Preprints of the 2012 20th Mediterranean Conference on Control & Automation (MED), Barcelona, Spain, July 3-6, 2012



# Predictive Control Oriented Subspace Identification Based on Building Energy Simulation Tools

Samuel Prívara, Zdeněk Váňa, Jiří Cigler, Lukáš Ferkl

*Abstract*— **Even though modern control has emerged in numerous control applications, a building automation is still a field where the position of the classical control is almost exclusive. The main reason is that for the synthesis of a predictive controller a decent model for control is needed. In the field of building climate control, it is still problem to obtain a model of large building in an explicit form suitable for control. Most of the approaches either use building modeling software to get detailed model, which is unfortunately in implicit form; or the model is built-up as a first principle model, which usually ends-up as an extreme simplification of the reality. In this paper, a building model identification procedure is presented, wherein the building model is built-up as a first-principle model using a simulation software (detailed, precise, however in implicit form), and then a state-space model is identified by means of subspace identification methods. The main focus of the paper lays on a case study of a large office building, and the entire process of its identification.**

## I. Introduction

The European Union (EU) presented targets concerning energy cuts defining goals until 2020 [1]: *i)* Reduction in EU greenhouse gas emissions at least 20 % below the 1990 levels *ii)* 20 % of EU energy consumption to come from renewable resources *iii)* 20 % reduction in primary energy use compared to projected levels, to be achieved by improving energy efficiency. Even though there are fierce debates about real impact of the renewable resources, because of e.g. the power grid issues, energy savings achieved by smart control algorithms are free of political controversy.

As the buildings account for about 40% of total final energy consumption[1] (and its amount has been increasing at a rate 0.5–5 *% per annum*)[2], an 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 heating, ventilation and air conditioning (HVAC) systems, as well as the blind positioning and lighting systems [3], [4].

One of the control strategies suitable for building automation is *Model Predictive Control* (MPC) [5], [6]; unfortunately, the modeling and identification of buildings is rather difficult and time-consuming [7] (large and complex systems with strongly coupled dynamics among the zones). MPC requires a model, which predicts outputs with reasonable precision on the control-relevant frequency range (see e.g. [8], [9]). One approach is to use the first-principles, however, most of the papers devoted to the modeling from the first principles provide only two-room or two-zone example even when the real building has tens of rooms. We would like to have a model of a large office building with a large number of rooms (i.e. hundreds of inputs and outputs). In case of using building simulation software such as TRNSYS, EnergyPlus (EP), ESP-r, etc. (see [10], [11]) for building modeling, the resulting models are in implicit form and cannot be used for predictive control. Therefore we have decided to use statistically-based, i.e. data-driven models [12], [13].

In this work, we combined the benefits of both above mentioned approaches. A physical model in a building simulation software is created, such that it describes the real building as closely as possible. Then identification signals are fed into the simulation software to obtain the high-quality identification data which are consequently used for obtaining a suitable control-oriented model. The uniqueness of this approach is in the combination of the real building data, the first-principle model in implicit form and statistical identification algorithms and interconnection of the different software tools with possibility of real-world operation.

The paper is structured as follows: next section provides the motivation and explains the background of the problem. Identification, modeling and all the relative issues are addressed in Section III. Section IV discusses the prediction properties of the resulting model and the last section concludes the paper.

## II. Problem Description and Setup

### A. Description of the building

The analysis deals with the third floor of a large office building in Munich (20 000 m$^2$ and six above-ground floors, see Fig. 1). Based on the usage, façade orientation and HVAC supply, the floor can be divided into 24 mutually interconnected zones. 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.

The building automation system contains several actuators, namely individually controlled convectors, 24 independently controlled radiant ceiling panels for cooling and heating, two air handling units (AHU) for control of the ventilation, and venetian blinds for all windows in all zones. Energy supply i.e. hot and chilled water supply for the entire building is provided by a central heating and cooling plant which is

S. Prívara, Z. Váňa, J. Cigler and L. Ferkl are with Department of Control Engineering, Faculty of Electrical Engineering of Czech Technical University in Prague, Technická 2, 166 27 Praha 6, Czech Republic

[1]The energy of the product delivered to the consumer.

978-1-4673-2529-5/12/$31.00 ©2012 IEEE

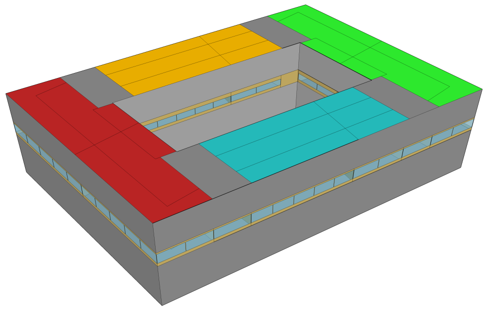


Fig. 1. 3D simulation model: Investigated zones were on the third floor, other floors are grayed out and used for shading purposes only. The zone layout is shown on top of the model for clarity. Zones of the same subsystem are colored alike. Core areas are grey.

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 modeling strategy and model inputs and outputs

The choice of model inputs and outputs should be in accordance with the choice of the identification method which is determined by the application[2]. As was briefly mentioned before, the usage of the first-principle models for our problem is limited [14] and the software packages do not provide model in a form suitable for control. Therefore, we have decided for subspace identification (4SID) which is able to identify large MIMO systems. Some key assumptions of 4SID [15], e.g. persistent excitation, open-loop data, etc. are always violated during building's normal operation. The identification procedure can be improved by including prior information [16] or by carrying out the identification experiment on a building. Depending on the building size, the experiment is expensive, but may bring significant improvements to the resulting model [12]. Therefore a new approach is introduced, which yields a model of a large multi-zone building. A very promising strategy seems to be a combination of a building simulation software used for identification experiments to get data for a standard statistical identification procedure. EP as the building simulation software and Building Controls Virtual Test Bed (BCVTB) as the middleware between EP and a controller written in Matlab is used. The description of the interconnection of the various simulation and computational tools is described more in detail in [17].

## III. Identification and Modeling, Issues and Proposed Solutions

### A. Problem statement

In the last two decades, 4SID algorithms have become an important tool for the system identification (SID). The objective of the 4SID, as will be used further on, is to find a linear time invariant (LTI) discrete time state-space model in an innovative form

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \\ y(k) &= Cx(k) + Du(k) + e(k), \end{aligned} \tag{1}$$

given the measurements of the inputs $u(k) \in \mathbb{R}^m$ and the outputs $y(k) \in \mathbb{R}^l$ with $e$ being zero-mean white noise. In other words, we want to determine the system order $n$ and to find the matrices $A$, $B$, $C$, $D$ and $K$. The order of the system is usually determined from an analysis of the singular values obtained from Singular Values Decomposition (SVD) of oblique projection matrix, see e.g. [15]. Especially in case of strong noise contamination, the determination of the system order is troublesome and the standard procedure is not working well. We have proposed the following heuristic formula for automatic selection of the order $n$ as follows

$$\begin{aligned} f(\sigma_j) &= \operatorname{grad}\log\left[\sigma_1, \quad \sigma_2, \quad \ldots, \quad \sigma_{i \cdot l}\right] \\ n &= \arg\min_j f(\sigma_j), \end{aligned} \tag{2}$$

which ensures better order selection even when the singular values are drown in the noise.

### B. Input signals

Generation of sufficiently exciting input signals is one of the key theoretical assumptions enabling reliable statistical identification. Under real operation, this request is almost infeasible due to technical, physical or economic constraints and limitations. As the image of the building modeled in EP is at hand, the identification experiment was proposed as follows.

Three different kinds of input signals have been constructed, namely pseudo-random binary signal (PRBS), sum of sinusoids (SINE) and multilevel pseudo-random signal (MPRS). Let $\tau_H$, $\tau_L$ denote the slowest and the fastest time constants of the system, respectively. Then the frequency spectrum to be covered is $(\omega_*, \omega^*)$ with $\omega_* = \frac{1}{\beta\tau_H} \leq \omega \leq \frac{\alpha}{\tau_L}\omega^*$, where $\alpha$ defines how fast will the closed-loop be with respect to the open-loop response, and $\beta$ specifies low frequency information corresponding to the settling time. The typical values are $\alpha = 2$ and $\beta = 3$ (95 % settling time) [18]. 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 $q \geq 2^h - 1$ holds and the minimum length $n$ is computed from

$$\omega_* \geq \frac{2\pi}{T_s(q^n - 1)}. \tag{3}$$

The length of a signal cycle is then $N_{cyc} = q^n - 1$, which (in time domain) represents a signal of duration $T_{cyc} = N_{cyc} \cdot T_s$, with $T_s$ being sampling time. The number of the signals to be generated $(m)$ does not need to be considered, as it is sufficient to generate a single signal and shift it $(m-1)$ times, which guarantees good statistical properties of the generated

[2]The whole set of inputs and outputs is a set resulting from the original non-linear system, however, for a linear model, only a subset is selected, which has desired characteristics.

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 arbitrary set within limits |
| ZCPCR | Zone ceiling panel cooling rate | Input | Yes | Supply water temperature and mass flow rate through pipes 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 arbitrary set within limits |
| DSRV | Direct solar radiation gains | Input | Yes | By means of blind control (position and angle), we can adjust solar gains influencing zone temperature. |
| DFSRV | Diffuse solar radiation gains | Input | Yes | Same as DSRV |
| 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 |

signals [19]. For the building under investigation, the fastest and the slowest time constants were acquired as 4 hours and 20 days, respectively.

### C. Specification of model inputs and outputs

The heat fluxes affecting zone temperatures were selected as system inputs and temperatures and illuminances as outputs, in total 288 inputs and 48 outputs lumped into the variable categories as described in Table I for easy comprehension. It is important to note here, that some signals are zone relevant (there exist a single signal for each zone), while others are not (which means, that a signal is common for multiple zones or the whole building as such). Moreover, the signals entering the LTI model and EP model are not the same (e.g. one signal entering the LTI model is composed of multiple EP signals, etc.), which is indicated in Table I by “EP equivalent” column. The key factor for such a selection was the linearity of the underlying physics. The complete set of inputs and outputs is described in Table I. Note that the model inputs are different from the inputs on side of the detailed EP model – direct manipulation of some heat fluxes is not allowed, and therefore, signals on lower level (description of these are given also in Table I) must be provided. The input set was divided into two categories: the former group represents the actuator heat fluxes and the latter disturbances affecting the system. The identification procedure does not distinguish between disturbances and manipulated variables, however, this categorizing is necessary for user orientation as well as consequent control (where the manipulated inputs and measured disturbances must be strictly divided).

### D. Analysis of EP model linearity

Even though the underlying physics of the process is linear, the data produced by EP do not need to be linear as well. The linearity should be hence verified in the first place. The verification of the linearity of the EP model according to the definition of linearity, i.e.

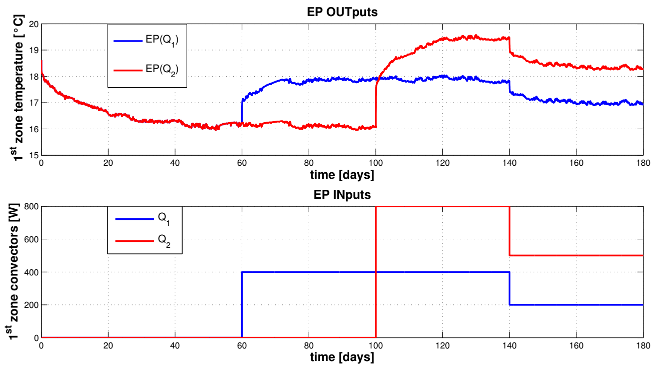
$$\alpha f(x_1) + \beta f(x_2) = f(\alpha x_1 + \beta x_2), \qquad (4)$$

had to be performed. That means, that independent inputs (for convectors – see Fig. 2(a) and Fig. 2(b) lower figures, for equipment and lightning gains see Fig. 3(a) lower figure) were designed, fed into the EP and the response of the EP was summed up and compared to the response of the EP fed-up by sum of the input signals. Linearity tests were intentionally performed at two different outside temperatures, namely $15°C$ and $20°C$. The results can be seen in Fig. 2(c), Fig. 2(d) and Fig. 3(b) for convectors, equipment and lightning gains, respectively. The maximum errors between the sum of EP model responses and the response of the sum of input signals are displayed in Table II. The growing error in case of multiple step in input signals can be explained as follows. It is caused by a cooling effect from the ambient temperature. This effect depends on the temperature difference between zone and ambient temperatures. While sum of input signals is applied (as an input), the actual zone temperature is higher than the level of zone temperature at the beginning of the experiment, thus the cooling effect is higher as well (ambient temperature is constant for the experiment) and above-mentioned error appears. Consequently, this can be written as $Q_{ss} - Q_{cool}(T_z) = Q_{EP}$,where $Q_{ss}$ denotes heat flux corresponding to the designed input (e.g. convectors), $Q_{cool}(T_z)$ is a flux altering (the actual size depends on the temperature difference between outside and zone temperatures) the requested value and $Q_{EP}$ is real value of the flux affecting the simulation software. When summing up two signals with different step size, there is different alternation by $Q_{cool}(T_z)$, hence a small difference between the sum of responses and response of the sums. Nevertheless, it can be concluded, that EP model response on selected inputs is indeed linear, hence the use of 4SID algorithm for creation of linear model is justifiable.

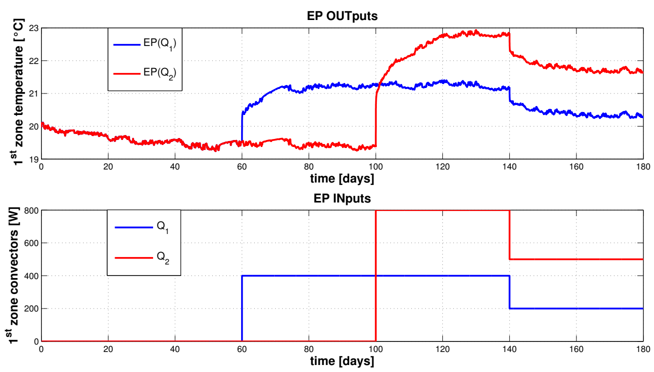
TABLE II

PERCENTAGE TEMPERATURE ERROR IN LINEARITY

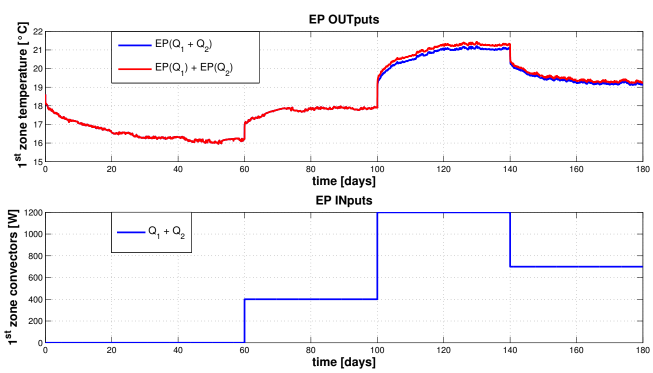
| Errors in % | EG | LG | $Q_{CONV}$ | |
|---|---|---|---|---|
| | | | $15°C$ outside | $20°C$ outside |
| $2^{nd}$ step | 3.9 | 2.1 | 4.4 | 5.2 |
| $3^{rd}$ step | 3.3 | 1.0 | 3.4 | 5.0 |



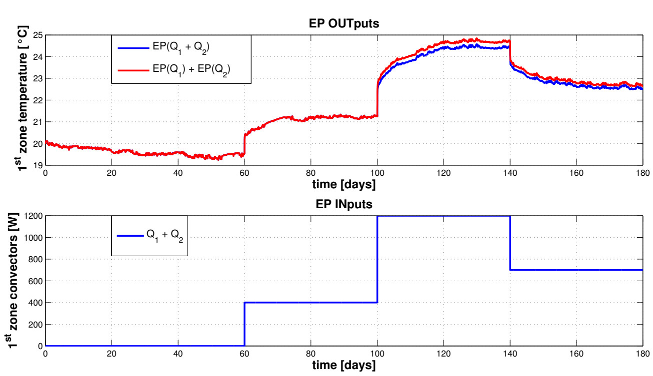

(a) Two convectors: input signals at outside temperature $15°C$ and the EP model response




(b) Two convectors: input signals at outside temperature $20°C$ and the EP model response

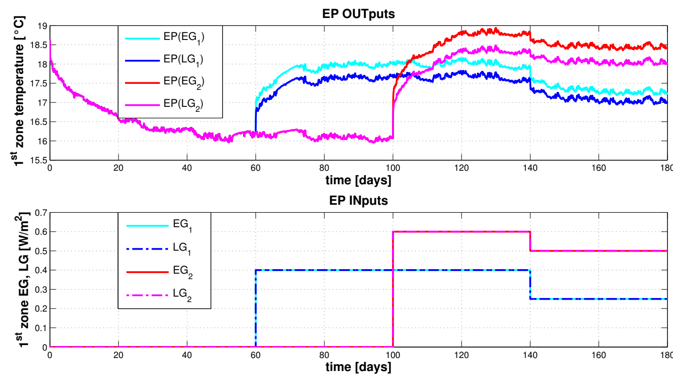



(c) Sum of two convectors: input signal at outside temperature $15°C$ and EP model response (response of sum and sum of responses)

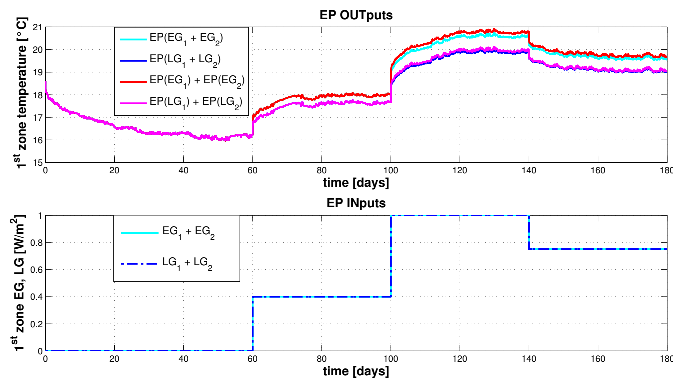



(d) Sum of two convectors: input signal at outside temperature $20°C$ and EP model response (response of sum and sum of responses)

Fig. 2. Convectors: EP model linearity test.




(a) Two convectors: input signals at outside temperature $15°C$ and the EP model response




(b) Sum of signals: input signals at outside temperature $20°C$

Fig. 3. Equipment and lightning gains: test of the EP model linearity

## E. Identification procedure settings

The next step in identification is to decide, how to choose a desired model order and size of the Hankel matrices that enter the algorithm.

1) **Identification algorithm:** There are several algorithms covered by subspace methods, which differ in applicability, numerical stability and computational demands. For this study, the N4SID was employed.
2) **Desired model order:** Although the order selection has already been implemented in N4SID, the problem of complexity of the large building remains. Even though the model was linear, and there are no unknown disturbing signals, the order obtained by original algorithm was very high. It is often enough to have smaller model, which still sufficiently describes the process. The physically-based assumption is that the change of temperature caused by a source of heat is a $1^{st}$ order process. Adding the walls of the leads to a $2^{nd}$ order process. Additional effects (see e.g. [20]) finally leads to a $3^{rd}$ order dynamics per output temperature. Extended to the whole building level (4 subsystems per floor each consisting of 6 zones), the model order should be between 48 and 72 (prior estimate of the system order). After employing N4SID algorithm and validation tests, $18^{th}$ order model (order selection algorithm Eq. (2)) per subsystem turned out to be indeed a good choice,

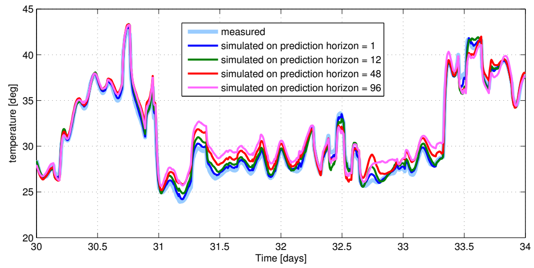


Fig. 7. Predicted outputs for different horizons.

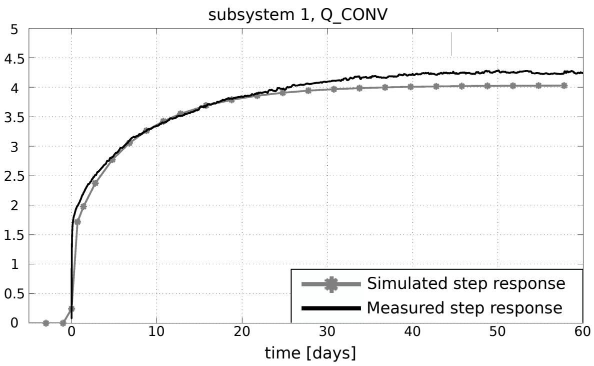



Fig. 5. Measured and simulates step responses from convectors.

considering both its simplicity and sufficient precision.

3) **Size of the Hankel matrices** is given by the number $i$ of block rows in said matrices. $i$ must be larger than the maximum model order [15]. Essentially, $i$ means how far into the past/future of the measured data is searched, therefore it may appear that the greater $i$ is, the better the result will be. However, a trade-off has to be done concerning the computation difficulties, especially for the case of a large MIMO systems (such as a building). Several numbers $i$ had been selected and their effect on the identification results were analyzed. Fig. 4 shows the step responses of several inputs for $i = 24, 30, 36$ and $40$. All these step responses possess good properties, such as reliable dynamics, the sign of the effect as well as its nominal value. Using greater $i$ (thus more information from data), only DC-gains change. Next, the measured step responses (the EP had been excited by step impulses into individual inputs and the step responses were measured) were analyzed. It turned out, that with greater $i$ the model step responses approach the measured step responses (see Fig. 5). Finally, $i = 40$ was selected for the size of Hankel matrices for one subsystem.

## IV. Prediction properties of the model

The good prediction properties of the model are crucial for MPC. As there is a single MPC for the whole building, the four submodels are joined together. Recall, that $1$ of the $4$ submodels, has an order $18$ (72 for joined model). 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. Even after joining partial submodels, the verification response stays very good not only for 1-step ahead prediction (Kalman filtering), but for longer predictions as well – see the comparison for all zones in Fig. 6. The graphical representation of the prediction properties of the model for different prediction horizon are depicted in Fig. 7. However, the precision of the prediction is not the only criterion. Additional request on the model is a good step response. Step responses of the identified model and the original systems were tested and it was found out that they complied each other, for step response of convectors see Fig. 5.

## V. Consequences for Control and Conclusions

This paper has introduced a new methodology of interconnecting building simulation software and traditional identification methods. The effort was to avoid the low-excited data, to have a complex MIMO model reliably reflecting reality and to be able to switch easily between control of simulation example (EP) and real building using BacNet. The building was modeled using EnergyPlus, which was excited by proposed input signals to get data of a good quality. Then subspace identification approach was applied to acquire a model suitable for predictive control. The last step of preparation of the model for control are selections and adjustments to inputs and outputs for obtaining the model corresponding to the variety of MPC problems. We investigated a number of properties and parameters of the identification algorithm and provided some hints for identification of the large MIMO systems such as buildings. Consequently, the resulting model is in a state-space form and can be readily used within the predictive control framework.

## References


[1] "Communication from the commission to the European Parliament, the Council, the European Economic and Social Comittee and the Comittee of the regions," Brussels, p. 14, 2010.

[2] 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.

[3] J. Široký, F. Oldewurtel, J. Cigler, and S. Prívara, "Experimental analysis of model predictive control for an energy efficient building heating system," *Applied Energy*, 2011.

[4] 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.

[5] F. Oldewurtel, D. Gyalistras, M. Gwerder, C. Jones, A. Parisio, V. Stauch, B. Lehmann, and M. Morari, "Increasing energy efficiency in building climate control using weather forecasts and model predictive control," in *Clima 10th REHVA World Congress*, 2010, pp. 9–12.

[6] F. Oldewurtel, A. Parisio, C. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, "Energy efficient building climate control using stochastic model predictive control and weather predictions," in *American Control Conference*. IEEE, 2010, pp. 5100–5105.

[7] Y. Zhu, *Multivariable system identification for process control*. Elsevier Science, 2001.

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

[9] 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.

[10] 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.

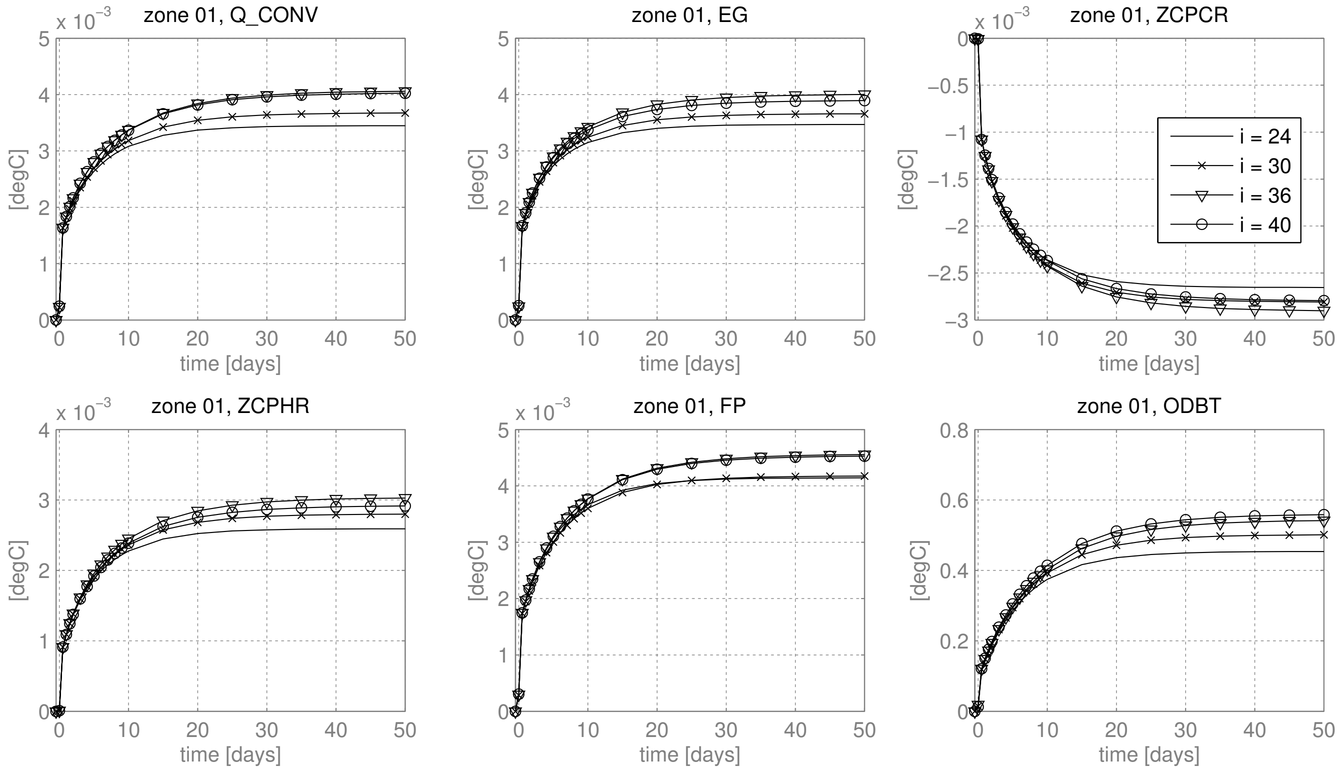



Fig. 4. Step responses of several inputs in zone 1 for different $i$-s. Vertical axes are particular contributions to zone temperatures.

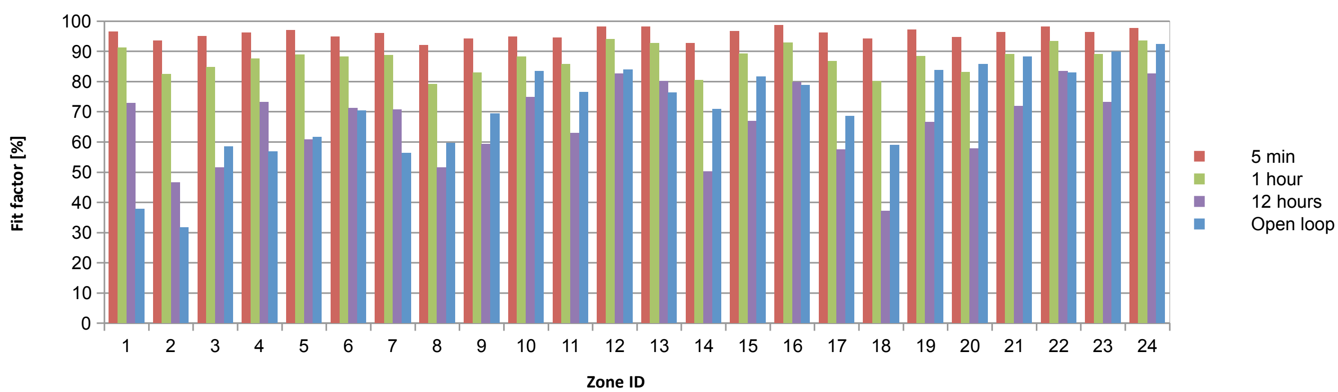



Fig. 6. Fit-factors for all zones for different $k$-step ahead predictions


[11] 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, Sep. 2009.

[12] 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.

[13] S. Prívara, J. Sirokı, L. Ferkl, and J. Cigler, "Model predictive control of a building heating system: The first experience," *Energy and Buildings*, 2010.

[14] 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.

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

[16] 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.

[17] S. Prívara, Z. Váňa, D. Gyalistras, J. Cigler, C. Sagerschnig, M. Morari, and L. Ferkl, "Modeling and identification of a large multi-zone office building," in *2011 IEEE Multi-Conference on Systems and Control, Denver, CO, USA*, 2011.

[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] D. Gyalistras and M. Gwerder, "Use of weather and occupancy forecasts for optimal building climate control (opticontrol): Two years progress report." Terrestrial Systems Ecology ETH Zurich, Switzerland and Building Technologies Division, Siemens Switzerland Ltd., Zug, Switzerland, Tech. Rep., 2010.