Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Black-box modeling of buildings thermal behavior using system identification ⋆

Sullivan Royer * **Stéphane Thil** * **Thierry Talbert** *
Monique Polit *

* *Laboratoire PROMES-CNRS (UPR 8521),*
Tecnosud, Rambla de la thermodynamique
66100 Perpignan, France
Tel: +334-6868-2222; e-mail: firstname.lastname@promes.cnrs.fr

Abstract: In this paper the modeling of buildings thermal behavior is studied. The main goal is to develop a modeling procedure that can be used at different scales (a thermal zone, a floor or a whole building) and on different buildings. The scalability of the chosen black-box model structure is first assessed; simulation experiments are then conducted in order to test if the modeling procedure is reusable. As these tests are hardly feasible in practice, a real university building is first modeled using an energy simulation software. This model is then used to validate the proposed approach.

Keywords: System identification, energy efficiency, building thermal modeling.

1. INTRODUCTION

As the building sector is the largest consumer of energy among all economic sectors and a greenhouse gases emitter, its energy consumption has to be reduced. A way of doing so, and a popular research subject, is the application of model predictive control (MPC) strategies to the building automation systems (see [Freire et al., 2008, Kolokotsa et al., 2009, Široký et al., 2011]). The dynamic model used to perform the predictions is of great importance and three categories regroup the numerous modeling approaches that have been considered:

- white box models are based on physical knowledge of the system and thermal balance equations: these are often obtained through energy simulation softwares like EnergyPlus [Crawley et al., 2001], TRNSYS [S.A. Klein *& al.*, 2010], etc;
- black box models use only measured input/output data and statistical estimation methods (e.g. [Cigler and Prívara, 2010, Ferkl and Široký, 2010]);
- grey box models, a mix of the first two categories: they use input/output data as well as some *a priori* knowledge on the system. A popular grey-box model is the equivalent RC networks (see [Wang and Xu, 2006, Široký et al., 2011, Bacher and Madsen, 2011] for example).

This work is a part of the RIDER project [RIDER Project, 2010], whose objective is to improve energy efficiency of buildings and groups of buildings while preserving the thermal comfort of the occupants and/or other (economic) criteria. The use of MPC is in this case natural. However, since the developed methodology should be applicable to different buildings, at different scales (from a single room to a whole building) and in different locations, the main issue becomes the model used to predict the thermal behavior of the studied zone: it needs to be reusable and scalable, while being simple enough to apply MPC. As a consequence, the model's structure should be complex enough to adapt to the various scales and the use of *a priori* knowledge should be avoided. Black box models are thus the most appropriate.

This paper presents, in the second section, how an scalable model of an university building has been obtained. The building was modeled using the energy simulation software EnergyPlus (see [Royer et al., 2013] for more details) and rich data have been generated to obtain a scalable model using system identification. In the third section, the generic feature of modeling procedure is tested by applying it in different locations and with another building.

2. STUDY OF A UNIVERSITY BUILDING

The studied building (see Figure 1) is a five years old university building located in the city of Perpignan (southern France). It is a two-storey building, with a total floor area of $514.9\,\mathrm{m}^2$. The ground floor is divided in five thermal zones: two classrooms, a space for students, an office, a corridor and toilets. The first floor is identical, except that the student space and the office are gathered into a media classroom.

The HVAC system is a Hitachi Variable Refrigerant Flow. The heat or cold production is performed by two outside condensing units connected to several inside air handling units (simultaneous cooling and heating is not allowed).

2.1 Inputs, outputs and informative experiments

The main signals influencing indoor temperature are: outdoor temperature, direct normal solar radiation and

⋆ This work is a part of RIDER project, supported by a fund from the *Fonds Unique Interministériel* (OSÉO – Région Languedoc-Roussillon).

Fig. 1. Photography of the two floors building under study and the corresponding Google Sketchup mock-up.

HVAC power in the considered thermal zone. Although several additional inputs can be considered, only these three signals have been considered to keep the model simple. The output $y(t)$ is the indoor temperature in the considered thermal zone (see Figure 2). Note that, in this paper, it is supposed that there are no internal gains; the reason is that – at least within the RIDER project – it can be considered as a known perturbation through *a priori* knowledge on occupation (time tables) and equipment.

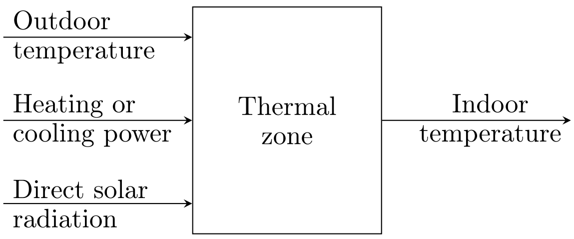

Fig. 2. Three inputs / one output model of a thermal zone.

As it relies on input/output data only, the identification of black-box models usually requires data of high quality. This is mathematically expressed through the notion of persistence of excitation (see e.g. [Ljung, 1999]): the input signal must persistently excite the system under study, at least in the frequency band of interest. As a consequence, input signals such as filtered white noise or pseudo-random binary sequences (PRBS) are commonly used.

Data collected from real operation of a building are usually insufficiently informative to reliably estimate the model [Prívara et al., 2013]. Informative experiments have thus to be designed. Among the various inputs affecting the thermal behavior of a building, the only controllable input is the heating/cooling system power; an input such as a PRBS can be impossible, too long or even too expensive to run. To overcome this problem, a model of the studied building has been constructed using EnergyPlus [Crawley et al., 2001], an energy simulation software. This model will in turn be used to generate the informative data used to perform system identification procedures. Note that an energy simulation software is used in this work only as a first step, to assess the feasibility of the problem and gain insight on the minimal design of experiments. In future work, no other model than the black-box model will be used, since it defeats the purpose of not relying on physical *a priori* knowledge on the studied thermal zone.

2.2 EnergyPlus model

A 3D model of the studied building has been created from the building architect book (containing drawings, dimensions, materials, ...) with the Google Sketchup 3D design software (see Figure 1). All geometry data have been imported in EnergyPlus using the OpenStudio plug-in; thermal parameters such as conductivity, specific heat, etc. come from ASHRAE's book [ASH, 2009]. The HVAC characteristics are provided by the manufacturer documentation book, except energy ratio curves which are derived from similar equipment found in [Raustad, 2012].

The validity of this model has been confirmed using measurements from installed sensors: the standard deviation between measured and simulated temperatures is less than 1 °C over a test period of one month with several types of excitation and weather (see [Royer et al., 2013]).

2.3 System identification

EnergyPlus and Matlab co-simulation The design of input signals in EnergyPlus is made through the use of schedules, which is quite impractical to generate a Pseudo-Random Binary Sequence (PRBS) for example. Co-simulation, that is the integration of different software components by run-time coupling [Sagerschnig et al., 2011], is a solution to overcome this problem. Co-simulation of EnergyPlus with Matlab is done through a software environment called Building Controls Virtual Test Bed (BCVTB) [Wetter, 2011], used as a middleware, and the toolbox MLE+ [Bernal et al., 2012], which provides Matlab functions and classes.

The procedure is as follows: the input signals are designed with Matlab and then fed to EnergyPlus; EnergyPlus then simulates the output signals that are used to estimate the model's parameters using Matlab system identification toolbox [Ljung, 2003].

Model structure A discrete-time state-space representation, more suitable for modeling MIMO systems, has been chosen. In innovation form, it is given by (see *e.g.* [Ljung, 1999]):

$$x(t+1) = Ax(t) + Bu(t) + Ke(t) \tag{1}$$
$$y(t) = Cx(t) + Du(t) + e(t) \tag{2}$$

where $u(t)$ is the input vector, $y(t)$ the output vector, $e(t)$ the disturbance and $x(t)$ the state vector. The order of the state-space model and the coefficients of the state-space matrices A, B, C, D and K are estimated using input/output data and system identification techniques.

Design of experiments Pseudo-random binary sequences is a common input signal in system identification: it is a periodic, deterministic signal with white-noise-like properties [Ljung, 1999]. The only controllable input of the input vector $u(t)$, is the heating/cooling power of the

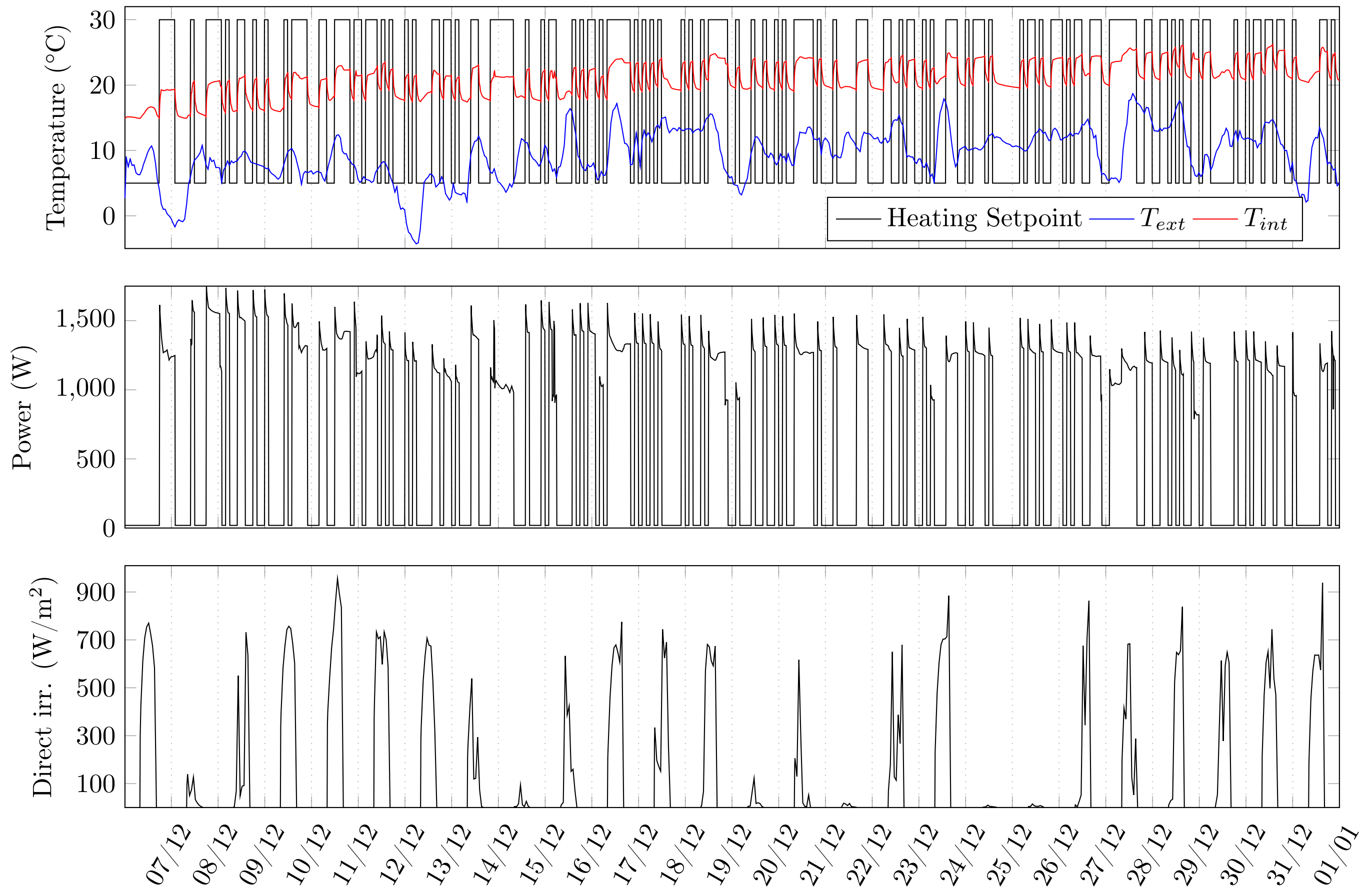

Fig. 3. Part of generated data for classroom 01. Top plot: indoor temperature, outdoor temperature and temperature setpoint. Middle plot: heating power. Bottom plot: direct normal radiation.

HVAC system. However, in EnergyPlus, it is currently impossible to directly control the heating or cooling; the solution is then to control the indoor temperature heating and cooling setpoints, which act directly on the HVAC power. In order to ensure that the HVAC power follows the setpoint PRBS as closely as possible, the binary levels of the PRBS are set to 5 and 30 °C: the HVAC should thus either be on at full power, or completely off. Moreover, the cooling setpoint is set to 40 °C, avoiding the HVAC system to switch in cooling mode. The EnergyPlus model is excited with a time-step of 10 minutes, during 36 days (from the 6th of December 2012 to the 11th of January 2013). Some of the generated data for one thermal zone can be found in Figure 3, where the input signals (direct normal solar radiation, HVAC power and outdoor temperature), the output signal (indoor ambient temperature) and the indoor temperature heating setpoint are plotted. One can see that, as expected, the indoor temperature never reaches the setpoints and the HVAC power follows the PRBS rather closely. Variations are due to internal rules of EnergyPlus regarding HVAC systems control. One can also notice that during the considered period, the weather is varied (sunny and cloudy days, low and high outdoor temperature, etc.), which means more information to construct the model than if there were only warm and sunny days.

Order and parameters estimation Tests on the generated data have shown that a second order model is sufficient estimated to model the dynamics of the indoor temperature (see also the discussion in [Prívara et al., 2013]). Hence, the model order has been set to two for all the models considered in the sequel.

The chosen model structure is a state-space model of second order, and to assess its scalability, several models have been identified:

- first, each room of the building is considered as a thermal zone separately;
- then the case of each floor as a unique thermal zone (whose temperature is defined as the average of each thermal zone temperature weighted by its volume);
- finally, the complete building as a unique thermal zone (same as above).

Each time, the considered thermal zone is excited using the same PRBS to control the heating setpoints, while the other thermal zones are left drifting. The system matrices A, B, C and D are determined using the classical prediction-error method in the system identification toolbox of Matlab [Ljung, 1999].

Model validation The model validation method is the cross-validation, the generated data is divided in two sets: the first two-thirds are used to estimate the model's parameters, while the last third of the data is used to validate the identified model.

Furthermore, the Normalized Root Mean Square Error (NRMSE) fit value appraises the models' quality, it is defined as:

$$\text{fit} = 100\left(1 - \frac{\|y-\hat{y}\|_2}{\|y-\text{mean}(y)\|_2}\right) \tag{3}$$

where y is the validation data and $\hat{y}$ is the output of the model. In addition to pure simulation (infinite prediction horizon), the ability of the identified model to predict indoor temperatures is tested on three prediction horizons: 24 h, 12 h and 3 h.

The comparison between the validation data and the temperature simulated models data is presented in Figure 4. Only four plots for pure simulation are printed here to save space but results are similar for the other zones. The fit values for the different prediction horizons can be found in Table 1.

Figure 4 and Table 1 show that the second order model structure is complex enough to account for the different scales (a room, a floor or the whole building): the indoor temperature predictions are satisfying for our needs (the temperature gap between them and the validation data is lower than 1 °C).

Table 1. Fit between validation data and models output (rounded to the nearest integer).

	Fit in %			
Prediction horizon	∞	24 h	12 h	3 h
Classroom 01	73	84	87	89
Classroom 02	71	79	84	88
Student space	69	83	87	90
Office	70	84	88	91
Classroom 11	70	84	88	91
Classroom 12	74	82	87	90
Media room	65	69	76	88
First floor	70	84	89	92
Whole building	71	84	89	92

3. ADAPTABILITY OF THE SYSTEM IDENTIFICATION PROCEDURE

In this section simulation experiments are conducted to test if the modeling procedure is reusable. First, the procedure is applied on the same university building in several new locations; the same modeling procedure is then applied on a different building.

3.1 The university building in different locations

Three locations have been chosen in addition to Perpignan: Tampa (Florida), San Francisco (California) and Madison (Wisconsin). These cities are geographically distant enough to have different meteorological conditions and the weather files are available within EnergyPlus. The Table 2 presents the fit values between validation data and models output for pure simulation of the university building, in Perpignan and in these three cities. It can be noticed that:

- the fit values are for the most part between 70 and 80 % or higher for pure simulation for all thermal zones except these of Madison, which are lower;

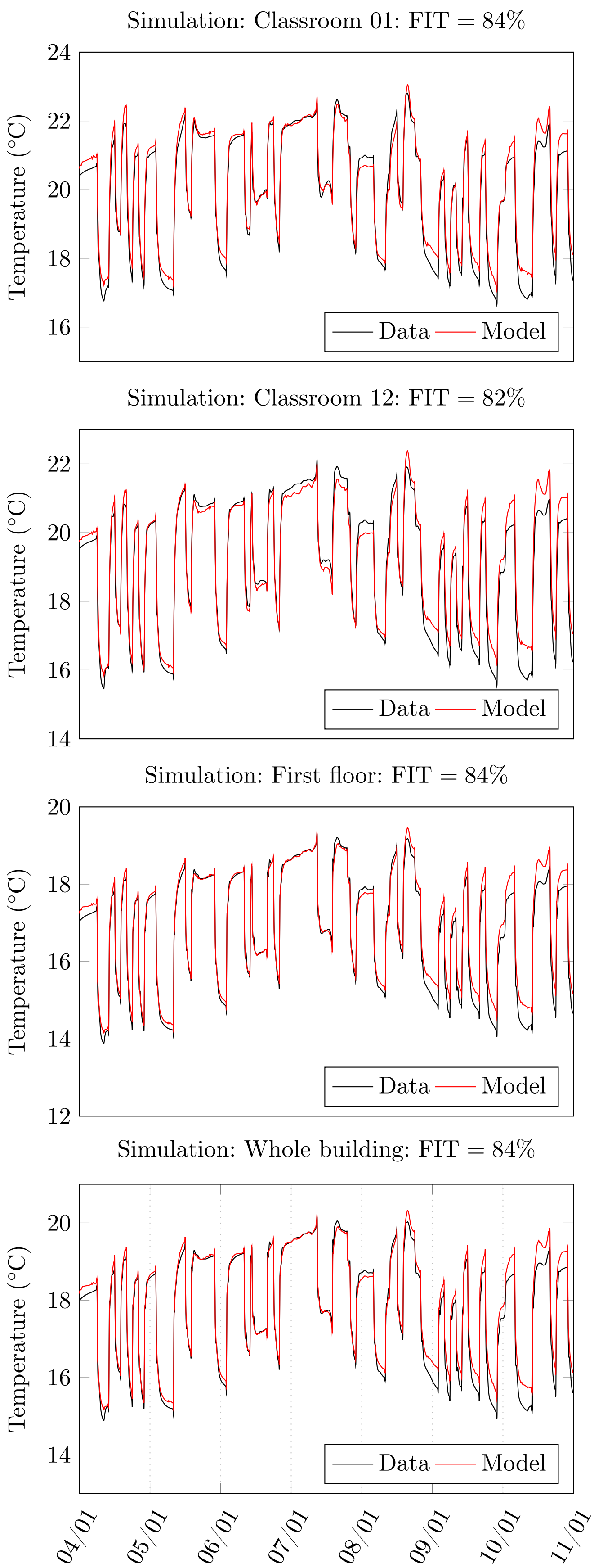

Fig. 4. Comparison between validation data and models output for pure simulation (infinite prediction horizon).

Table 2. Fit between validation data and models output: case of the university building in different locations (pure simulation).

	Fit in %			
Location	Perpignan	Tampa	SF	Madison
Classroom 01	73	73	78	72
Classroom 02	71	74	75	69
Student space	69	78	87	56
Office	70	78	85	70
Classroom 11	70	68	86	67
Classroom 12	74	77	81	66
Media room	65	87	92	51
First floor	70	80	90	60
Whole building	71	79	88	62

- the fit values for the building in Perpignan are not the best while the identification procedure was made from it. Sorting the cities in best results order is: San Francisco, Tampa, Perpignan and Madison.

The differences of values can be explained in part by the location and the weather differences of the four cities. Geographical coordinates are presented in the Table 3 and we can see that all the cities have things in common: Perpignan and Madison have a close latitude, Tampa and Madison have a close longitude, Tampa and San Francisco (SF) have a close elevation and Perpignan is not much distant. That last property distinguishes Madison: with an elevation of 262 m against 35 m, 6 m and 2 m. Madison is considerably higher than the three other cities. Longitude differences impact only the time zone whereas latitude and elevation impact the outdoor temperature. One can see in Figure 5 that Madison has lower outdoor temperature than the others, Tampa has the highest, and variations between Perpignan and San Francisco are rather close. Direct solar radiations of the cities are not plotted because they belong to the same range (between 0 and 1000 $\mathrm{W/m^2}$), all cities have sunny and cloudy days.

Table 3. Geographical coordinates of the four cities.

City	Perpignan	Tampa	San Francisco	Madison
Latitude	42.65	27.97	37.62	43.13
Longitude	2.90	-89.53	-122.40	-89.33
Elevation	35	6	2	262

So, the university model structure is still scalable and satisfying for three of four cities, the identification procedure proposed in this paper is adaptable, but can be improved.

3.2 Test with another building

The objective is to know if the modeling procedure is usable with another entirely different building (geometry, materials, HVAC system, . . .). This building has been created with the EnergyPlus example file generator. It is a U-shape, one floor building, divided in three thermal zones (see Figure 6): two large identical zones of 600 $\mathrm{m^2}$ each and a middle zone of 150 $\mathrm{m^2}$; the total floor area is 1350 $\mathrm{m^2}$.

The HVAC system is a Packaged Terminal Air Conditioner (PTAC) consisting of an outdoor air mixer, a direct expansion cooling coil, an hot water heating coil (via a gas boiler) and a supply air fan.

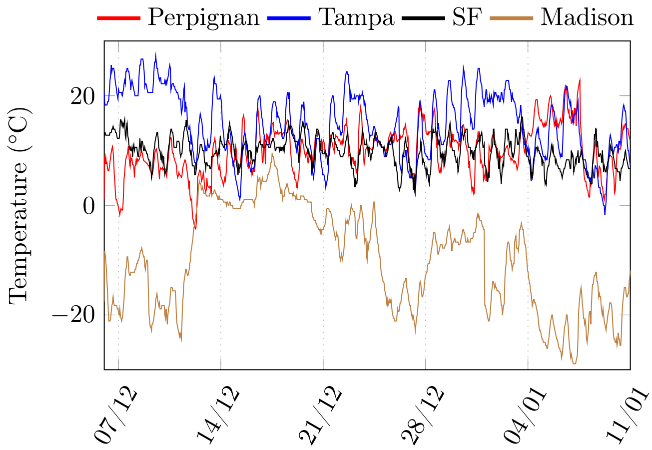

Fig. 5. Outdoor temperatures of the four selected cities.

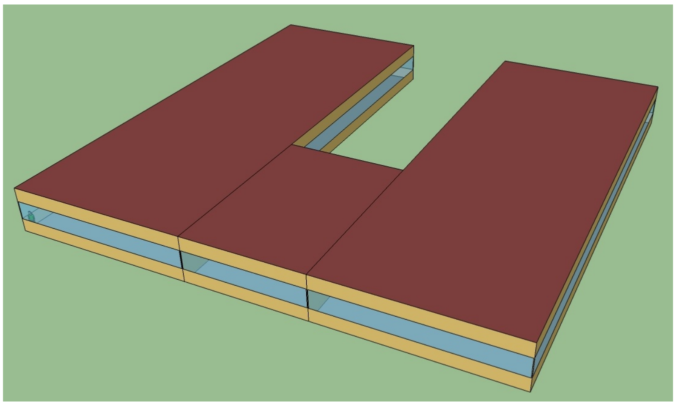

Fig. 6. The three zones of the U-shape example building.

This building is submitted to the same conditions than the first we have studied: the EnergyPlus model is excited during 36 days (from the 6th of December 2012 to the 11th of January 2013), with a time-step of 10 minutes (see subsection 2.3.3).

The Table 4 presents the fit values between validation data and models' output for pure simulation of the example building, in the four same cities than the first building. It can be noticed that:

- once again, the fit values are close or higher to 70 % for pure simulation for all thermal zones except these of Madison, which are close to 60 %;
- this time, the fit values for the building in Perpignan are the best. Sorting the cities in best results order is: Perpignan, Tampa, San Francisco and Madison.

The fit values for simulation prediction are satisfying for every thermal zone in each of the four cities. The model structure and the identification procedure suit this other building with a different HVAC system.

4. CONCLUSION

In the first part of this paper, a black-box model of a building and its thermal zones has been developed. System identification techniques have been applied to rich

Table 4. Fit between validation data and models output: case of the U-shape building in different locations (pure simulation).

	Fit in %			
Location	Perpignan	Tampa	SF	Madison
Zone 1	78	71	73	60
Zone 2	75	72	71	56
Zone 3	74	73	70	55
Whole building	77	73	72	58

input/output data generated using co-simulation between EnergyPlus and Matlab. The model structure, a state-space dynamical model of second order, has been shown to be scalable: the identified models of rooms, floors and the building as a whole exhibited rather satisfying prediction of indoor temperature of the considered thermal zones.

Then in the second part, the model structure has been reused by applying this procedure in the same building but located in three american distant cities, and has showed satisfying results. Finally, the procedure has been repeated on another type of building with a different heating system, in the same cities, and results are satisfying too.

Obviously, it exists several building geometries and heating systems; future work will repeat this procedure and model structure on other or same buildings, may be in other cities, and equipped with the most usual heating system. Future work will also assess the input/output influence and selection in function of the location.

ACKNOWLEDGEMENTS

The authors would like to thank Polytech Montpellier and Pyrescom for their involvement in this project, and also the french meteorological organization 'MeteoFrance', who provided Perpignan's weather data.

REFERENCES

ASHRAE Handbook—Fundamentals. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2009.

Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7):1511–1522, 2011. ISSN 0378-7788. doi: 10.1016/j.enbuild.2011.02.005.

Willy Bernal, Madhur Behl, Truong X. Nghiem, and Rahul Mangharam. MLE+: A tool for integrated design and deployment of energy efficient building controls. In *4th ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings, (BuildSys '12)*, Toronto, Canada, 2012.

Jiři Cigler and Samuel Prívara. Subspace identification and model predictive control for buildings. In *11th International Conference on Control, Automation, Robotics and Vision*, Singapore, December 2010.

Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y. Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. Energyplus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, 2001. ISSN 0378-7788.

Lukáš Ferkl and Jan Široký. Ceiling radiant cooling: Comparison of ARMAX and subspace identification modelling methods. *Building and Environment*, 45(1):205–212, 2010. ISSN 0360-1323. doi: 10.1016/j.buildenv.2009.06.004.

Roberto Z. Freire, Gustavo H.C. Oliveira, and Nathan Mendes. Predictive controllers for thermal comfort optimization and energy savings. *Energy and Buildings*, 40(7):1353–1365, December 2008. ISSN 0378-7788. doi: 10.1016/j.enbuild.2007.12.007.

D. Kolokotsa, A. Pouliezos, G. Stavrakakis, and C. Lazos. Predictive control techniques for energy and indoor environmental quality management in buildings. *Building and Environment*, 44(9):1850–1863, 2009. ISSN 0360-1323. doi: 10.1016/j.buildenv.2008.12.007.

Lennart Ljung. *System Identification—Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999. ISBN 9780136566953.

Lennart Ljung. *System identification toolbox for use with Matlab*. The MathWorks, 6th edition, 2003.

Samuel Prívara, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, Carina Sagerschnig, and Eva Žáčeková. Building modeling as a crucial part for building predictive control. *Energy and Buildings*, 56:8–22, 2013. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.10.024.

Richard Raustad. Creating performance curves for variable refrigerant flow heat pumps in EnergyPlus. Technical report, Florida Solar Energy Center/University of Central Florida, November 2012.

RIDER Project. http://www.rider-project.com/, 2010.

Sullivan Royer, Michaël Bressan, Stéphane Thil, and Thierry Talbert. Modelling of a multi-zone building and assessment of its thermal behaviour using an energy simulation software. In *IEEE International Conference on Automation Science and Engineering*, Madison, Wisconsin, USA, August 2013.

S.A. Klein *& al.* TRNSYS 17: a transient system simulation program, 2010. Solar Energy Laboratory, University of Wisconsin, Madison, USA.

Carina Sagerschnig, Dimitrios Gyalistras, Axel Seerig, Samuel Prívara, Jiři Cigler, and Zdeněk Váňa. Co-simulation for building controller development: the case study of a modern office building. In *Cleantech for Sustainable Buildings – From Nano to Urban Scale (CISBAT)*, volume 2, Lausanne, Switzerland, September 2011.

Jan Široký, Frauke Oldewurtel, Jiří Cigler, and Samuel Prívara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9):3079–3087, 2011. ISSN 0306-2619. doi: 10.1016/j.apenergy.2011.03.009.

Shengwei Wang and Xinhua Xu. Simplified building model for transient thermal performance estimation using ga-based parameter identification. *International Journal of Thermal Sciences*, 45(4):419–432, 2006. ISSN 1290-0729. doi: 10.1016/j.ijthermalsci.2005.06.009.

Michael Wetter. Co-simulation of building energy and control systems with the building controls virtual test bed. *Journal of Building Performance Simulation*, 4(3): 185–203, 2011.