

# Application of system identification methods to implement Comsol models into external simulation environments

M. Gontikaki & A.W.M. van Schijndel\*

Eindhoven University of Technology

\*P.O. Box 513, 5600MB, Eindhoven Netherlands, A.W.M.v.Schijndel@tue.nl

**Abstract:** Full coupling of distributed parameter models, like Comsol, with the lumped models often lead to very time-consuming simulation duration times. In order to improve the speed of the simulations, the idea of using system identification methods to implement the distributed parameters models of Comsol into external simulation environments (SimuLink), is explored. It is concluded that the system identification methods are a valuable tool for such applications, and result in models that not only are easy and fast to derive using the appropriate tools (SI Toolbox of Matlab), but also in models with a very satisfactory performance that offer great reduction of the simulation times.

**Keywords:** system, identification, Comsol, SimuLink

## 1. Introduction

Building performance simulation has important applications such as optimal operation strategies and optimal design of the building and systems. Especially for design purposes it is important that the design parameters of interest have physical meaning, i.e. reflect dimensions, material properties, mass flows and time constants of the building. In many building performance simulation tasks, where the indoor climate and/or the energy demands are to be predicted, certain elements of the building that are expected to have an important contribution to the overall behavior of the system, need to be modeled separately using distributed parameter models like for example Comsol. Full integration of such models with the lumped models that are related to the indoor climate and the building systems often lead to very time-consuming simulation duration times. As an alternative to the above procedure and in order to improve the speed of the simulations, the idea of using system identification methods to implement the distributed parameters models of Comsol into external simulation environments (SimuLink), is

explored [1-3]. Our research methodology was as follows:

### (1) Definition of the problem and goal:

A system whose behavior we would like to study is chosen. The aspects of the system that are of interest in a certain problem context are chosen. The model's structure and input/output variables are specified in relation to the intended future use of the model.

### (2) Creation of data sets:

Input/output data sets are created with the use of appropriate tools that are to be used for the identification. Some of the data sets will be used for the identification (estimation data sets) while others will be used for the verification of the identified model (validation data sets) [4,5].

### (3) System Identification:

A model is estimated in the System Identification Toolbox of MATLAB (SI Toolbox). The 'best' model is chosen among many, according to a selection criterion available in the SI Toolbox ('Fits') [6].

### (4) Model Verification:

The model is introduced into the environment of Simulink in MATLAB, and verified for two distinct cases: when it is used separately and when it is coupled with a lumped model.

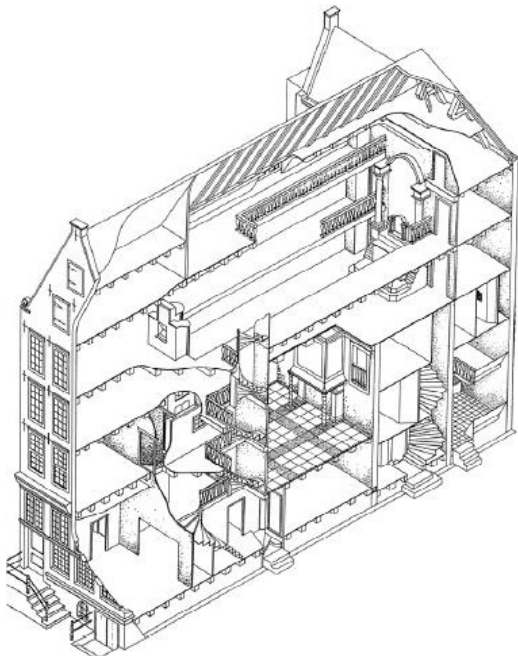
In the following sections these steps will be presented in this sequential order.

## 2. Definition of the problem and goal

The system identification procedure will be implemented to identify a model of the thermal response of a building component belonging to a certain building that is of interest. This building is the Amstelkring Museum, located in Amsterdam. The building is a 350-year old canal house that has been open to visitors for more than 120 years now, and records a large number of visitors annually (in 2004, 75.000 people visited the museum). Considerations regarding the indoor climate conditions (stuffy in the summer, often too dry or moist air) and if these could jeopardize the preservation of the collection led the authorities responsible for the

museum to request an advisory report from the BPS department of the faculty of Architecture of TU/e. In the framework of this investigation all relevant data were gathered and a report was completed in 2006 [7].

As can be seen in Fig.1, the structure of the floors consists of wooden beams running in the transverse direction to the exterior walls and at small spacing distances with one another. Thus in all interior spaces, a significant surface of the structure consists of these wooden beams, which moreover present indications of rot due to the excessively humid air and surface condensation, as reported in the investigation aforementioned.

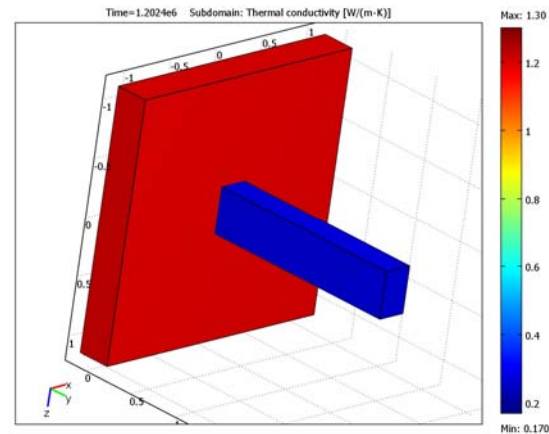


**Figure 1.** Axonometric cross-section of the Amstelkring museum.

For this reason it was decided that a model of the thermal response of these wooden beams would be of interest. More specifically, the surface temperature at the point of the intersection of the wooden beam with the external wall (made of traditional brickwork) is chosen as one of the output signals of the model. Due to the different conductivities of the materials of the beam and wall, a difference in surface temperatures is expected between the two elements, which results in a thermal bridge during winter and possibly to the development of mould on the beam due to moisture condensation. Therefore

the simulation of the surface temperature at that point is considered useful knowledge. Additionally, the second output signal of the model is chosen to be the heat flux to the outside of that area (of a specified surface) consisting of a part of the external wall and the transverse wooden beam, which is supported in the depth of the wall. Although the external wall (and its thermal properties) will certainly control the transmission losses to the outside, and the contribution of the beam to the heat losses/gains will most probably be of minor importance, due to its small surface area in comparison to the total wall surface area, this second output signal is considered because the heat flux variable is almost always of interest in heat transfer problems, even though it is not of real practical use in this case.

A schematic representation of the ‘system’ is given in Fig.2.



**Figure 2.** Geometry of the ‘system’ under study.

The problem is a heat transfer through conduction problem and is represented by the heat equation

$$\rho C \frac{\partial T}{\partial t} - \nabla \cdot (k \nabla T) = Q$$

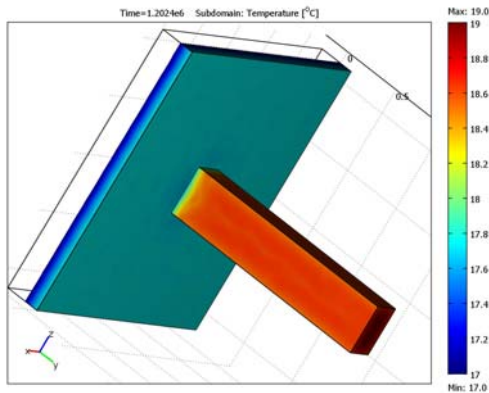
The boundary conditions are represented by:

$$n \cdot (k \nabla T) = h C_e (T - T_e)$$

and by:

$$n \cdot (k \nabla T) = h C_i (T - T_i)$$

where  $T_e$ ,  $T_i$  are the outdoor and indoor air temperatures respectively. The values of the heat transfer coefficients  $h$  are considered constant and equal to:  $h_i=7.7 \text{ W/m}^2\text{K}$  and  $h_e=25 \text{ W/m}^2\text{K}$  for the internal and external surfaces respectively. The model will therefore have two input signals, one for the indoor air temperature and one for the outdoor air temperature, and two output signals, one for the surface temperature at a specified point and one for the heat flux as calculated over the surface area of the wall. The dimensions of the beam are  $0.25 \times 0.40 \text{ m}$  and the surface area of the wall considered (fraction of the real wall dimensions) is  $6.25 \text{ m}^2$ , while its depth is  $0.30 \text{ m}$ . The beam is supported in the wall until a  $0.15 \text{ m}$  depth. The temperature distribution is solved (see Fig3).



**Figure 3.** The temperature distribution

Since the model is created so that at a later stage it is coupled with an indoor climate of the building, which uses one hour time intervals for the input and output variables, the input and output signals of the models are also chosen to have sampling intervals of one hour (i.e. the values of the signals will be calculated in one hour time steps). Furthermore, since the model will be a linear parametric model, we choose the structure of state-space for its representation, because it is generally considered one of the most useful forms for a model, and especially because this model structure can be used in the Simulink environment of MATLAB, where the coupling of the identified model and the indoor climate model will be realized in the final stage. To sum up, after the definition of the problem, we set the goal to be the identification of a linear parametric model in a state-space structure

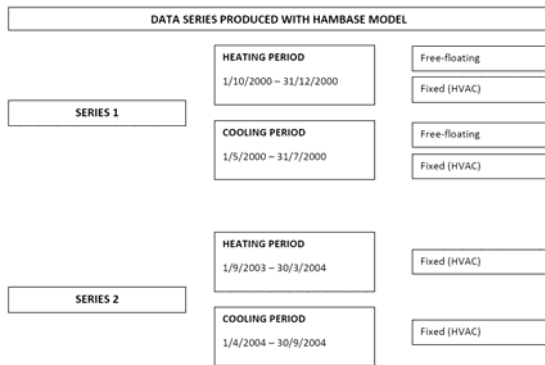
which will include two input and two output signals (multivariable system). The data sets used for the identification procedure will be in hourly intervals, but the model –once identified– should be able to work with continuous time input data, when coupled to the indoor climate model in Simulink.

### 3. Creation of estimation & validation data sets

To proceed with the identification procedure the next step is to create estimation and validation data sets for the input and output signals of the model. Two groups of data sets will be created, which we will call ‘Series 1’ and ‘Series 2’. The data sets of Series 1 will be used to estimate the model in the SI toolbox while the data sets of Series 2 will be used to verify the performance of the estimated model in the environment of Simulink and when the model is coupled with the indoor climate model of the building. The data belonging to Series 1 correspond to a 90-days simulation period, while those of Series 2 to an approximately six months simulation period.

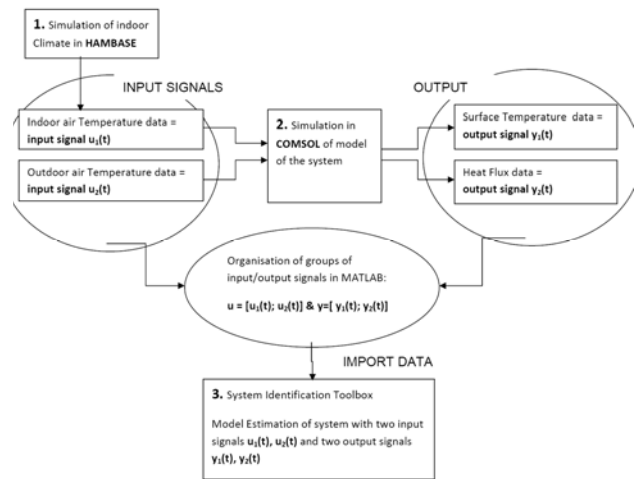
To create data for the two input signals of the system, i.e. data of the indoor and outdoor air temperatures, a model of one zone of the building is developed in HAMBASE [4]. This model (in the form of an m-file in the MATLAB environment) is differentiated to include two cases: one with assumed operation of the HVAC systems available in the building, which have only heating capacities (no cooling), and one without the operation of the heating system. We will refer to the data of the former case as ‘fixed’, and to those of the latter as ‘free-floating’. Two periods are simulated in HAMBASE for each of the above cases: a ‘heating’ period and a ‘cooling’ period, with the first one consisting of winter days and the second one consisting of summer days of the year 2000. The distinction between heating/cooling, fixed/free data was made to explore possible implications of these different data sets to the performance of the estimated model (e.g. if free-floating data were able to give a more accurate model in comparison to fixed data etc). The findings during system identification and if this distinction was necessary or not, will be reported in the next sections. The data of Series 2 are also divided into ‘heating’ and ‘cooling’ data, but the model assumes the operation of the heating

system in both data sets, i.e. Series 2 is only made of ‘fixed’ data. A schematic representation of the created data series for the inputs of the model using HAMBASE is given in Fig.4.



**Figure 4.** Data series created with simulation in HAMBASE of the indoor climate.

After the data series for the input signals of the model are created using HAMBASE, a model of the ‘system’ is made in COMSOL in order to get the corresponding data series for the two output signals (surface temperature and heat flux). The geometry of the system will be represented in 3 dimensions, while the problem to be solved is a heat transfer through conduction problem. The Comsol model is simulated for the indoor & outdoor temperature data of Series 1 and Series 2, as taken from HAMBASE, and gives as output the surface temperature and heat flux data that correspond to these input data. Thus, for every data series we now have the two input signals (taken from HAMBASE) and the two output signals (taken from Comsol) of the model. These input and output signal data series will be properly arranged in MATLAB, in order to proceed with the next step, which is the system identification in the SI Toolbox. Each data set will be divided in half, so that the first half works as the estimation data and the second half as the validation data. Fig. 5 presents the procedure of creating the data required for the system identification schematically.



**Figure 5.** Procedure of creating data for the system identification.

#### 4. System Identification

The data of Series 1 are arranged in input/output groups in MATLAB, in appropriate form to be imported in the SI Toolbox. The input is an array with 2 columns, each corresponding to one input signal, and with as many rows as are the hours of the simulation period. The output is accordingly an array with 2 columns, each corresponding to one output signal, and with the same number of rows as the input array. In total four groups are formed (heating-fixed, heating-free, cooling-fixed, cooling-free) and each one of them is later split in half to form the estimation and validation data sets. In total, and for each one of the above groups, the estimation data and the validation data consist of 40-days period data. For each group state-space models of different orders are estimated in the SI Toolbox. The various models are compared on the basis of how well the estimated output for the validation data set agrees with the ‘measured’ output of the validation data set. The comparison is performed by a function available in the SI Toolbox, which plots the ‘measured’ and estimated output for each model and gives a numerical value of the ‘fit’ of the model. The estimated output of many models can be plotted simultaneously and the one with the highest value of ‘fit’ is the most reliable. In the end of this procedure, we get four state-space models which resulted from the four data groups of Series 1. An outline of the procedure followed for the system identification is given below:

- Opening the application.

By typing ‘ident’ in the MATLAB command window, the SI Toolbox GUI opens up. To proceed, the data to be used (for estimation and validation) must be present as variables in the Workspace. It is useful that the

variables have names such as u1, u2, etc for the input signals and y1, y2, etc for the output signals. In our case u is an array with two columns, one for each input signal, and y is an array with two columns, one for each output signal. In any case, the input must be an array with as many columns as are the input signals to the system. The same applies for the output array in respect to the number of output signals.

- Importing the data.

The estimation and validation data sets are imported in the SI Toolbox in groups of input/output. (Pull down menu Import Data →Time-Domain Signals. Names are given for the input&output, the same as the name of the corresponding variables in the workspace. A name is also given for the data, e.g. 'estimation data', and the starting time of the data and its sampling interval, here equal to 3600s, are stated).

- Examination of the data and pre-processing.

The data are arranged automatically in the icons of the 'data section' of the window. By selecting the 'Time-plot' check-box we can examine the data and decide whether to preprocess them or not. In our case, preprocessing was not necessary.

- Selection of estimation & validation data.

By drag & drop we move the estimation data into the 'working data' icon under the 'operations section' of the window, and the validation data in the 'validation data' icon under the 'models' section of the window.

- Model Settings.

We select to estimate a linear parametric model (pull-down menu Estimate →Linear parametric Models) and choose a state-space structure in the new window that opens. In the 'orders' data field we select a range of orders by typing for e.g. [1:20]. This possibility is offered for certain structures and the SI Toolbox presents you with the order that is the best to start with in a separate window. We state the order that the previous step indicated, and choose 'Simulation' under the 'Focus' pull-down menu. K is fixed to zero, so that no model for the noise is created but instead the disturbances in the system are treated as 'white' noise. The rest of the choices are left to their default values (Initial State: Auto, Covariance :Estimate). By clicking on 'Estimate' a model is estimated and is placed automatically in one of the icons of the 'models section'.

- Examination of the model's properties.

By selecting the 'model output' option, a plot of the 'measured' output (values of the output signals of validation data set), and the estimated output that the model gives when the input data are those of the validation data set, is presented. A numerical value of the fit appears of every model. The higher the value, the better the fit, and the better the model estimates the 'real' values of the output signals of the system, for the specific input signals of the validation data set. The output of many models can be plotted simultaneously, making the comparison an easy task. Models of different orders are compared, and the one

with the better fit is chosen as the best one, to be used later on. Plots of the models' residuals, transient and frequency responses, zeros and poles, are also an option, however here the model selection is based on the numerical value of the fit of the model's estimated output to the 'measured' one.

The above procedure is performed for all four data groups of Series 1 and at the end we have four different estimated models.

In order to evaluate if the models estimated from the four different data groups (heating-fixed, heating-free, cooling-fixed, cooling-free) 'behave' differently, and if the type of data has an influence over the performance of the model, each model is validated in the SI Toolbox with the validation data belonging to the rest of the data groups (e.g. the model estimated from heating-fixed data is also validated\* with heating-fixed, cooling-free and cooling-fixed validation data). By validation we mean that the other validation data are placed in the 'validation data' icon in the SI Toolbox and we get the fit of the model with the model output plot.

Besides the 'fits' of the models with the validation data as an indicator of the models' performance, a calculation of the Crest factor of the input and output signals of each estimation data set was also calculated. The Crest factor is used as an indicator of the power delivered to the system by the signal and is found by dividing the absolute higher value with the RMS of the signal. Table 1 shows the values of the Crest factor of each of the estimation data sets of Series 1. Lower values of the Crest factor are considered to indicate better signal quality. The value of the Crest factor for the estimation data set of every group is presented in Table 1.

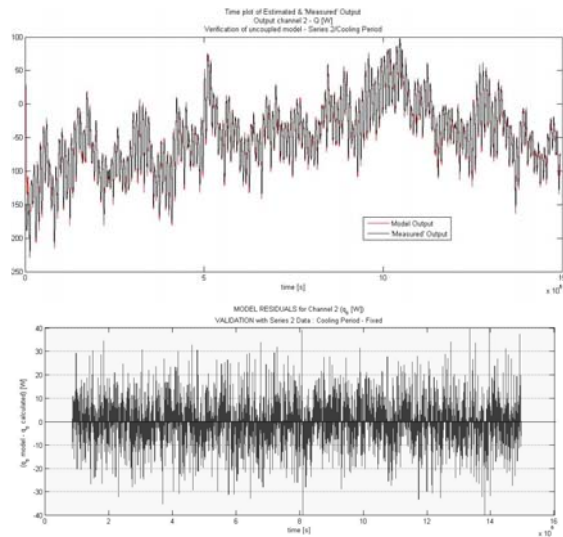
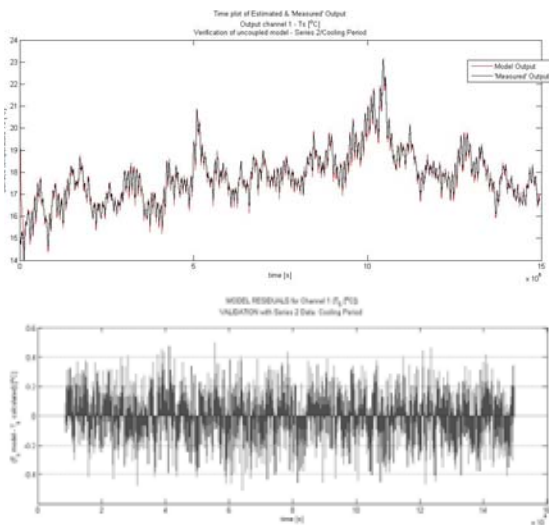
**Table 1.** Crest factor for all estimation data sets.

| CREST FACTOR OF OUTPUT SIGNALS OF ESTIMATION DATA SETS |            |              |
|--|------------|--------------|
|  | Input - Ti | Input 2 - Te |
| Heating period Free-floating data                      | 1,3765     | 1,9165       |
| Heating period fixed data                              | 1,0352     | 1,9165       |
| Cooling period free data                               | 1,3831     | 2,0596       |
| Cooling period fixed data                              | 1,1386     | 2,0596       |

| CREST FACTOR OF OUTPUT SIGNALS OF ESTIMATION DATA SETS |              |               |
|--|--------------|---------------|
|  | Output1 - Ts | Output 2 - qs |
| Heating period free data                               | 1,3852       | 2,7493        |
| Heating period fixed data                              | 1,1104       | 1,628         |
| Cooling period free data                               | 1,4399       | 3,4093        |
| Cooling period fixed data                              | 1,2632       | 2,2305        |

The models are thus considered to perform almost excellent for simulating the first channel (i.e. the surface temperature Ts of the specified

point of the wooden beam) and satisfactory enough for the second channel (i.e. the heat flow  $Q_s$  through the surface of the wall). Between the four models small differences of the fits are observed, with the exception of the model identified with estimation data the Cooling Period/Free-Floating data set (see Table 2.3), which is generally found to have poorest fits. This estimation data set also has the highest values of the Crest-factor for both the input and output signals (see Table 1). The best (lowest) values of the Crest-factor belong to the Heating Period/Fixed data set, however this quality is not reflected in the ‘fits’. Due to the small differences among the different models, it is assumed that whichever model is chosen, besides the one with the highest values for the Crest factors and lowest fits, will perform satisfactorily. Since the model is intended to be used for a building with HVAC function, and thus the input signal of channel 1 (indoor air temperature) will represent ‘fixed’ temperatures, the model with the best fits for the validation data with ‘fixed’ indoor temperatures is chosen: Model identified with estimation data the Cooling Period/Fixed data set To further demonstrate the good fits that this model presents, we import in the SI Toolbox the data from Series 2/Cooling Period and use them as validation data. The ‘model output’ plots for the two channels in this case are shown in Figure 6.



**Figure 6.** Time Plot of the Reference and Estimated Model Output for Output Channel 1 (first) & 2 (third), with corresponding residuals (second) & (fourth)

Finally, it should also be noted, that a certain pattern cannot be distinguished that would testify that for example, a model estimated from the fixed data performs well only when the validation data is fixed as well, or that the model estimated from data of the heating period works only for data belonging to the heating period. This is quite logical, since the system’s function is governed by the input signals but the system’s dynamics are obviously not affected by the way these inputs were generated, or if they belong to a certain range (e.g. the heating period indoor temperatures range between 18-20°C). Therefore the ‘testing’ with the four different data groups is considered an unfortunate choice and an unnecessary procedure. The power that the signals delivered into the system, as expressed by the Crest factor should qualify to determine whether the model will be of acceptable performance or not.

## 5. Model Verification

The procedure of system identification provided a discrete-time state-space model representing the thermal response of the system under study. The intended use of this model is to be coupled with the indoor climate model of the building in the environment of Simulink. In Simulink, the implementation of HVAC controllers has as a

result that the indoor air temperature of the building is subject to changes in time intervals smaller than the hourly intervals and therefore it is important to verify that the model's performance remains satisfactory when the input data are arranged in time-variant intervals. In order to investigate that, the following procedure is applied:

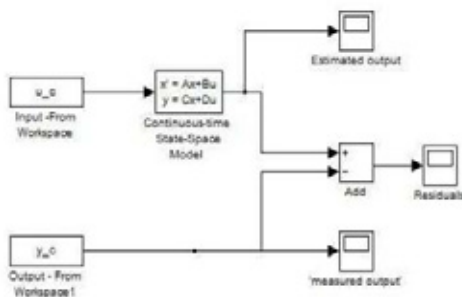
- Transformation of the discrete-time model to a continuous-time model.

This transformation is possible with a single MATLAB command. Initially we export the model from the SI Toolbox by dragging the model's icon to the box 'To Workspace', and then by typing in the Command Line :

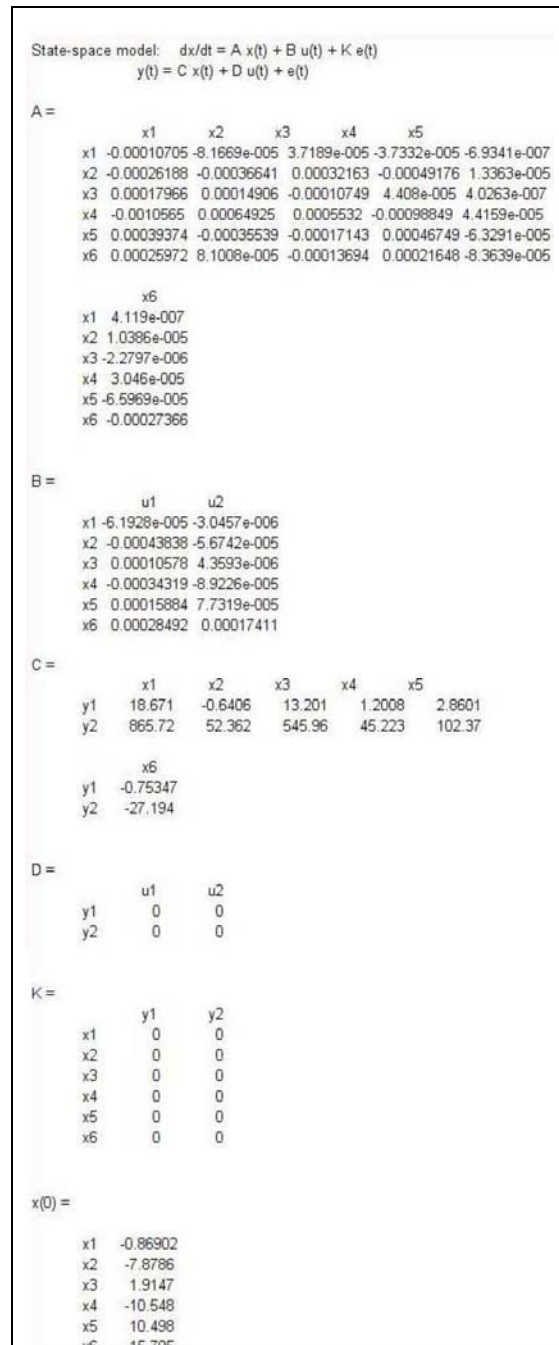
'name of continuous-time model' = d2c('name of the discrete-time model'),

we get the model in continuous-time structure. By double clicking the corresponding icons in the Workspace, we can view the matrices A,B,C,D,K and x(0) of the continuous-time model in the array-editor window. Figure 7 presents the estimated model in discrete-time structure (left) and in continuous-time structure (right).

- Import of the continuous-time model in Simulink and simulation for input data those of Series 2. Free time steps are used (selected automatically by the solver). The model's predicted output is compared with the 'measured' output. It is noted that here the 'measured' output is the one as estimated from Comsol. An interpolation of the 'measured' data for the new times used in Simulink is required before the comparison is possible. Figure 8 presents the uncoupled model in Simulink.



**Figure 8.** View of Simulink diagram with the uncoupled model (state-space block).



**Figure 7.** A,B,C,D,K and x(0) matrices of estimated model in Discret-time and Continuous-time structure

A state-space block is used to import the model. The matrices A,B,C,D and x(0) were created as array variables in the workspace. The input and

output channels of Series 2 data are imported in the model with 'From Workspace' blocks. The input data are used as input in the state-space block which produces the 'estimated output' data series. From this data series the 'measured output' data are subtracted in order to get the residuals. The 'estimated output' is saved in the Workspace. Time plots for the two output channels showing the estimated output from the model and the 'measured' output are produced. Plots of the residuals for the two output channels are also created. The results are similar to Figure 6.

## 6. Application

The state-space block model of the 3D beam represented in figure 8 with matrices from figure 7, was integrated into a indoor climate model of the Amstelkring museum (see figure 1). Due to the significantly reduced calculation time, it was possible to simulate all the necessary scenario's for optimizing the HVAC system. Hereby taking the minimal surface temperature of the monumental wooden beams into account.

## 7. Conclusions

This study had set as a goal to explore the possibilities of applying the system identification procedure to implement distributed parameters models into building models. A system of interest belonging to a certain building was selected and a linear state-space model describing the thermal response of the system was identified. The model was derived using the well-known System Identification (SI) Toolbox in Matlab and was then coupled with the building's indoor climate model in the Simulink environment. The model's performance was verified in the SI Toolbox as well as when performing in the Simulink environment, both in an uncoupled condition and when coupled to the indoor climate model of the building. From the procedure outlined here (explained in detail in the former parts of this report) we have demonstrated that the system identification methods are indeed a valuable tool for such applications, and result in models that not only are easy and fast to derive using the appropriate tools (SI Toolbox of Matlab), but also in models with a very satisfactory performance that offer great reduction of the simulation times.

For the two output channels of the model in the coupled condition, one of them referring to surface temperatures and the other one referring to heat flux data, the time plots of the estimated model's output and the 'measured' one, clearly demonstrate the accuracy and satisfactory performance of the model. The residuals between the estimated and measured output are between -0.4 to 0.4 °C for the surface temperatures and between -30 to 30 W for the heat flux. The verification of the model in the SI Toolbox, produced fits of up to 98% for the first (surface temperature) and up to 87% for the second output channel (heat flux). For a linear system with input signals of similar strength, there is no reason to expect a lower performance.

## 8. References

- [1] A.W.M van Schijndel, P.W.M.H. Steskens, 2006, *The application of system identification for climate control*, 7<sup>th</sup> International Conference on System Simulation in Buildings, Liege December 11-13
- [2] L. Ljung, 1982, *Aspects on the system identification problem*, Signal Processing 4 pp. 445-456
- [3] M.J Jimenez, H. Madsen, K.K. Andersen, 2008, *Identification of the main thermal characteristics of building components using MATLAB*, Buildings and Energy Vol. 43, pp. 170-180
- [4] Martin de Wit, HAMBASE – Heat Air and Moisture for Building and Systems Evaluation – .pdf available at: <http://sts.bwk.tue.nl/hamlab/readers>
- [5] User's Guide: COMSOL Multiphysics, Version 3.4
- [6] Lennart Ljung, System Identification Toolbox For Use with Matlab
- [7] Neuhaus, E., Schellen, H.L., 2006, *Museum Amstelkring TE Amsterdam, Advisory Report, indoor climate control*, TU/e Eindhoven University of Technology BPS-report 05.20.K, Subdepartment BPS