Description and Instructions

for the

Second System Identification Competition

Tests for estimation techniques for the thermal characterisation of buildings and building components. This text is based on the original text published in 1996, however updated for publication on the Web-site of PASLINK in 2002.

Contents

1. Objective

The objective of the second competition is to set up a comparison between alternative techniques and to clarify particular problems of system identification applied to the thermal performance of buildings. The present competition is concerned with wall components and solar radiation. Four different cases are provided for estimation and prediction.

Philosophy

This competition has been organized to help clarify the conflicting claims among many researchers who use and analyze building energy data and to foster contact among these persons and their institutions. The intent is not necessarily only to declare winners but rather to set up a format in which rigorous evaluations of techniques can be made. Because there are natural measures of performance, a rank-ordering will be given. In all cases, however, the goal is to collect and analyze quantitative results in order to understand similarities and differences among the approaches.

2 General Information

To participate in a proper way in the competition these are the steps to follow. The steps are:

(1) read the instructions carefully,
(2) analyze the data from one or more cases, and
(3) send in your results along with an entry form.

Instructions on submitting a return disk with the analysis of the data will be included in section 5. The mailing will also include an entry form that each entrant will need to complete and submit along with the results.

Overview
Research on energy savings in buildings can be divided in three major areas: building components, test houses in real climates and occupied buildings. Three competitions are planned along this line of which this will be the first one.

Data
Five different cases are provided for estimation (see chapter 4). The accuracy of estimations and predictions is one of the criteria for judging this competition. Another criterion will be the methodology which has been applied to solve the problems.
One case is intended for training or self study. The values of the parameters that the organizers used for creating the data series, are given. This Case A, is based on the first competition.
Three cases (cases B, C and E) have been designed in order to test identification methods. The purpose is to test the accuracy of the parameter estimates. One more case has been designed for prediction exercises (Case D) and is based on real data from a well controlled experiment.

Entries
The competition will start on January 1, 1996 and will end on September 30, 1996. Of course people may submit there exercise at any time after this date and ask for an evaluation. The format for the entries, described in the following sections and in the entry form supplied on the data must be followed exactly, or the entry will regretfully have to be rejected.

One can participate in the competition with any number of cases, however the results of every case will be evaluated only when all required output for that case has been submitted.

Results
The results to be produced by the competitors are in the form of estimations or predictions. These results will be submitted to the organizers on the supplied output form and in the required output format. The organizers will evaluate them using the same methods for all submissions.

Following the close of the competition, the results will be analyzed and published in a book. The organisers will not participate in the competition. Selected entries will be sent to a reviewing committee. This committee will invite some ten participants to write formal papers, describing the applied methodology. The overall results will be presented in a book and published by the JRC.

#
Prizes
There are no monetary prizes in the competition (to prevent unnecessary disagreements).

Secrecy
Because this is an open scientific study, entries that provide results without describing the methods used, are not acceptable. On the other hand we recognize that a great deal of labor might have been applied to develop commercially useful applications and full details of those need not be revealed. Sufficient information has to be supplied so that the results can in principle be independently verified. It is not necessary to submit practical implementation details or the computer code. However, we encourage sharing the software at the end of the competition. At a minimum, each participant should supply a flow chart of their methodology together with a one page description of the method.

Future Plans
If the outcome of this competition is promising for a follow up, another competition has been planned on test houses, unoccupied buildings and occupied buildings under real climate conditions. If interest warrants, it is planned that a computer server will operate after the close of the competition as a central repository of interesting data, analysis programs, and the results of other comparative studies.

Organisation

J. Bloem, EC - Joint Research Centre, Ispra, Italy
U. Norlén, KTH Gävle, Sweden
H. Madsen, IMM Technical University of Denmark, Lyngby, Denmark
J. Kreider, JCEM, University of Colorado, U.S.
P. Wouters, WTCB-CSTC, Limelette, Belgium

3. Definitions

Although the nomenclature below, is common in some approaches and not in others, it will provide an understandable nomenclature for this competition.

For each case you are asked to estimate parameters or to predict the heat flow. Consider the following variables as illustrated in figure 1.

$q_i(t)$ is the density of the heat flow rate at the internal surface of the wall, positive from the internal to the external side of the wall, at time *t*, in W/m^2

$q_e(t)$ is the density of the heat flow rate at the external surface of the wall, positive from the internal to the external side of the wall, at time *t*, in W/m^2

$q_i(t)$ is the internal surface temperature at time *t*, in °C

$q_e(t)$ is the external surface temperature at time *t*, in °C

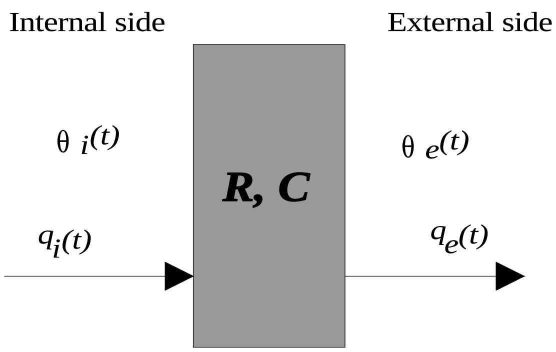

Figure 1. Notation for wall components

The data sets given for estimation consist of simulated measurements of the density of heat flow rate at the internal surface and the two temperature variables. The data sets given for the prediction cases consist of simulated measurements of the two temperature variables only.

For example, the one dimensional solution of the Fourier equation for a homogeneous single layer wall as given by

$$Q(s) = \frac{\sqrt{st}}{R\sinh\sqrt{st}}\cosh\sqrt{st}\,q_i(s) - \frac{\sqrt{st}}{R\sinh\sqrt{st}}\,q_e(s) \qquad (1)$$

where $s = iw$, $i = \sqrt{-1}$ and w is the angle velocity has been used for simulation of the heat flow in Case A and B.

Definitions of the parameters R and C, may be found in DEFSTATEST.PDF as well as the statistical tests which are applied for evaluation are presented.

4. Data Sets

The data sets provided, are compressed DOS-readable ASCII text files and ZIPped. In this chapter the general features of the data sets are described. The order of the signals as they appear in columns in the data file are respectively T_e (external temperature), T_i (internal temperature) and q (heat flow density). In the cases A, B, C and E the time interval of the data is one hour. In case D it is 30 minutes.

4.1 Test on the data series

When starting the analysis of the data sets it is advised to check the data before starting. For that purpose one file *DATASTAT.TXT* with statistical information for all the data series is given too. For every signal in the data set the following statistical information is given :

Maximum, Mean, Minimum and Variance for each signal and Length for the data file

4.2 Description of the cases

The data sets have been chosen to address different sorts of system identification and prediction cases.

Case A is a relatively simple case and intended for training purposes. It is a combination of Case 1 and 2 of the first competition. A challenge for the new participants and a test case for the participants of the first competition. In case A the one dimensional solution for the heat conduction equation for a multi layer wall is applied. The values for R and C which have been used to create the data are given.

Case B is based on simulated data and intended to test any estimation method on the confidence interval of the estimates. In this case the one dimensional solution for the heat conduction equation for a multi layer wall is applied. The problem of identifying the time of a sensor failure is included.

Case C. For this case, 20 series of data are supplied from the same physical system. They differ in the level of noise which has been added to the generated time series for all 3 signals.

Case D. A serious but practical problem is Case D for both estimation as prediction. The collected data are from a field experiment and contain real data. The signals are averaged to 30 minute data from 5 minutes observations of several temeprature and heatflux sensors. Participants are encouraged to try their estimation methods on this case. The second part is dedicated to prediction; the data for the heat flow rate is withheld and should be predicted for the same length of period.

Case E is a one-zone object with a certain volume and walls consisting of two different composition of materials. At a certain moment in the second part of the data series, the capacity will be changed while the thermal resistance remains the same.

The withheld data used for evaluating the submitted estimates and prediction after the close of the competition, will not be available to any of the entrants before the results of the competition are published.

The organizers will compare the estimations and predictions by each contestant with the true parameter values and with the true values of the dependent variable, respectively, that are known only to the organizers.

Case A. A training case

Estimation of R and C of a three-layer wall and estimation of the standard error of these estimates and time of a sensor failure

Part 1:

Real R : 8.651 Real C : 53.472 Real sensor failure time : 668 hours

Data_a92 are identical to *Data_a91* to which have been added data for fourteen additional days. *Data_a92* contains data for 38 days. At some point of time after three weeks, one of the temperature signals changes because of a failure of one of the sensors. This change lasts for the rest of the measurement period.

Sought :

- Estimates of R and C and their standard deviation sR and sC respectively, from *Data_a91* and time of sensor failure in *Data_a92* (minimum requirement is 5 values).

Part 2:

Real R : 3.229 Real C : 81.000

G*iven:* Twenty data-sets (called *Data_a01* to *Data_a20)* for a symmetric three-layer wall (brick-insulation-brick). The heat flow density is assumed to be the average of several heat flow sensors. The external and internal temperatures are assumed to be averages of several temperature sensors each. The data are corrupted by white noise.

Method for data generation: First noise-free data were generated from the solution of the heat conduction equation using the theory of heat conduction through multi-layer walls. After transformation to the time domain noise was added to all three variables. The twenty data-sets are the same except for twenty different realisations of the noise. Each set contains data for 25 days.

Sought :

- Twenty sets of estimates of R and C and estimates (sR and sC) of their standard deviations (minimum requirement is 80 values).

Note :

Case 1 from the first competition has been shortened by about 30 days for the purpose of demonstrating the ability of dynamic methods to assess reliable results over the traditional average method.

The main purpose of Case 2 was to test the confidence interval of the estimates produced by the methods, which should be an important aspect to guarantee a good quality of the analysis result. However evaluation of the F- and t-statistics show that the results from most participants attending the first competition are not that good. It is obvious from the results that more attention should be given by the user of estimation methods to this fact and is therefore included as a training case.

Hints :

- **Study the confidence interval of the estimates generated by the applied method**
- **Study the effect of filtering**
- **Analyse the obtained residue**
- **What is the minimum number of observations necessary to achieve an accuracy of 2% on the estimate for R and 10% on the estimate for C**
- **How well is the used model performing? Cross validation might give additional information**

Case B

Estimation of R and C of a three-layer wall and estimation of the standard error of these estimates and time of a sensor failure

***G**iven:* Three data-sets (called *Data_b1n, Data_b01* and *Data_b2*) for a symmetric three-layer wall (brick-insulation-brick). The heat flow density is assumed to be the average of several heat flow sensors. The external and internal temperatures are assumed to be averages of several temperature sensors each. The data are corrupted by white noise except for *Data_b1n* which is the noiseless generated data series.

Method for data generation: First noise-free data were generated from the solution of the heat conduction equation using the theory of heat conduction through multi-layer walls (*Data_b1n*). After transformation to the time domain noise was added to all three variables *(Data_b01).* Each set contains data for 28 days.

Data_b2 are identical to *Data_b01* to which have been added data for fourteen additional days. *Data_b2* contains data for 42 days. At some point of time after four weeks, one of the temperature signals changes because of a failure of one of the sensors. This change lasts for the rest of the measurement period.

Sought :

- Estimates of R and C and their standard deviation sR and sC respectively, from *Data_b01* and time of sensor failure in *Data_b2* (minimum requirement is 5 values).
- What is the minimum number of observations necessary to achieve an accuracy of 2% on the estimate for R <u>and</u> 20% on the estimate for C (required 1 value, eq. data point) using *Data_b01*

The purpose is to test the ability of the method to give an accurate estimate for R and C and to identify a unexpected behaviour in the data series. Furthermore a comparison with the traditional average method concerning the number of observations necessary for estimating R, can be made

Note that :

Data_b01 is supposed to be for the R and C and sR and sC
Data_b2 for the time of sensor failure
Data_b1n for the minimum length

Case C

Estimation of R and C of a three-layer wall and estimation of the standard error of these estimates

***G**iven:* Twenty data-sets (called *Data_c01* to *Data_c20)* for a symmetric three-layer wall (brick-insulation-brick). The heat flow density is assumed to be the average of several heat flow sensors. The external and internal temperatures are assumed to be averages of several temperature sensors each. The data are corrupted by white noise.

Method for data generation. In this case the method of generation of the data is not given. It can be the the heat conduction equation, a thermal network or a simulation tool. The twenty data-sets are the same except for twenty different realisations of the noise. Each set contains data for 25 days.

Sought :

- Twenty sets of estimates of R and C and estimates of their standard deviations sR and sC (required are 80 values).

The purpose is to test the ability of the method to give reliable information about the accuracy of the estimate.

Case D

Estimation of the R and C of a three-layer wall
Prediction of heat flow density through a three-layer wall under real weather conditions

Data are obtained from a well controlled experimental set up, designed for the purpose of this competition. The tested wall is obtained through adding extra insulation to a normal external wall at a laboratory. The Gas Concrete is a part of the normal construction of the laboratory, while the two insulation layers have been added to create the test wall. The testing area is 2m. wide and 1.2m. high, while the concrete wall is 5m wide 2.5m. high. The wall is facing south-west.
The test wall is a symmetric three layer wall with Gas Concrete insulated on both sides with a glass fibre board layer, see Figure 2. The wall is 204 mm thick. The temperature sensors are positioned at the external side of the glass fibre boards. The Heat Flux meters are mounted in a 3mm wood fibre board which thickness and thermal conductivity were nearly equal to the Heat Flux meter. Data was collected every 5 minutes.

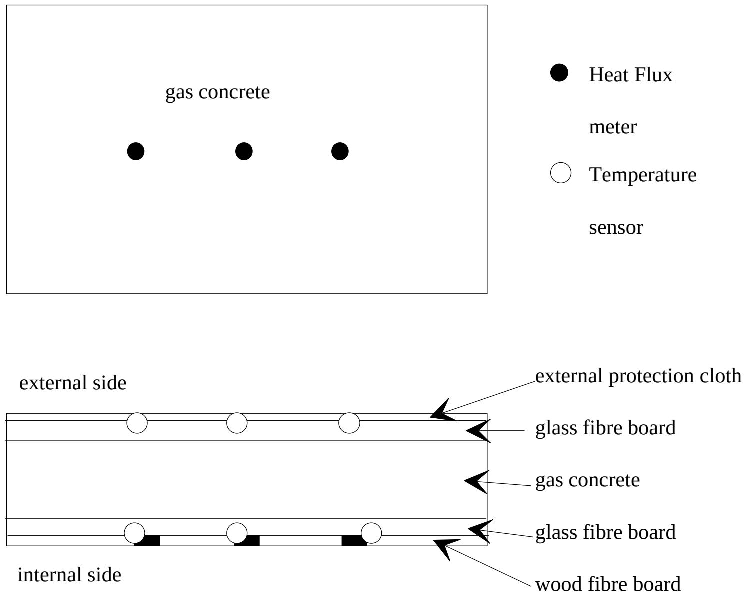

Figure 2. Position of Heat Flux meters and temperature sensors

***G**iven:* Two data sets (called *Data_d1* and *Data_d2*) are created for this three-layer wall.

Data_d1 (the estimation data set) consists of measurements of all three variables (two temperatures and one heat flow). During the measurement period the wall was exposed to solar radiation.

Data_d2 (the prediction data set) consists of measurements of only the internal and external temperatures. During the measurement period the wall was not exposed to solar radiation. A sun shield was used to protect the wall for direct solar radiation. These data are not necessarily a continuation of *Data_d1* temperature sequences. The data sets therefore should be considered as two separated series but from the same physical system.

The heat flow density is assumed to be the average of three heat flow sensors. The external and internal temperatures are assumed to be averages of three temperature sensors each. The external temperature sensors were not directly exposed to the sun but shielded with a special cloth. Each data set contains data for about 31 days and contain each 1500 observations of calculated mean values for every 30 minutes.

Note: concerning the density and moisture in the gas concrete: The density is quite constant and the moisture content is higher at the external side than at the internal side.
Mass transfer in gas concrete is an effect which complicates the analysis from thermal in-situ measurements. Regular measurements have been performed during both periods to obtain information about this phenomena. When the external temperature is high, there is vapour diffusion inwards, and the vapour content increases. When the external temperature is low, there is vapour diffusion outwards, and the vapour content decreases. One might try to assess an effect in the heat flow due to this phenomena.

Sought : Estimates of R and C and estimates of their standard deviation, sR and sC, for *Data_d1* (minimum requirement is 4 values)
Prediction of the heat flow density for *Data_d2* (required is a data set of the heat flow, containing 1500 observations)

Case E

Estimation of R and C from a one zone and the change of thermal mass in the zone.

Given: One data-set (called *Data_e01*) for a one-zone building (3m by 3m) without window apertures. The external temperature is assumed to be averages of several temperature sensors each. There is no solar radiation or other thermal gains to be taken into account.
The one-zone can be a building with very light weight outer walls with far most of the thermal mass inside the building. The heat supplied to the air node in the zone, comes from an electric heater which contributes all its heat to the air by convection.
At some point of time after four weeks, the thermal mass in the zone is changed. One may assume that it is changed by bringing in flags (or stones, bricks), which results in an increase of the thermal mass. This change lasts for the rest of the measurement period.

Method for data generation:
The zone can be presented by the following network model, consisting of two transmittances and two capacitances.

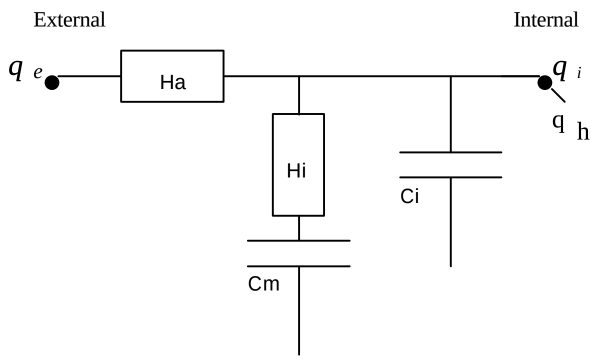

Figure 3. Thermal network for Case E.

The data set contains data for 60 days (1440 hourly observations). The data are corrupted by white noise. The data given are hourly values for the ambient temperature, the indoor temperature and the heat supplied to the air node. The indoor temperature is the simulated output of the applied second oreder model.

Sought :
- Estimates of R and C and their standard deviation sR and sC respectively, from *Data_e01* and time of change of thermal mass (minimum requirement is 5 values).

Optional :
- An estimate of the increased capacity C and the standard deviation sC

Note that the model is given the transmittances but the parameter sought is the ‘overall thermal resistance’ of the envelope of the zone .
The e1 and e11 indicate that two results can be given and is only for our analysis program. So, there are NOT two data series.

Data_e1 (give R, C, sR, sC, time; in this order)
Data_e11 (optional: give C, sC; in this order)

I want to make sure what the symbols mean:

theta-e = external-wall surface temperature

theta-i = internal-wall surface temperature, interpret it as the only temperature that is given is the surface temperature of the thermal mass, which is the same as the internal air temperature. We had to simplify it here and we have not considered any process of convection between the air and the thermal mass (flagstones). It is a simulation and not a real situation.

Concerning Ci and Cm = the internal air Capacitance Ci, and the heavy thermal mass of the building Cm

At some point of time after 4 weeks, the thermal mass has increased.
This would be Cm, some flagstones have been put in the space

The overall capacitance is then C=Cm+Ci

5. Submittals

The minimum required submittal to the organizers consists of :

1. A general description sheet/file. A dummy can be found in the Appendix. Part of the entry form will include your name and address and describe the machine type, operating system and software used.
2. The result format sheet/file. A dummy can be found in the Appendix
3. Method description text sheet/file. In case different methods are applied for the cases it should be made for every method!!!!
4. A hardcopy of a flow chart describing the method.
5. Graphical outputs for illustrative purposes are optional. At maximum one A4 sheet per case.

All output for R, C and the predicted heat flow in Case D, should be given to 4 significant digits. All output for sR and sC should be given in 6 significant digits.

The predicted data series for the Case D, should be put as an additional column added to the right of the two columns in the supplied data file Data_d2.txt. That is, submit a data file that are three columns wide, the rightmost column being the heat flow prediction that corresponds to the temperature values in columns 1 and 2. Supply the data in ASCII format, readable for DOS machines.

The organizers wish the contestants a lot of success.

For more information

Further questions about this competition should be directed to:

Joint Research Centre
Institute of System Engineering and Informatics
J.J. BLOEM Building 45
I - 21020 ISPRA (VA), Italy
tel: +39 0332 789842/789145
fax: +39 0332 789992
E-mail: hans.bloem@jrc.it

APPENDIX

Format of general description sheet/file "DESCRIB.TXT"

Give general information about your name, etc. in a file with the following format (already provided on the diskette, or on FTP, named "DESCRIB.TXT").

Name:
University/Company:
Address:
City:
Country:
Telephone:
Fax:
E-mail:
Type of computer:
Operating system:
Applied software: (give here the name and origin of the program)
Case1 submitted results? (give yes or no)
Case1 method? (if yes, give some keywords on this line)
Case2 submitted results? (give yes or no)
Case2 method? (if yes, give some keywords on this line)
Case3 submitted results? (give yes or no)
Case3 method? (if yes, give some keywords on this line)
Case4 submitted results? (give yes or no)
Case4 method? (if yes, give some keywords on this line)
Case5 submitted results? (give yes or no)
Case5 method? (if yes, give some keywords on this line)
Additional comments ! (up to 20 lines will be read, including this one)

Example of the complete filled sheet "describ.txt". Do not use tab's (only space) and do not add lines. Our software will read this file as 21 lines with an additional text block for comments of at maximum 20 lines!

Name: Hans Bloem
University/Company: Joint Research Centre
Address: Building 45
City: I-21020 Ispra
Country: Italy
Telephone: +39.332.789842
Fax: +39.332.789992
E-mail: hans.bloem@jrc.it
Type of computer: PC 486 DX
Operating system: MS-DOS
Applied software: MATLAB, SIT
Case1 submitted results? yes
Case1 method? prediction error method
Case2 submitted results? yes
Case2 method? prediction error method
Case3 submitted results? no
Case3 method? (if yes, give some keywords on this line)
Case4 submitted results? no
Case4 method? (if yes, give some keywords on this line)
Case5 submitted results? yes
Case5 method? neural network
Additional comments ! (up to 20 lines will be read, including this one)
I don't have any comments

APPENDIX

Format of result sheet/file "RESULT.TXT"

Give your results in a file with the following format (already provided on the diskette, or on FTP, named "RESULT.TXT"). Use a 'space' as separator of the values and no tabs. Do not add lines. The file should exist of 28 lines, not more and not less. The program that will be used to read this file reads these 28 lines in the order given below. All output for R and C should be given to 4 significant digits. All output for sR and sC should be given to 6 significant digits.

Data_b01 (give R, C, sR, sC; minimum length; in this order)
Data_b2 (give data point of sensor failure)
Data_b1n (give minimum length of data series)
Data_c01 (give R, C, sR, sC; in this order)
Data_c02 etc.
Data_c03
Data_c04
Data_c05
Data_c06
Data_c07
Data_c08
Data_c09
Data_c10
Data_c11
Data_c12
Data_c13
Data_c14
Data_c15
Data_c16
Data_c17
Data_c18
Data_c19
Data_c20
Data_d01 (give R, C, sR, sC; in this order)
Data_d02 (give length, maximum, mean, minimum, variance)
Data_e1 (give R, C, sR, sC; in this order)
Data_e11 (optional, give C, sC; in this order)
Data_e2 (give length, maximum, mean, minimum, variance)