description of the building

The 20 000 m^2 office building has six floors above ground. For this study, the entire third floor (as depicted in Figure 1), which is representative for all office floors, was modeled. Based on usage, façade orientation and HVAC supply, the floor can be divided into 24 zones which are mutually interconnected. Most zones are used as open-space offices; for modeling reasons, single offices were always lumped to a bigger zone.

The total floor area of the simulation model is approx. 2 800 m^2. The façade of the building has a window-to-wall ratio of approx. 70 %. Façades to the atrium have a glazing ratio of approx. 50 %. Roughly 50 % of the windows have interior blinds, remaining blinds are in-between-glass blinds of double windows.

S. Prívara and J. Cigler are with Department of Control Engineering (DCE), Faculty of Electrical Engineering (FEE) of Czech Technical University (CTU) in Prague, Czech Republic and with Automatic Control Laboratory (ACL), Department of Information Technology and Electrical Engineering (DITEE), ETH Zurich, Switzerland samuel.privara@fel.cvut.cz

Zdeněk Váňa and Lukáš Ferkl are with DCE, FEE CTU in Prague, Czech Republic

Carina Sagerschnig is with Gruner AG, Gellertstrasse 55, CH-4020 Basel, Switzerland

Dimitrios Gyalistras and Manfred Morari are with Automatic Control Laboratory, DITEE, ETH Zurich, Switzerland

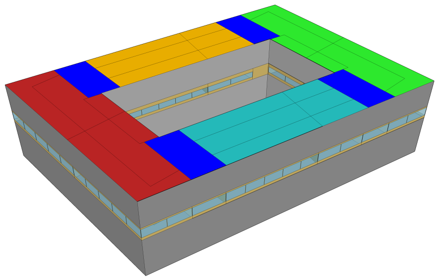


Fig. 1. 3D simulation model of the building: Investigated zones were on the third floor, other floors are greyed out. The zone layout is shown on top of the model for clarity. Zones of the same sub-system are colored alike. The core zones enabling the decoupling are dark blue.

There are the following actuators installed in the building:

- Convectors: individual convector control is possible.
- Radiant ceiling panels for cooling and heating: for control purposes ceiling panels of the floor are grouped into 24 zones that are controlled independently of each other.
- Ventilation: There are two air handling units (AHU)s – for the north, and the south. The temperature of supply air can be set independently in both AHUs.
- Venetian blinds are available for all windows in all zones. Controllable blinds of individual windows within the same zone are grouped together as one control input.

Energy supply, i.e. hot and chilled water supply for the entire building, is provided by a central heating and cooling plant, which is located partly in the basement and partly on the roof. District heating is used for the building's heat supply. Chilled water is provided locally by mechanical chillers.

B. *Choice of a modeling strategy*

As already stated, one of the objectives of this project was to find a convenient MPC-oriented modeling strategy suited for buildings, which would balance accuracy with the design-time demand.

The first possible approach is based on detailed physical modeling, represented by e.g. equivalent RC-network [11]. Unfortunately, fitting of parameters of differential equations is infeasible for large-scale problems [12].

The second approach is based purely on measurements collected during the building operation, which are used for input-output statistical identification. Even though this procedure looks simple, the results are oftentimes far from good – some important assumptions, such as persistent excitation [13], are nearly always violated during building's normal operation. The identification procedure can be improved by including some prior information [14] or by carrying out the identification experiment on the building which would excite all important system modes. Depending on the building size, the experiment might be rather expensive, but can bring high improvements to the resulting model [10].

When the building is brand new, real data are not available. From the aforementioned discussion, the only possible approach might be modeling of the building using RC network, which is quite challenging for multiple zone buildings.

Therefore a new approach had to be introduced, which yields a model of a large multi-zone building. A very promising strategy might be a combination of a building simulation software (to have an implicit model of the building) used for identification experiments to get data for a standard statistical identification procedure. We used EP as the building simulation software and Building Controls Virtual Test Bed (BCVTB) as the middleware between EP and a controller written in Matlab (in terms of excitation signal generator).

C. *Software tools*

A widely used tool for building energy performance simulation is EnergyPlus by the Lawrence Berkeley National Laboratory, which can be used for thermal load simulation and energy analysis of buildings. Besides the simulation itself, EP has a built-in energy management system that allows for integration of a rule-based control. Traditionally, EP is a stand-alone simulation engine which processes text-based input files. For developing and testing MPC models, co-simulation was necessary to allow more flexible simulation input. Co-simulation describes the integration of different tools by run-time coupling. This allows for example to couple building energy performance simulation tools to Matlab, and thus provide new possibilities to building simulation. Co-simulation fundamentals for building simulation such as coupling strategies and data transfer are described in [15]. Extensive capabilities for coupling simulation tools are provided by the Building Controls Virtual Testbed (BCVTB) [16]. It is a middleware tool that allows to couple different simulation programs for distributed simulation. Programs to be linked via the BCVTB are EP, Matlab, Modelica and Radiance. Data exchange with BACnet building automation systems is also featured. The BCVTB plays a master role in the data exchange, as depicted in Figure 2. For the entire simulation study, hourly weather data for Munich were used. The statistical weather data used were provided by the weather database of the US Department of Energy and prepared by ASHRAE[1] based on International Weather for Energy Calculations (IWEC) data.

III. Identification and modeling

As was mentioned in previous sections, the identification and modeling is one of the most demanding tasks. We will describe the whole procedure of getting a building model in the following steps. Firstly, we describe the choice of suitable inputs and outputs for the identification, then we present software tools needed for handling and keeping all

[1]American Society of Heating, Refrigerating and Air-Conditioning Engineers

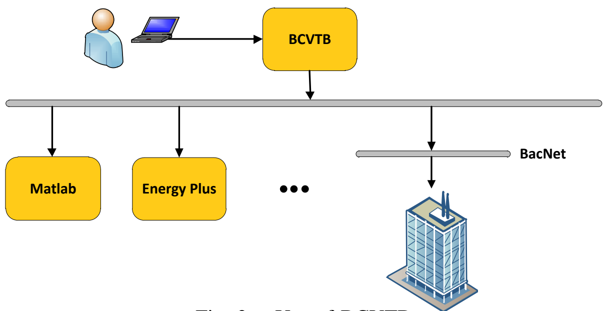


Fig. 2. Use of BCVTB

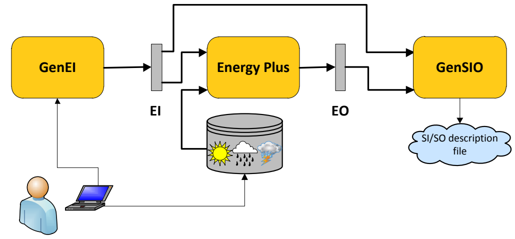


Fig. 3. Preparation of data for identification.

information about system consistent, and finally mathematical tools necessary for successful system identification (SID).

A. Choice of model inputs and outputs

The choice of model inputs and outputs plays an important role for the particular identification procedure. According to the physical relationship between chosen inputs and outputs, one should opt for a suitable procedure which is able to handle underlying physics. In other words: if the input-output relation is non-linear, then linear identification methods may fail. The size of the problem is also quite an important factor – especially in the presented problem.

Based on the aforementioned observations, we decided to choose heat fluxes affecting zone temperatures as system inputs and temperatures as outputs. The key benefit is that underlying physics is linear. Complete sets of inputs and outputs are described in Table I. Note that the model inputs are different from the inputs of the detailed EP model – direct manipulation of some heat fluxes is not allowed, and therefore we have to provide signals on a lower level (see Table I). The input set was divided into two categories: the first group represents the actuator heat fluxes, whilst the second represents disturbances affecting the system. The identification procedure does not distinguish between disturbances and manipulated variables, however, is needed for user orientation and consequent control as well.

B. Step-by-Step to get a model

Each of the following steps is actually a stand-alone software package which enables a specific task as follows.

1) GenEI: The main task of GenEI is a generation of sufficiently exciting input signals. Such signals are needed to satisfy key theoretical assumptions on reliable statistical identification – persistent exciting signals [17]. In real operation, this request is almost infeasible due to technical, physical or economical limitations. As the image of the building modeled in EP is at hand, a proper identification experiment can be designed. Obviously, when the objective is to build-up of a model suitable for control, the generated inputs do not need to cover the entire frequency domain, but rather some control-relevant selection. The prior knowledge of the time constants of the system is often known or at least possible to estimate using some preliminary tests, thus the input signal is generated according to this information. We have proposed three different kinds of input signals, pseudo-random binary signal (PRBS), sum of sinusoids (SINE) and multilevel pseudo-random signal (MPRS). All of them have similar settings as follows. Let τ_H, τ_L denote the slowest and the fastest systems time constants, respectively. Then the required frequency spectrum to be covered by the generated signal is (ω_*, ω^*) and the following equation holds:

$$\omega_* = \frac{1}{\beta \tau_H} \leq \omega \leq \frac{\alpha}{\tau_L} \omega^*, \qquad (1)$$

where α defines the ratio of closed-and-open loop responses and β specifies the settling time. Typical values are $\alpha = 2$ and $\beta = 3$, which corresponds to 95 % of settling time [18]. Due to the Nyquist-Shannon-Kotelnikov theorem, frequency range of the generated signal cannot be as in (1), but must be larger, and the range (1) should bear majority of the power of the signal. Furthermore, the choice of switching time is based on $T_s \leq \frac{2.78}{\omega^*}$.

In case of MPRS, the input sequence is computed by Galoise fields [18] with the number of shift registers n and the length q, which defines the maximum possible multiple of harmonics to be suppressed. In the opposite way, let h be the maximum possible multiple of the harmonics to be suppressed. Then q has to be chosen such that $q \geq 2^h - 1$. Next, the length n can be computed according to $\omega_* \geq \frac{2\pi}{T_s(q^n-1)}$. Further on, the length of a signal cycle is $N_{cyc} = q^n - 1$, which, in time domain, represents a signal of duration $T_{cyc} = N_{cyc} \cdot T_s$. The number m of the signals to be generated has to be considered as well, but in practical applications, it is sufficient to generate only 1 signal, and shift it $(m-1)$ times afterwards. It is indeed a very suitable solution as the signal generation is time consuming. Moreover, this technique guarantees the sufficient lack of cross-correlation between the respective signals [19].

2) GenSIO.: This block processes outputs produced by GenEI (inputs to EP), outputs of EP and some variables from schedules and databases, and produces the input and output data sets used in SID. The respective inputs, outputs and disturbances, as used in identification, are described in Table I and the procedure of data generation and preparation for identification is schematically depicted in Figure 3 and Figure 4, respectively.

TABLE I

NOTATION OF THE VARIABLES USED FOR SYSTEM IDENTIFICATION

ID	Variable Category	Type	Zone relevant	EP equivalent
Q_{CONV}	Convector heating rate	Input	Yes	Same quantity, power can be arbitrarily set within limits
ZCPCR	Zone ceiling panel cooling rate	Input	Yes	Supply water temperature and mass flow rate through plumbing can be adjusted. Together with return water temperature, they stand for heat flux of radiant ceiling
ZCPHR	Zone ceiling panel heating rate	Input	Yes	Same as ZCPCR
LG	Lighting gains	Input	Yes	Same quantity, power can be arbitrarily set within limits
NRF	Net radiation flux	Disturbance	Yes	Partly by means of blinds control
FP	Fan power	Input	Yes	Air flow rate (which is either 55 or 0 m^3/h) and supply air temperature. Together with return air temperature, they stand for heat flux of fans.
ODBT	Outdoor dry bulb temperature	Disturbance	No	Same quantity
EG	Equipment gains	Disturbance	Yes	Same quantity
OG	Occupancy gains	Disturbance	Yes	Same quantity
ZT	Zone temperature	Output	Yes	Same quantity
ZI	Zone interior illuminance	Output	Yes	Same quantity

3) Splitter.: Even the powerful servers (64bit machines, 16 cores @ 2.6 GHz and 24 GB RAM) are not able to compute the identification procedure because of the size of the problem. However, due to the floor layout, façade orientation and floor usage, the model of the third floor can be looked at as four decoupled subsystems. The ring-shaped layout of the floor houses also four cores (hosting infrastructural supply such as elevators, staircases, etc.) which separate the office spaces from each other. Thermal coupling of the investigated office zones via the cores is very loose and can be neglected, as far as control issues are concerned. The distributed heating, cooling and ventilation supply of the zones also support the idea of the system division said four subsystems. Consequently, each of the subsystems contains its zone-relevant signals and a copy of the signal, which was originally common for all subsystems. In other words, when a satisfactory computational power is at hand, the proposed procedure does not request special knowledge of the building's physics, etc. On the other hand, it does not exclude the possibility of the system division due to computational or other reasons.

4) SID.: The choice of the identification method was determined by the factors described in previous sections, namely the size of the problem with a vast number of inputs and outputs, implying the multiple input multiple output (MIMO) system, and on the other hand, a huge set of generated (and/or measured) data suggesting the use of statistical identification procedures. Two different choices for subspace identification algorithm were implemented

- `N4SID` function from System Identification toolbox for Matlab, (see [20], [13]).
- Combined deterministic-stochastic algorithm [17], [21].

5) Joiner.: The four resulting subsystems are merged together, when all the subsystems retain their zone-specific signals, whilst the common signals are joined.

6) Verification and validation.: Each of the identified system was verified using the verification data sets and residual and correlation analyzes. The joined system was repeatedly verified.

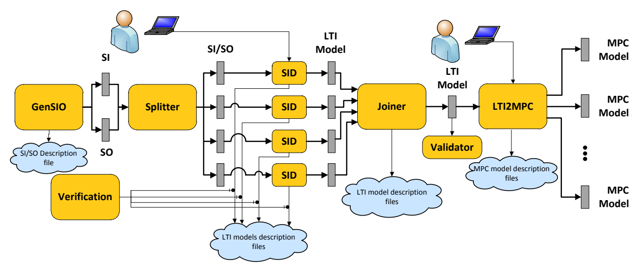


Fig. 4. System identification procedure

7) LTI2MPC.: For optimization requirements, there are several variables added to the model, e.g. total Energy Power Demand (totEPD[2]) or total Heat Power Demand (totHPD[3]). For this reason, the model provided by Joiner must be transformed according to control requirements. This transformation is actually determined by the MPC variant, e.g. optimization objective. Furthermore, for purposes of predictive optimization (cost function and particular bounds), the B and D matrices must be split for the deterministic (manipulated variables) and stochastic (disturbance variables) counterparts, as the SID identified all the inputs (no matter the deterministic and stochastic parts) together.

IV. IDENTIFICATION RESULTS

Because of the size of the system matrices, we will omit them here and show only the resulting model validations, which were carried out using comparison of k-step ahead predictions, as well as by the analysis of the system structure. The first identification attempts, which seem to be straightforward, were to excite the system by generated SINE signals with $\tau_L = 60$ and $\tau_H = 240$ minutes, which is sufficient in sense of building dynamics. Since EP needs different input signals than our identified model (Table I), not all model inputs are able to excite the system in an arbitrary way (see Figure 5

[2]totEPD is a sum of lighting and equipment gains and ceiling cooling

[3]totHPD is a sum of convectors' heating rate and a ceiling heating

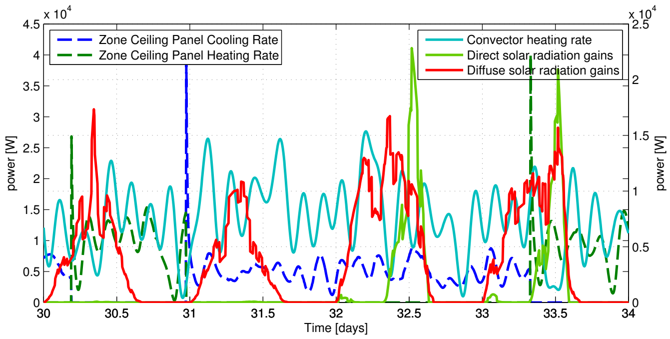


Fig. 5. Part of model inputs.

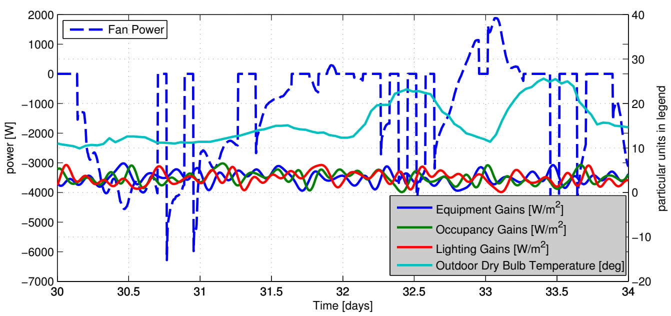


Fig. 6. Part of model disturbances.

and Figure 6). Anyway, these data are still excited enough to describe the system behavior well. This statement can be deduced from the response of the identified model to the verification data set. The part of output data corresponding to the time axes of Figure 5 and Figure 6 is depicted in Figure 7. The model, or to be more specific "1 of the 4 submodels", has an order around 20. This value depends on the type of input excitation, identification data length, time period of the year for which the identification is computed, focus on either simulation or prediction, and on the users choice (since the data is disturbed by noises). Even after joining partial submodels into one big model, the verification response stays great not only for 1-step ahead prediction (Kalman filtering), but for longer predictions as well – see comparison for all zones in Figure 8.

However the system response to verification data is nice, a significant (and surprising) drawback of the statistical identification has come out. The step responses did not satisfy our expectations in both DC gains and signs. To ensure not only good verification response, but step response as well, it was necessary to reshape the experiment inputs. So far, all inputs have been excited in parallel, and therefore partial zones in one subsystem could affect each other. Thus for the new identification, the experiment inputs for one subsystem have been torn in time in such a way, that at any time instant, only one input category in only one zone is excited; all other inputs are set to a constant "stand-by" value. "Stand-by" values have been selected to ensure no active heating nor cooling into the building, but only natural behavior. Improvement of the resulting system structure was significant, and the resulting model is now valid also from physical point of view, which can be illustrated by step responses from subset of system inputs to zone temperatures in Figure 9. All the zones have correct step response dynamics as well as sign. The step responses from a specific energy source or outdoor temperature do not have the same impact for all zones (each zone has different size, orientation and equipment), but should be similar:

- Ceiling heating rates (see Figure 9(a)) present correct structure with an appropriate impact of energy sources – the larger the zone is, the smaller the temperature impact of 1 W of input signal.
- Ceiling cooling (see Figure 9(b)) has, in all cases, correct sign of the step response (positive power demand should affect the zone negatively).
- From Figure 9(c), we can see quite a high impact of the outdoor temperature on the zone temperatures. It has of course, slower dynamics than ceiling panels shown in Figure 9(b) and Figure 9(a), respectively.

V. Conclusions and and future development

This paper has introduced a new methodology of interconnecting building simulation software and traditional identification methods in order to avoid the statistical problems with data gathered from the real building. The building was modeled using EnergyPlus, which was excited by specially proposed signals to get data of a good quality. Then the subspace identification approach (with some modifications) was applied to acquire a model suitable for predictive control. To the authors' best knowledge, there was no detailed building modeling intended for predictive control of such a size. The last step of preparation of the model for control are adjustments of inputs and outputs, in order to obtain a model corresponding to the variety of MPC problems (according to the control criteria).

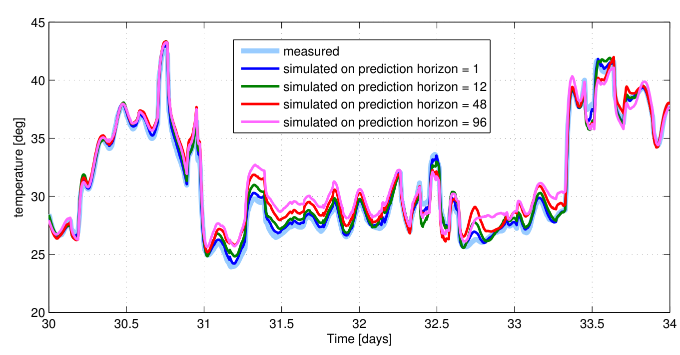


Fig. 7. Part of model outputs.

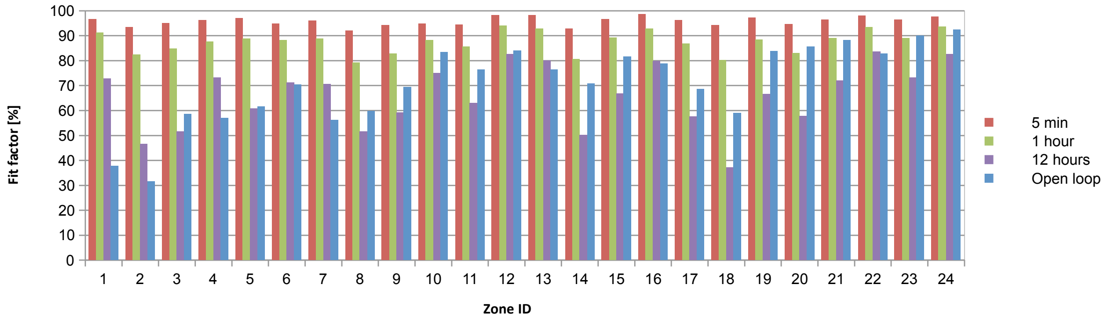
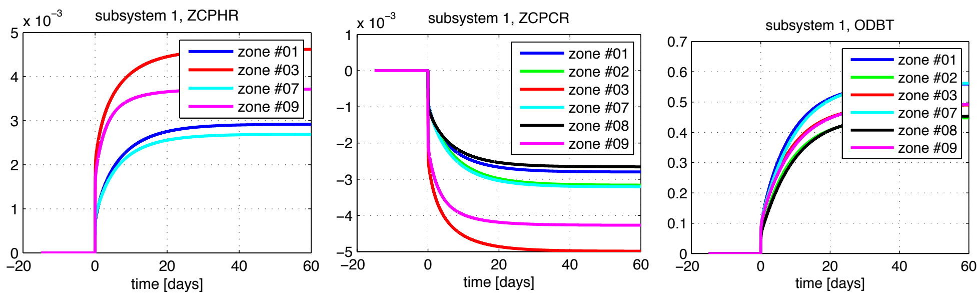


Fig. 8. Fit-factors for all zones for different k-step ahead predictions



(a) Zone ceiling panel heating rate (b) Zone ceiling panel cooling rate (c) Outdoor dry bulb temperature

Fig. 9. Step responses from a subset of inputs to particular zone temperature (inputs and outputs are paired by zone name)

Acknowledgments

The work was supported from the state budget of the Czech Republic, through the Ministry of industry and commerce, in the scope of grant 5.1 SPK01/024 ENERGOKLASTR.

References

[1] L. Perez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[2] J. Wong, H. Li, and S. Wang, "Intelligent building research: a review," *Automation in Construction*, vol. 14, no. 1, pp. 143–159, jan 2005.

[3] H. Han, Y. Jeon, S. Lim, W. Kim, and K. Chen, "New developments in illumination, heating and cooling technologies for energy-efficient buildings," *Energy*, 2009.

[4] H. Hjalmarsson, "System identification of complex and structured systems," *European journal of control*, vol. 15.

[5] R. Gopaluni, R. Patwardhan, and S. Shah, "MPC relevant identification–tuning the noise model," *Journal of Process Control*, vol. 14, no. 6, pp. 699–714, 2004.

[6] D. Lauri, J. Salcedo, S. Garcia-Nieto, and M. Martinez, "Model predictive control relevant identification: multiple input multiple output against multiple input single output," *Control Theory & Applications, IET*, vol. 4, no. 9, pp. 1756–1766, 2010.

[7] D. Shook, C. Mohtadi, and S. Shah, "A control-relevant identification strategy for GPC," *IEEE Transactions on Automatic Control*, vol. 37, no. 7, pp. 975–980, 2002.

[8] H. Madsen and J. Holst, "Estimation of continuous-time models for the heat dynamics of a building," *Energy and Buildings*, vol. 22, no. 1, pp. 67–79, 1995.

[9] K. Andersen, H. Madsen, and L. Hansen, "Modelling the heat dynamics of a building using stochastic differential equations," *Energy and Buildings*, vol. 31, no. 1, pp. 13–24, 2000.

[10] J. Cigler and S. Prívara, "Subspace identification and model predictive control for buildings," in *The 11th International Conference on Control, Automation, Robotics and Vision – ICARCV2010*, pp. 750–755.

[11] M. Gwerder, B. Lehmann, J. Tödtli, V. Dorer, and F. Renggli, "Control of thermally-activated building systems (tabs)," *Applied energy*, vol. 85, no. 7, pp. 565–581, 2008.

[12] T. Bohlin and S. Graebe, "Issues in nonlinear stochastic grey box identification," *International Journal of Adaptive Control and Signal Processing*, vol. 9, no. 6, pp. 465–490, 2007.

[13] L. Ljung, *System Identification: Theory for user*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 1999.

[14] S. Prívara, J. Cigler, Z. Váňa, L. Ferkl, and M. Šebek, "Subspace identification of poorly excited industrial systems," in *Proceedings of the 49th IEEE Conference on Decision and Control*, 2010, pp. 4405–4410.

[15] M. Trčka, J. Hensen, and M. Wetter, "Co-simulation of innovative integrated HVAC systems in buildings," *Journal of Building Performance Simulation*, vol. 2, no. 3, pp. 209–230, Sept. 2009.

[16] W. M. and H. P., "A modular building controls virtual test bed for the integration of heterogenous systems," in *Third National Conference of IBPSA-USA Berkeley/California*, 2008.

[17] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 101 Philip Drive, Assinippi Pard, Nowell, Massachusetts: Kluwer Academic Publishers, 1999.

[18] M. Braun, D. Rivera, A. Stenman, W. Foslien, and C. Hrenya, "Multi-level pseudo-random signal design and model-on-demand estimation applied to nonlinear identification of a RTP wafer reactor," in *American Control Conference, 1999. Proceedings of the 1999*, vol. 3. IEEE, 2002, pp. 1573–1577.

[19] S. V. Gaikwad and D. E. Rivera, "Control-relevant input signal design for multivariable system identification: Application to high-purity distillation," 1996.

[20] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[21] ——, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 31, no. 12, pp. 1853–1864, 1995.