4. THE STATISTICAL EVALUATION

J.J. Bloem, M. de Graaf

ABSTRACT

This chapter describes the statistical evaluation of the submittals. Definitions are given about the applied statistics and the results of the evaluation can be found separately for each case. This chapter concludes with the remark that the applied identification technique can be known, but that the applicant needs a certain level of skill to apply the method successfully in order to perform well. Furthermore the competition has shown that the estimation of the uncertainty of the parameter estimates is a difficult one.

4.1. INTRODUCTION

At the closure of the competition, 20 entries were registered as participating officially in the evaluation of the System Identification Competition. Most participants came from Europe, although a considerable requests for the data came from the US. A general division can be made concerning the software that has been applied.

- ready to use software tools, like MRQT, CTLSM, PENTAUR
- calculation environments, like MATLAB, LabView
- self written programs in a language such as QBasic, Fortran, C++

The MATLAB environment (including the toolboxes) appears to be the most popular in the competition. It was registered 6 times. The programs MRQT and CTLSM each scored 3 times.
A submission for all 5 cases was presented by 8 participants and 4 submitted results for only 1 case. The results of the evaluation can be found separately for each case. For reasons of discretion you will find an identity number in the table. A simple ranking system has been used to get insight into the performance of the different submittals (read: participant and his applied method).
An important conclusion can be made when the statistical evaluation is observed: "One needs a certain level of skill in order to apply the method well".

4.1.1 Estimation

Case 1. About 7 performed well. This case was meant as an easy case, with a lot of data and consistent over the whole supplied period. Assessing the time of sensor failure was a quite difficult task for most of the participants.
Case 2. Only 4 performed well and 2 reasonably. To our surprise not all the 7 "best performers of case 1" are among these.
Case 3. Here it turns out that some (9) are able to identify the overall value for the H, but the estimation of the individual thermal parameters is quite difficult. Only 4 came close enough to be successful.

Case 4. This Case and Case 5 were optional for estimation but were evaluated as well. Case 4 in comparison with Case 5 was meant to be easier and this turns out to be the case in the results.

Case 5. Fewer data were supplied but 5 participants were able to give reasonable estimates for both R and C.

4.1.2 Prediction

Case 4. Prediction of the heat flow rate was performed good enough by 9 participants to score half of the points or more. 4 did score the maximum.

Case 5. Although it was intended to be more difficult for prediction, due to the change of the external temperature behaviour, it appeared to be easier than Case 4.

4.1.3 Conclusion

The organisers evaluated all submittals using the same software program. A ranking of performance has been made based on standard statistical measures for each individual case. The estimation and prediction results are evaluated separately. For all cases an overall check has been performed by comparing the estimated result of the overall values for R and C with the values used for the creation of the data. The difference appears in percentages which makes a direct comparison possible among participants. The applied technique can be generally understood, but the applicant needs a certain level of skill to apply the method successfully to perform well. Some have demonstrated this in the competition.

Name	Nationality	Institute or Company
ANDERLIND, Gunnar	Sweden	Gullfiber, Isolation Industry
ARGIRIOU, Thanos	Greece	University of Athens
De MOOR, Maarten	Belgium	University of Leuven, Buildings
GALATA, Alfio	Italy	Conphoebus, Research
GUTSCHKER, Olaf	Germany	University of Cottbus
GUY, Alan	Great Brittain	Pilkington, Glass Industry
HANDEL, Peter	Sweden	University of Uppsala
HANSEN, Lars	Denmark	Technical University of Denmark
HOOYKAAS, Herbert	Netherlands	Technical University of Twente
NEIRAC, François	France	Ecole des Mines de Paris
NIELSEN,Henrik Aalborg	Denmark	Technical University of Denmark
NIELSEN, Jan Nygaard	Denmark	Technical University of Denmark
OESTREICHER, Yves	Switzerland	CUEPE, University of Geneve
OKADA, Eiji	Japan	Waseda University
OULADSINE, Mustafa	France	ENSEM/INPL CNRS
PFLUGER, Rainer	Germany	ITW, University of Stuttgart
REYNOLDS, Odell. R	United States	Airforce Institute of Technology
TROEDSSON, Daniel	Sweden	Linköping Institute of Technology
Van DIJK, H.A.L.	Netherlands	TNO Building Research
XU, Mingzhe	Finland	VTT Building Technology

Table 4.1 List of participants

4.2 PARTICIPANTS

In table 4.1 are listed in alphabetical order the participants and a description of their regular work environment.

In table 4.2 the Identity numbers can be found associated with a description of the applied method. It enables the reader to judge in another way the tables showing the results for the individual cases. The table is ordered following the identity number and is not necessarily in the same order as table 4.1

Identity nr	Method name	Software	Method description
1001	MRQT	program	Thermal network
1002	CTLSM	program	Thermal network, CT
1003	Pentaur	program	Multiple regression, QBasic
1004	IDENT	specific	Thermal network
1005	LabView	toolbox	Thermal network, FFT
1006	MRQT	program	Thermal network
1007	MRQT	program	Thermal network
1008	NN	toolbox	MATLAB, NN
1009	XPrisma	specific	sing. value decomposition
1010	SIMPLEX	toolbox	MATLAB
1011	PEM	toolbox	MATLAB, SIT
1012	CT, own Continous time	toolbox	MATLAB
1013	NN	specific	QBasic
1014	PEM	toolbox	MATLAB SIT
1015	HC,own heat cond. eq.	specific	TPascal
1016	LADY	program	model reduction
1017	PEM	toolbox	MATLAB, SIT
1020	CTLSM	program	Thermal network, CT
1021	CTLSM	program	Thermal network, CT
1022	PEM	toolbox	MATLAB, SIT

Table 4.2 Overview of the applied methods by the participants

4.3 DEFINITIONS

To obtain a correct evaluation the organisers have defined the parameters and tests as given below. The statistical evaluation is used for giving points to the individual results. The following tests are considered:

the **t-test**, which is used for test of unbiased estimates, i.e. to test for correct mean value of the estimates.

the **F-test**, which is used to test whether the variance of the parameter estimates as provided by the estimation method is equal to the empirical variance between stochastic independent runs.

the c^2 **test**, which is used to test whether the ratio between the observed deviation and the estimated standard deviation of the parameter estimates is reasonable.

x_r : the real value x_r is the value that the organisers have used to generate the data
x_e : the estimated value x_e is a point estimate of x_r given by the participant
s_e^2 : the variance of the estimate, also given by the participant

The empirical variance of N point estimates is defined by:

$$s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_{e,i} - \bar{x}_e\right)^2 \quad (4.1)$$

where $x_{e,i}$ is the point estimate in data series *i* and $\bar{x}_e$ is the empirical mean, i.e.

$$\bar{x}_e = \frac{1}{N}\sum_{i=1}^{N}\left(x_{e,i}\right) \quad (4.2)$$

N corresponds to the total number of data series.
The mean of the estimated variance of the parameter estimates equals :

$$\bar{s}^2 = \frac{1}{N}\sum_{i=1}^{N} s_{e,i}^2 \quad (4.3)$$

with $s_{e,i}^2$ defined as the variance of the estimate in data series *i*

Percentage comparison
To obtain a value for inter comparison of the results a percentual deviation from the real value is calculated:

$$\frac{\bar{x}_e - x_r}{x_r} \times 100\% \quad (4.4)$$

this measure appears in the tables for cases 1, 4 and 5.
A first analysis has been made using a percentual comparison of the estimates, R and C with the value as the organisers have used to generate the data. Concerning the thermal resistance R, a good estimate is within 2%, a reasonable estimate is within 5%. For the estimate of the thermal capacity these percentages are 20% and 60%, respectively.

t-test
In order to test whether the estimates are unbiased (have the correct mean value) we define the t-test quantity:

$$z_t = \frac{\left|\bar{x}_e - x_r\right|}{s_x / \sqrt{N}} = \frac{\sqrt{\left(\bar{x}_e - x_r\right)^2}}{\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(x_{e,i} - \bar{x}_e\right)^2}} \quad (4.5)$$

It can be stated that $z_t \in t(N-1)$. This means that the critical region for testing on level a is $\{z > t(N-1)_{1-a/2}\}$. Which results that for N=20 (in the cases 2 and 3) the following critical regions are obtained:

a = 20%	z >1.328
a = 10%	z >1.729
a = 5%	z >2.093
a = 1%	z >2.861
a = 0.1%	z >3.883

Concerning the evaluation, 2 points are given if not rejected on a = 20% and 1 point if not rejected on a = 0.1%.

F-test

In order to test whether the variance of the parameters provided by the estimation tool is equal to the empirical variance the F-test quantity is defined. The F-statistic is given by:

$$z_f = \frac{s_x^2}{\bar{s}^2} = \frac{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{e,i} - \bar{x}_e\right)^2}{\frac{1}{N}\sum_{i=1}^{N} s_{e,i}^2} \tag{4.6}$$

Since the number of observations used in the calculation of $s_{e,i}^2$ is large, the situation exists that $\{z_F \in F(N-1,\infty)\}$. This means that the critical region for testing on level a is $\{z < F(N-1,\infty)_{a/2}$ V $z > F(N-1,\infty)_{1-a/2}\}$.

This means that for N=20 (cases 2 and 3) the following critical regions are obtained:

a = 20%	z <0.613 V z >1.432
a = 10%	z <0.532 V z >1.586
a = 5%	z <0.469 V z >1.729
a = 1%	z <0.372 V z >2.031
a = 0.1%	z <0.258 V z >2.420

Concerning the evaluation, 2 points are given if not rejected on a = 20% and 1 point if not rejected on a = 0.1%.

In the table (for the cases 2 and 3) will appear the percentage deviation, the empirical variance of the estimated parameters, the mean of the estimated variance of the parameters, the F-statistic and the t-statistic.

The c^2 test

For cases were only one data series is available, another test quantity is necessary. The c^2 test is used to test for a reasonable ratio between the deviation and the estimated standard deviation of the parameters and hence to obtain a quantity to evaluate the estimated standard deviation. The test has been applied to the cases 1, 4 and 5.

The c^2 test quantity as defined by:

$$z_{c^2} = \frac{(x_e - x_r)^2}{s_e^2} \tag{4.7}$$

Since the number of observations in each test is large, the approximation is used that $z_{c^2} \in c^2(1)$. The critical region for testing on level a is $\{ z < c^2(1)_{a/2}$ V $z > c^2(1)_{1-a/2} \}$.

For level a = 1% and 20% the critical regions are:

a =20%	z <0.016 V z >2.706
a =1%	z <0.001 V z >7.879

2 points are given if the test quantity is not in the critical region for a = 20% and 1 point if not in the critical region for a = 1%. The estimation of the parameters R and C using data series Data41 and Data51 was optional.

Prediction evaluation

In the two prediction cases 4 and 5 the following variables and measures are defined:

L : the time period of prediction measured in data points

$y_{r,t}$: the real value at time t, as generated by the organisers

$y_{e,t}$: the predicted value at time t, as given by the participant

The accuracy of the predicted data series has been calculated by three statistical measures; the RMSE, the CV and the NMBE.

The RMSE, the root mean square error, can be calculated from :

$$RMSE = \sqrt{\frac{\sum_{t=1}^{L} \left(y_{e,t} - y_{r,t} \right)^2}{L}} \tag{4.8}$$

The CV, the coefficient of variation, is a dimensionless measure of the RMS error:

$$CV = \frac{RMSE}{\bar{y}_r} \tag{4.9}$$

where $\bar{y}_r$ is the average of the L data points

The NRMS, the normalised root mean square error is :

$$NRMSE = \frac{RMSE}{\sqrt{s_r^2}} \tag{4.10}$$

where s_r^2 is the empirical variance of the L data points

Finally the NMBE, the normalised mean bias error, is a dimensionless estimate of the bias of the prediction :

$$NMBE = \frac{\sum_{t=1}^{L}\left(y_{e,t} - y_{r,t}\right)}{\sqrt{s_r^2}} \qquad (4.11)$$

In general the rule is that the smaller these values are, the better that the prediction has performed and approximates the real data series. The NRMSE and NMBE have been used for the prediction evaluation and can be found in the tables. A rule of thumb is: a NRMSE value less than 50% performs reasonable. For the NMBE a value of 8% can be applied for a reasonable result.

Ranking points are given as below indicated

NRMSE	points
< 25%	6
25% < < 50%	4
50% < < 100%	2

NMBE	points
< 2%	6
2% < < 8%	4
8% < < 20%	2

Note concerning the tables: -99 signifies that the value could not be calculated. xxbigxx signifies that the value is too big and irrelevant to be printed.

Points assignment for ranking

A simple straight forward method to assign points to the results was applied. Table 4.3 gives an overview. It should be stressed that the points are evaluated per individual case and does not reflect any degree of difficulty between the different cases. For the organisers it was used to analyse the results in an easy and reliable way and to compare them based on statistical evaluation. It should be noted too that all results which were optional are evaluated half the normal points.

Case 1	**Points**
Evaluation R and C	4
Time sensor failure	4
c^2 test (optional)	2
Total	10
Case 2	
Evaluation R and C	4
F and t-tests R and C	8
Total	12
Case 3	
Evaluation H1, H2, H3, G1, G2	10
F and t-tests H1, H2, H3, G1, G2	20
Total	30
Case 4 and 5	
NRMSE	6
NMBE	6
Evaluation R and C (optional)	2
c^2 test (optional)	2
Total (each case)	16

Table 4.3 Overview of points.

4.4 CASE 1

A total of 15 submissions were available for case 1, of which 12 did give the standard deviation for the estimates as well, which was optional. Only one participant did not give an answer for time of the sensor failure.

Examining the table, one can note that 3 participants did not give correct answers for R, and are even more than 40% wrong. R is almost in all cases over estimated. Also one can conclude that in most cases the C is under estimated, in 5 cases by more than 50%.

Real R : 8.65100 Real C : 53.47200 Real failure time : 1440 hours

	Thermal resistance			Thermal capacitance			Time failure	Ranking criteria
ID no	Est. R	Diff R in %	c^2 test	Est. C	Diff C in %	c^2 test	Diff. hours	Total points
1001	8.658	0.081	0.02463	52.99	-0.901	0.2273	106	6
1002	9.356	8.149	1.23808	32.4	-39.408	17.983	12	4
1003	8.636	-0.173	0.02206	53.778	0.572	0.0886	15	8
1004	9.171	6.011	6.76	50.05	-6.400	2.36113	240	3.5
1005	9.138	5.629	18.249	54.17	1.305	0.26419	-1	7
1006	8.945	3.398	n.a.	34.389	-35.688	n.a.	32	2
1007	8.726	0.867	6.6884	0.1248	-99.767	big nbr	113	2.5
1009	8.775	1.433	n.a.	3.043	-94.309	n.a.	179	2
1010	0.0078	-99.910	n.a.	3.67	-93.137	n.a.	110	0
1011	13.2	52.584	517.335	13.97	-73.874	0.15921	n.a.	1
1015	8.8	1.722	0.35522	57	6.598	5.5319	29	5.5
1016	8.666	0.176	0.13444	60.49	13.125	14.5475	59	5
1020	9.841	13.756	2.18471	43.07	-19.453	3.40758	16	5.5
1021	8.655	0.051	0.00028	55.89	4.522	0.45961	0	9
1022	12.936	49.539	468.009	17.191	-67.850	7788.588	15	2

Table 4.4 Overview of results for Case 1.

The detected time of the sensor failure is considered to be good when it is within 6 hours and awarded 4 points. A reasonable estimated failure is awarded 2 points when it is within 24 hours. Apparently most of the participants had some problems to give a reasonable answer; only two performed very well.
A second analysis has been applied using the c^2 test for those 12 participants who give the standard deviations. The results appear in columns 4 and 7 of the table. For the c^2 test each time 2 points are given if the value is not in the critical region on a =20% and 1 point if not in the critical region on a =1%.
In the table are given the total number of ranking points for the obligatory 3 results, and the optional gained points. From this we may conclude that case 1 has 7 good performers (5 points or more).

4.5 CASE 2

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. 13 submissions for case 2 were evaluated using the tests described in section 4.2 The estimation of R and C was performed by 6 participants quite good and scored half the points or more. However evaluation of the F- and t-statistics show that the results from most participants are not that good. Seven fulfil the test criteria to get some points and only one scores reasonable good. It is obvious from the competition results that more attention should be given by the user of estimation methods to this fact.

Real R : 3.229 Real C : 81.000

ID	Param.	Estim.	Diff.in %	Sx^2	S^2	F-test	t-test	Total points
1001	R	3.232	0.093	0.0	0.00002	4.7127	1.507	
1001	C	79.815	-1.464	0.190	0.4562	0.4166	12.161	6
1002	R	1.580	-51.055	0.0	0.00723	0.0132	753.66	
1002	C	38.766	-52.141	0.065	0.35234	0.1839	741.89	1
1003	R	3.221	-0.228	0.0	0.00139	0.1005	2.778	
1003	C	80.254	-0.921	1.520	2.21765	0.6856	2.705	8
1004	R	3.062	-5.171	0.0	0.02462	0.0023	97.746	
1004	C	756.359	833.776	178.1	7408.38	0.0240	226.298	1
1005	R	2.987	-7.472	0.0	0.01456	0.0148	73.405	
1005	C	79.630	-1.691	0.223	6.03564	0.0369	12.988	2
1007	R	3.193	-1.097	0.0	0.00026	0.3241	17.382	
1007	C	0.154	-99.810	0.0	0.00003	0.0215	448826	2
1010	R	0.005	-99.835	0.0	-99	-99	286791	
1010	C	12.878	-84.102	0.146	-99	-99	798.21	0
1011	R	3.291	1.933	0.0	0.00769	0.0546	13.617	
1011	C	78.275	-3.364	0.217	9801	0.0	26.178	4
1014	R	3.460	7.174	0.004	0.0292	0.1306	16.769	
1014	C	28.859	-64.372	0.062	0.08331	0.7475	934.42	2
1015	R	3.194	-1.065	0.006	0.1255	0.0467	2.010	
1015	C	98.786	21.957	7.352	21.9875	0.3344	29.334	5
1016	R	3.268	1.235	0.011	0.09528	0.1109	1.734	
1016	C	88.028	8.677	134.6	1813.44	0.0743	2.709	6
1021	R	3.239	0.335	0.0	0.00244	0.1413	2.603	
1021	C	39.676	-51.018	0.751	0.11733	6.4021	213.24	5
1022	R	5.287	63.742	0.427	0.06051	7.0541	14.086	
1022	C	53.926	-33.425	8.736	0.0	-99	40.965	1

Table 4.5 Overview of results for Case 2.

4.6 CASE 3

In table 4.6 a and b the results from the different tests for the individual thermal parameters are listed. The last column shows the total number of points for the F- and t-test and the percentual deviation for each parameter.

Real H1 : 1.000 Real H2 : 10.000 Real H3 : 0.100 Real Htot : 0.090
Real G1 : 100.000 Real G2 : 50.000 Real Gtot : 150.000

ID	Param.	Est.	Diff.in %	Sx^2	S^2	F-test	t-test	Total
1001	H1	1000.0	99900	0.0	0.0	-99	-99	
1001	H2	0.091	-99.094	0.0	0.00012	22.64511	70397.27	
1001	H3	11.060	10960.0	0.0	0.0	0.0	-99	
1001	G1	91427.8	91327.8	15158085	0.00012	2.1954	104.905	
1001	G2	49.948	-0.103	0.001	0.00012	0.07716	8.128	7
1001	Htot	0.090	-0.25 %					
1001	Gtot	91477.749	60885 %					
1002	H1	1.053	5.335	0.027	0.0144	1.84338	1.464	
1002	H2	9.948	-0.520	0.020	0.01286	1.55789	1.643	
1002	H3	0.1	0.0	0.0	0.0144	0.0	-99	
1002	G1	100.366	0.366	200.611	0.01286	1.3689	0.116	
1002	G2	50.0	0.0	0.0	0.01286	0.0	-99	20
1002	Htot	0.090	0.45 %					
1002	Gtot	150.366	0.24 %					
1004	H1	74.813	7381.300	1.846	9422.1	0.0002	242.940	
1004	H2	0.083	-99.166	0.0	0.0	0.1007	116058.7	
1004	H3	9.877	9776.950	0.00015	9422.1	0.01184	3625.840	
1004	G1	370.150	270.150	105.375	0.0	0.0005	117.693	
1004	G2	50.995	1.989	0.00494	0.0	0.01412	63.267	2
1004	Htot	0.083	-8.29 %					
1004	Gtot	421.145	180.76 %					
1006	H1	10.953	995.300	0.00004	-99	-99	6775.461	
1006	H2	0.092	-99.079	0.0	-99	-99	80197.23	
1006	H3	10.953	10853.0	0.00004	-99	-99	7388.132	
1006	G1	51.901	-48.099	0.015	-99	-99	1759.848	
1006	G2	51.901	3.803	0.015	-99	-99	69.573	3
1006	Htot	0.091	0.54 %					
1006	Gtot	103.803	-30.79 %					
1007	H1	xxbigxx	xxbigxx	xxbigxx	xxbigxx	0.00013	xxbigxx	
1007	H2	0.090	-99.099	0.000	0.00016	8.66583	70233.06	
1007	H3	11.059	10959.0	0.000	xxbigxx	0.00373	15923.05	
1007	G1	xxbigxx	2454750	xxbigxx	0.00016	0.00017	6.703	
1007	G2	0.179	-99.641	0.000	0.00016	0.03767	1780560	0
1007	Htot	0.089	-0.78 %					
1007	Gtot	xxbigxx	xxbigxx %					
1010	H1	0.000	-99.963	0.0	-99	-99	809128	
1010	H2	7347.950	73379.5	10139088	-99	-99	10.306	
1010	H3	15585.2	15585100	17955931	-99	-99	16.448	
1010	G1	27630.6	27530.6	1067720	-99	-99	119.152	
1010	G2	11150.4	22200.8	56436121	-99	-99	6.608	0
1010	Htot	0.000	-99.59 %					
1010	Gtot	38781.150	25754 %					

Table 4.6 a. Overview of results for Case 3.

ID	Param.	Est.	Diff.in %	Sx^2	S^2	F-test	t-test	Total
1012	H1	4.174	317.390	0.002	0.00257	0.82312	308.684	
1012	H2	5.736	-42.641	0.001	0.00074	1.44729	584.043	
1012	H3	1.211	1110.800	0.001	0.00257	0.83229	178.152	
1012	G1	24.263	-75.737	0.567	0.00074	0.70509	449.896	
1012	G2	56.569	13.138	0.066	0.00074	0.81276	114.590	11
1012	Htot	0.807	795.28 %					
1012	Gtot	80.832	-46.11 %					
1014	H1	0.110	-88.990	0.0	0.0	-99	xxbigxx	
1014	H2	7.190	-28.099	0.011	0.001	16889.4	122.261	
1014	H3	3.652	3551.555	0.019	0.0	17838.1	116.304	
1014	G1	45.361	-54.639	0.00001	0.001	-99	108147	
1014	G2	35.240	-29.519	2.273	0.001	3747.9	43.780	2
1014	Htot	0.105	16.89 %					
1014	Gtot	80.601	-46.26 %					
1015	H1	11.001	1000.100	0.00001	0.00023	0.04085	14531.11	
1015	H2	0.089	-99.110	0.0	0.0	0.87947	39468.31	
1015	H3	11.001	10901.00	0.00001	0.00023	0.04085	15838.78	
1015	G1	50.158	-49.842	0.00027	0.0	0.04339	13578.5	
1015	G2	50.158	0.316	0.00027	0.0	0.04339	43.044	6
1015	Htot	0.088	-2.82 %					
1015	Gtot	100.316	-33.12 %					
1017	H1	1.052	5.190	0.028	0.01232	2.26535	1.389	
1017	H2	9.950	-0.497	0.021	0.00975	2.13409	1.541	
1017	H3	0.100	0.0	0.0	0.01232	0.0	-99	
1017	G1	101.382	1.382	198.536	0.00975	1.56196	0.439	
1017	G2	50.0	0.0	0.0	0.00975	0.0	-99	19
1017	Htot	0.090	0.442 %					
1017	Gtot	151.382	-0.921 %					
1020	H1	1.055	5.519	0.027	0.01811	1.4695	1.513	
1020	H2	9.946	-0.541	0.020	0.01465	1.37554	1.705	
1020	H3	0.100	-0.005	0.0	0.01811	0.71248	4.067	
1020	G1	100.197	0.197	199.566	0.01465	1.12277	0.062	
1020	G2	50.0	0.0	0.0	0.01465	0.0	-99	21
1020	Htot	0.091	0.46 %					
1020	Gtot	150.197	-0.13 %					
1021	H1	0.992	-0.774	0.069	0.00817	8.38524	0.132	
1021	H2	9.967	-0.329	0.0	0.0	5.55241	122.126	
1021	H3	0.101	0.914	0.00001	0.00817	10.09221	1.637	
1021	G1	99.722	-0.278	584.846	0.0	8.33893	0.051	
1021	G2	50.004	0.008	0.0	0.0	0.71459	8.879	17
1021	Htot	0.091	0.75 %					
1021	Gtot	149.726	-0.18 %					
1022	H1	1.070	6.952	0.047	0.02779	1.69323	1.433	
1022	H2	9.919	-0.810	0.044	0.02657	1.6524	1.729	
1022	H3	0.110	10.000	0.0	0.02779	0.0	xxbigxx	
1022	G1	98.296	-1.704	381.138	0.02657	0.11069	0.390	
1022	G2	45.360	-9.280	0.0	0.02657	0.0	-99	14
1022	Htot	0.099	9.61 %					
1022	Gtot	143.656	-4.23 %					

Table 4.6 b. Overview of results for Case 3.

Conclusion

This case was introduced to see if the methods are able to estimate the individual parameters of a thermal network. Only 4 of the participants have obtained reasonable estimates. This is somehow disappointing since many of the available tools claim to be able to estimate the parameters of a thermal network.

The overall value for case 3 is calculated too. This is a direct result of the poor outcome for case 3 for the estimation of the requested parameters. Calculated are $1/H_{tot}=1/H_1+1/H_2+1/H_3$ and $G_{tot}=G_1+G_2$ for both true and estimated values. Note: that the H is a conductance and not a resistance.

4.7 CASE 4

11 estimation submissions and 14 prediction submissions for case 4. Most participants did not have difficulties with the estimation of the R, but showed more problems to obtain a good estimate for R. Once a good model is known from the first data series, the prediction of the heat flow rate should not be a problem, as is demonstrated. Four score the maximum points for the prediction case. Two participants have applied a neural network to predict, but did not perform that good. The points are given in the table in different columns for estimation and prediction respectively. Since estimation of R and C was optional only half of the points is taken into account. A total of 23 out of the maximum of 44 points were scored for estimation and 94 out of 168 points for prediction.

Real R : 3.140 Real C : 26.778

Estimation **Prediction**

ID	Est. R	Diff R in %	c^2 test	Est. C	Diff C in %	c^2 test	points	NRMSE in %	NMBE in %	points
1001	3.108	-1.019	7.11111	5.867	-78.090	149952	1.5	261.52	-13.28	2
1002	3.202	1.975	2.62325	8775.0	32669	1.67771	3	99.76	-3.27	6
1003	3.147	0.223	0.1225	26.497	-1.049	0.01338	3.5	14.46	-0.09	12
1004	3.117	-0.732	0.00588	1021.4	3714.3	1.22133	2.5	35.36	1.31	10
1005	3.132	-0.255	0.04453	28.80	7.552	111.6	3	11.87	0.34	12
1006	3.117	-0.732	-99	33.50	25.104	-99	1.5	146.98	-107.73	0
1008	3.112	-0.892	0.00266	0.0	-100.0	-99	1.5	94.57	-3.29	6
1010	0.003	-99.908	-99	0.069	-99.742	-99	0	109.89	-37.84	0
1011	3.142	0.064	0.01778	4.270	-84.054	0.05169	3	45.26	3.10	8
1013	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	27.12	-19.41	2
1014	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	16.73	1.54	12
1015	3.069	-2.261	8.0656	24.00	-10.374	7.71617	2	21.73	3.04	10
1016	3.135	-0.159	6.25	6.432	-75.980	459946	1.5	15.04	0.01	12
1022	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	51.75	-49.17	2

Table 4.7 Overview of results for Case 4.

4.8 CASE 5

11 estimation submissions and 14 prediction submissions for case 5. It appeared to be a more difficult case for estimation and somewhat easier for prediction. A total of 21.5 points out of 44 were scored for the estimation part of this case and 104 out 168 for the prediction part.

Real R : 2.268 Real C : 22.000

	Estimation							Prediction		
ID	Est. R	Diff R in %	c^2 test	Est. C	Diff C in %	c^2 test	points	NRMSE in %	NMBE in %	points
1001	2.173	-4.201	82.377	21.600	-1.818	3.8522	2	116.72	-60.28	0
1002	1.130	-50.183	13748.45	9159.0	41531	0.40998	1	33.51	-17.65	6
1003	2.273	0.207	0.8836	20.892	-5.036	1.35174	4	15.92	0.17	12
1004	2.243	-1.115	0.02845	33.250	51.136	10.5653	2	62.96	25.75	2
1005	2.257	-0.498	3.05317	43.780	99.00	8242.43	1.5	36.03	-0.83	10
1006	2.422	6.776	-99	18.590	-15.50	-99	1	116.99	-60.77	0
1010	0.003	-99.850	-99	0.160	-99.27	-99	0	96.34	1.29	8
1011	2.270	0.075	0.00723	6.610	-69.95	0.02417	2	83.70	3.89	6
1013	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	22.71	6.46	10
1014	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	16.50	4.69	10
1015	2.260	-0.366	0.07654	21.690	-1.409	0.0961	4	9.95	0.09	12
1016	2.268	-0.013	0.0225	22.955	4.341	24.4845	3	16.37	0.37	12
1020	2.021	-10.902	-99	23.140	5.182	-99	1	18.32	4.24	10
1022	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	76.58	6.80	6

Table 4.5 Overview of results for Case 5.

4.8 OVERVIEW OF ALL RESULTS

		Case 1		Case 2		Case 3		Case 4		Case 5	
		R, C	time	R, C	F, t	R, C	F, t	Est	Pred	Est	Pred
	Maximum points	6	4	4	8	10	20	4	12	4	12
Identity nr	**Method name**										
1001	MRQT	6	0	4	2	2	5	1.5	2	2	0
1002	CTLSM	2	2	1	0	8	12	3	6	1	6
1003	PENTAUR	6	2	4	4			3.5	12	4	12
1004	IDENT	3.5	0	0	1	2	0	2.5	10	2	2
1005	LabView	3	4	2	0			3	12	1.5	10
1006	MRQT	2	0			3	0	1.5	0	1	0
1007	MRQT	2.5	0	2	0	0	0				
1008	NN							1.5	6		
1009	XPrisma	2	0								
1010	SIMPLEX	0	0	0	0	0	0	0	0	0	8
1011	PEM	1	0	4	0			3	8	2	6
1012	CT					2	9				
1013	NN								2		10
1014	PEM			0	2	2	0		12		10
1015	HC	5.5	0	3	2	4	2	2	10	4	12
1016	LADY	5	0	4	2			1.5	12	3	12
1017	PEM					8	11				
1020	CTLSM	3.5	2			8	13			1	10
1021	CTLSM	5	4	3	2	10	7		2		6
1022	PEM	0	2	1	0	6	8				
Total number of points		47	16	28	15	55	67	23	94	21.5	104
Number of participants		15		13		13		11	14	11	14

Table 4.9 Overview of statistical evaluation

4.10 CONCLUSION

Examining the table 4.9 one might conclude from comparison of the same programs or tools that the results obtained by the participants differ. Take for example case 1 and compare the results obtained by the participants using MRQT (Id.nr 1001, 1006 and 1007) or CTLSM (Id.nr 1002, 1020 and 1021). This inter comparison leads us to an important conclusion, that is that one needs a certain level of skill to apply a particular method to obtain good results.

Comparing the results from the prediction cases 4 and 5 one wonders why participants have more problems with one case than the other. The best performers are those who obtain for both predictions a good result.

4.11 REFERENCES

[1] Bloem J.J., Kreider J.F, Madsen H, Melgaard H and Norlen U., ”System Identification Competition; Instructions“, 1994. European Commission, JRC Ispra.

[2] Kreider, J.F. and Haberl, J.S., 1994. “Predicting Hourly Building Energy Usage:’The great Energy Predictor Shootout’-Overview and discussion of results.” ASHRAE Transactions, Vol 100, Pt.2.

[3] Bendat, J.S. and Piersol, A.G., 1986. Random data analysis and measurement procedures, John Wiley and Sons, New York, ISBN 0-471-04000-2