Technical University of Denmark

DTU

Optimal experiment design for identification of grey-box models

Sadegh, Payman; Melgaard, Henrik; Madsen, Henrik; Holst, Jan

DTU Library
Technical Information Center of Denmark

Optimal Experiment Design for Identification of Grey-Box Models

P. SADEGH*, H. MELGAARD*, H. MADSEN* and J. HOLST†

* Inst. of Mathematical Modeling, Technical University of Denmark,
DK-2800 Lyngby, Denmark.

† Dept. of Electrical and Computer Engineering, University of Newcastle,
Newcastle, NSW 2308, Australia; On leave from:
Dept. of Mathematical Statistics, Lund Institute of Technology,
S-22100 LUND, Sweden.

Abstract

In this paper optimal experiment design is investigated for stochastic dynamic systems where the prior partial information about the system is given as a probability distribution function in the system parameters. The concept of information is related to entropy reduction in the system through Lindley's measure of average information, and the relationship between the choice of information related criteria and some estimators (MAP and MLE) is established. A continuous time physical model of the heat dynamics of a building is considered and the results show that performing an optimal experiment corresponding to a MAP estimation results in a considerable reduction of the experimental length. Besides, it is established that the physical knowledge of the system enables us to design experiments, with the goal of maximizing information about the physical parameters of interest.

1. Introduction

The term 'grey-box' is inspired by the use of 'black-box' in connection with identification of systems where any knowledge of the system under consideration from other sources than measured data is not taken into account. In contrast, the grey-box identification is based upon the belief that partial knowledge about the actual system obtained from other sources in a beneficial manner can improve the result of the modeling procedure. Hence the partial information can influence the representation of the system, the design of experiments or the validation of the resulting model, just to mention a few of the points during the total modeling procedure where extra information might prove to be useful.

A basic reason for the interest in promoting the grey-box identification area is a desire to collect and systematize much of the procedures that earlier have been used to incorporate external knowledge, then often termed, e.g. physical modeling or mere intuition. The basic questions are then e.g. how the external knowledge is represented, how it is included in the modeling procedure as a whole or in a part of this procedure or what are the main advantages or disadvantages of the grey-box approach.

In recent years, some efforts have been reported to answer the questions around the grey-box modeling approach, see [1], [2], [3], and [4]. These references view different aspects of the same problem of incorporating partial prior knowledge in the identification and modeling procedure.

This paper focuses on the experiment design problem, i.e. on a part of the total grey-box modeling issue. This is a part where the partial knowledge may prove to be important, since it may help to design informative experiments with the goal of e.g. reducing the experiment time or decreasing the use of external excitation. However, this is also a very crucial issue, since the design is solely based on prior knowledge. The approach to the problem is statistically based. The concept of information is quantified using Lindley's measure of information, and the partial information is represented in a Bayesian framework as a prior distribution of the parameters that are to be estimated. This formulation of partial prior information is well suited to our adopted information criterion in the sense that the parameters are regarded as random variables. This also means that the prior information is embedded in a well founded procedure. Hence, it is possible within this procedure to discuss those qualities of the partial information that really are important, and to see also the limitations of the grey-box approach. It is also shown that this approach to experiment design leads to a more fundamental question about the whole experimental strategy, e.g. the

132

choice of estimators. The same approach to the problem has been pursued in [5], [6], [7]. Finally, and equally important, this can be used to develop methods that for a given system can show where to try to obtain partial information in order to improve the total modeling and identification procedure.

The methods are applied to a physical model of heat dynamics of a building, where exploring the physical knowledge of the system is indispensable in order to estimate some physical parametric constants. It is then investigated how this collection of prior knowledge and partial prior information enables us to design efficient informative experiments.

2. Experiment Design For Grey-Box Models

The basic concept behind the design of optimal experiments is to maximize a function which is a measure of the "goodness" of those experiments.

In a traditional approach, one may regard a suitable scalar function of Fisher's information matrix as the criterion. This is related to maximization of the posterior precision of the unknown parameters, where the posterior precision is merely obtained by performing measurements during the experiment and processing the data statistically. However, the purely statistical processing of data is in many cases not compatible with the physical reality [4].

In this paper, MAP estimator is introduced as a grey-box estimator. However, the grey-box approach can also be used in connection with other estimators where partial prior information about the value of the estimates does not exist or is ignored in the estimation procedure, e.g. in connection with (unconstrained) ML estimator. The latter approach has been thoroughly discussed in [8].

It is well known that if Y is a set of i.i.d. random variables the covariance of any unbiased efficient estimator of the unknown parameter β will asymptotically reach the lower bound of the Cramer-Rao inequality given by the inverse of Fisher's information matrix [9]. However, in *finite* experiment time, the posterior precision will be dependent upon the choice of the efficient estimator.

Fisher's information matrix is defined as

$$M_F = E_{Y|\beta}\{(\frac{\partial \log p(Y|\beta)}{\partial \beta})^T (\frac{\partial \log p(Y|\beta)}{\partial \beta})\} \qquad (1)$$

where $p(Y|\beta)$ is the probability density function of the random variable Y, given β. The following theorem establishes a relation between the posterior covariance of the parameters and Fisher's information matrix when a MAP estimator is used.

Theorem 2.1 *Assume that the prior information about system parameters is embodied in a Gaussian distribution with covariance Σ_{pre}. Also assume that a MAP estimator is used to estimate the unknown parameters based on sampled observations for a model brought into the regression form*

$$y_t = \varphi_t^T \beta + \epsilon_t$$

where $\{\epsilon_t\}$ is a sequence of Gaussian distributed random variables with known covariance, uncorrelated with $\{\varphi_t\}$.
Then the posterior covariance matrix Σ_{post} is given by

$$\Sigma_{post}^{-1} = \Sigma_{pre}^{-1} + M_F \qquad (2)$$

or equivalently

$$\Sigma_{post}^{-1} = \Sigma_{pre}^{-1} + N\bar{M}_F \qquad (3)$$

where M_F is Fisher's information matrix, $\bar{M}_F$ the average information matrix, and N the length of experiment.
PROOF: *See [7]*

The theorem is stated under very restrictive assumptions, e.g. that the output should be written as a linear regression in parameters. However, it can be shown that the theorem is true in more general cases. For a formal proof see [6].

The MAP estimator is asymptotically efficient, since the posterior covariance matrix approaches the lower bound of the Cramer-Rao inequality for $N \to \infty$, see (3).

If the posterior information is obtained applying a ML estimator, Σ_{pre} should be set equal to ∞ since ML and MAP estimators coincide in case the prior distribution of parameters is non-informative [10].

3. Information Theoretic Approach

In this section, the concept of information is related to Lindley's measure of information. The information related criterion is shown to depend upon the choice of estimator.

Definition 3.1 *The entropy of a random variable X having probability density function $p(X)$ is defined to be*

$$H_x = -E_X[\log p(X)]$$

133

Definition 3.2 *Lindley's measure of the average amount of information provided by an experiment ε with data y and parameters β is defined to be*

$$\mathcal{J}(\varepsilon) = H_\beta - E_y[H_{\beta|y}]$$

Theorem 3.1 *With the same assumptions as in Theorem 2.1, maximizing Lindley's measure of the average amount of information, $\mathcal{J}(\varepsilon)$, is equivalent to solving the optimization problem*

$$\begin{aligned} &min_\varepsilon J \\ &J = -E_\beta[\log det\{\Sigma_{pre}^{-1} + M_F\}] \end{aligned} \tag{4}$$

PROOF: *We regard the estimation as a means by which further information about the system parameters is provided. Recalling that MAP is the mode of the posterior distribution, the maximum amount of information with respect to Lindley's measure is obtained by MAP for each experiment.*

Since the prior distribution of the parameters and the distribution of observations given β is Gaussian and the covariance of the residuals is known, the posterior distribution of the parameters will also be Gaussian. Let's denote the posterior mean and covariance by $\bar{\beta}$ and Σ_{post}. From (2) we have

$$\Sigma_{post}^{-1} = \Sigma_{pre}^{-1} + M_F$$

From Definition 3.2

$$\begin{aligned} &\mathcal{J}(\varepsilon) \\ &= -E_\beta\{\log p(\beta)\} + E_y E_{\beta|y}\{\log p(\beta|y)\} \\ &= -E_\beta\{\log p(\beta)\} + E_y E_{\beta|y}\{-\tfrac{n_p}{2}\log(2\pi)\} \\ &\quad -E_y E_{\beta|y}\{\tfrac{1}{2}\log det\Sigma_{post} \\ &\quad +\tfrac{1}{2}(\beta - \bar{\beta})^T \Sigma_{post}^{-1}(\beta - \bar{\beta})\} \end{aligned}$$

As all the other terms obviously are constant, we only focus on the last two terms

$$\begin{aligned} &E_y E_{\beta|y}\{\tfrac{1}{2}\log det\Sigma_{post}\} \\ &= E_\beta E_{y|\beta}\{\tfrac{1}{2}\log det\Sigma_{post}\} \\ &= E_\beta\{\tfrac{1}{2}\log det\Sigma_{post}\} \\ &= -E_\beta\{\tfrac{1}{2}\log det\Sigma_{post}^{-1}\} \end{aligned}$$

The other term can be written as

$$\begin{aligned} &E_y E_{\beta|y}\{\tfrac{1}{2}(\beta - \bar{\beta})^T \Sigma_{post}^{-1}(\beta - \bar{\beta})\} \\ &= E_y E_{\beta|y}\{\tfrac{1}{2}trace\Sigma_{post}^{-1}(\beta - \bar{\beta})(\beta - \bar{\beta})^T\} \\ &= E_y\{\tfrac{1}{2}trace\Sigma_{post}^{-1}E_{\beta|y}[(\beta - \bar{\beta})(\beta - \bar{\beta})^T]\} \\ &= E_y\{\tfrac{n_p}{2}\} \\ &= \tfrac{n_p}{2} \end{aligned}$$

where n_p is the number of parameters. Now using (2) establishes the theorem.

In many practical applications, an approximation of the mean value in (4) is computed by setting the parameters equal to their prior mean values. This approximation simplifies the study considerably. Thus depending on the estimation method or the choice of a Bayesian or non-Bayesian criterion, the following criteria would be of interest:

$$\begin{aligned} J_1 &= -[\log det(\bar{M}_F)]_{\beta=E\{\beta\}} \\ J_2 &= -[\log det(N\bar{M}_F + \Sigma_{pre}^{-1})]_{\beta=E\{\beta\}} \\ J_3 &= -E_\beta[\log det(\bar{M}_F)] \\ J_4 &= -E_\beta[\log det(N\bar{M}_F + \Sigma_{pre}^{-1})] \end{aligned} \tag{5}$$

These criteria demonstrate different levels of including partial prior information about parameter values. Optimization with respect to J_1 results in designs which are strongly dependent upon the prior information. The dependence is even more pronounced in J_2. Therefore, it might be beneficial to perform a sensitivity analysis, i.e. to determine the sensitivity of the design to changes in parameters, when using these criteria. An alternative approach could be choosing Bayesian design criteria like J_3 and J_4.

Table 1 summarizes the relation between the choice of estimators and the optimization criterion, expressed in both a Bayesian and non-Bayesian form.

Table 1: Summary of the optimality criteria.

	ML	MAP
Non-Bayesian	J_1	J_2
Bayesian	J_3	J_4

4. Frequency Domain Design

Consider a linear discrete time system of the form

$$y_t = G_1(q)u_t + G_2(q)\epsilon_t \tag{6}$$

where $\{u_t\}$ and $\{y_t\}$ are the input and output sequences, respectively, and $\{\epsilon_t\}$ is a sequence of i.i.d. Gaussian random variables having zero mean value and covariance Σ. $G_1(q)$ and $G_2(q)$ are transfer functions where q is the forward shift operator. For simplicity, restrict attention to a single input, single output case with $G_2(\infty) = 1$.

Assume that

1) The experiment time is large.

2) The input $\{u_t\}$ is restricted to the class of signals, admitting a spectral representation with spectral distribution function $F(\omega)$, $\omega \in [-\pi, \pi]$.

3) The constraint is taken to be the allowable input

power.
Defining ξ as

$$d\xi(\omega) = \begin{cases} \frac{1}{2}dF(\omega) & \omega = 0 \cup \omega = \pi \\ dF(\omega) & \omega \in]0, \pi[\end{cases} \qquad (7)$$

the power restriction of the input signal can be stated by

$$P_u = \frac{1}{\pi} \int_0^\pi d\xi(\omega) = 1 \qquad (8)$$

It can be shown that the set of all average information matrices is the convex hull of the average information matrices corresponding to single frequency designs [10]. Based on this fact, the following theorem can be proved.

Theorem 4.1 *For the design criteria J_1, J_2, J_3, and J_4 given in (5), optimal designs exist comprising not more than $\frac{n_p(n_p+1)}{2}$ sinusoidal components, where n_p is the number of parameters.*
PROOF: *See [10] for a proof in the non-Bayesian case and [5, 8] for a proof in the Bayesian case.*

Knowing that a finite number of sinusoidal components suffice to reach optimality, numerical optimization algorithms can be easily applied to find the frequencies corresponding to the optimal input sinusoidal components and their corresponding power shares.
For instance corresponding to the criterion J_4, one can formulate the optimization problem

$$min_{\alpha_1,\ldots,\alpha_n,\omega_1,\ldots,\omega_n} \quad f(\alpha_1,\ldots,\alpha_n,\omega_1,\ldots,\omega_n)$$

$$\sum_{i=1}^n \alpha_i = 1$$
$$0 \leq \alpha_i \leq 1 \qquad i = 1,\ldots,n$$
$$\omega_l \leq \omega_i \leq \omega_u \qquad i = 1,\ldots,n$$

where the cost function is

$$f(\alpha_1,\ldots,\alpha_n,\omega_1,\ldots,\omega_n) =$$
$$-E_\beta \log det[\Sigma_{pre}^{-1} + \sum_{i=1}^n N\bar{M}_F(\omega_i)\alpha_i] \qquad (9)$$

$\bar{M}_F(\omega_i)$ is the average information matrix corresponding to single input frequency ω_i, and ω_l and ω_u are our initial guess of lower and upper bounds of the optimal design frequencies. n is the maximum number of sinusoidal components which suffice for optimality. In many cases n is less than the number $\frac{n_p(n_p+1)}{2}$ which is predicted by Theorem (4.1). The methods of including the prior knowledge by considering Bayesian criteria are thoroughly investigated in

[6] and [5].
The technique can be extended to continuous time models with minor modifications: replace the forward shift operator with differentiation operator, and the frequency range $[0, \pi]$ with $[0, +\infty]$.
In the following, the application of MAP estimation and its corresponding optimal experiment based on a physical model of the heat dynamics of a building are illustrated.

5. Application

This application shows how the optimal design depends upon the choice of the estimators and the prior information about the system parameters. It also illustrates that the physical knowledge of the system enables us to design experiments which are informative about the physical parameters of interest.
Throughout the example, the non-Bayesian criteria are considered. The Bayesian versions of the criteria can be investigated as well. However, previous experiments on the model show that the optimal input spectrum exhibits small sensitivity to perturbations in parameter values. Hence, a Bayesian study will in this case only contribute to computational complexity. Besides, there are some other interesting aspects of experimental design that can be studied, even within a non-Bayesian framework, e.g. a study of experimental design using MAP estimators, and these aspects are outlined in the application as follows.
The following set of stochastic differential equations, which originates from physical considerations, provides a model of heat dynamics of a building [11].

$$\begin{bmatrix} dT_m \\ dT_i \end{bmatrix} = \begin{bmatrix} \frac{-1}{R_i C_m} & \frac{1}{R_i C_m} \\ \frac{1}{R_i C_i} & -(\frac{1}{R_a C_i} + \frac{1}{R_i C_i}) \end{bmatrix} \begin{bmatrix} T_m \\ T_i \end{bmatrix} dt$$
$$+ \begin{bmatrix} 0 & 0 & \frac{A_w p_s}{C_m} \\ \frac{1}{R_a C_i} & \frac{1}{C_i} & \frac{A_w(1-p_s)}{C_i} \end{bmatrix} \begin{bmatrix} T_a \\ Q_h \\ Q_s \end{bmatrix} dt$$
$$+ \begin{bmatrix} dw_m(t) \\ dw_i(t) \end{bmatrix} \qquad (10)$$

with the following parameters
- R_i The resistance between the room air and the large heat accumulating medium.
- R_a The resistance between inner walls and the ambient air.
- C_i The capacity of the small heat accumulating medium (inner part of the walls).
- C_m The capacity of the large heat accumulating medium (due to concrete flags placed on the floor). The indoor temperature is measured, and both process and measurement noise terms are considered.
- T_i The temperature of the small heat accumulating medium (indoor temperature).

135

•T_m The temperature of the large heat accumulating medium.
•A_w Effective window area.
•p_s The part of solar radiation which reaches the building.
•T_a The ambient air temperature.
•Q_h Heat supplied by heater (controlled input).
•Q_s Heat supplied by solar radiation.
•$w_m(t), w_i(t)$ Elements of a vector Wiener process.
Assume that the vector

$$\beta = \begin{bmatrix} R_i \\ R_a \\ C_i \\ C_m \end{bmatrix} \tag{11}$$

is chosen as the parameter vector, i.e. the goal of the experiment is to gain information about the elements of β. In other words, the problem of designing the optimal experiments can be formulated as follows. Find an input signal, expressed as a function of time or specified by its power spectrum, that maximizes information about β, or a sub-set of β, according to some suitable information measure, when the means by which the behavior of the system can be observed is recording the indoor temperature (corrupted with noise) as the system's response to excitations in the indoor heat flow supplied by the heater.

In [10], methods to derive the information matrix are presented, and in [5] these methods are extended and applied to the model of heat transfer in a building, where the model is non-linear in parameters. Experiments are suggested which minimize the posterior uncertainty in the elements of β.

For some typical values of the thermal parameters:

$$\begin{aligned} m_{R_i} &= 0.4788(^\circ C/KW) \\ m_{R_a} &= 29.38(^\circ C/KW) \\ m_{C_i} &= 1.183(KWh/^\circ C) \\ m_{C_m} &= 3.987(KWh/^\circ C) \end{aligned} \tag{12}$$

and some typical values of variances of process and measurement noise terms, an optimal power distribution shown in Fig. 1 is obtained. The design is based on the criterion J_1. The figure shows that the optimal power distribution has components both in low and high frequencies, and it corresponds to a slow and a fast dynamics in the system. The slow dynamics is due to the large capacitor C_m, and the fast dynamics is due to the small capacitor C_i.

Now the role of adding prior information is studied by considering different prior uncertainties and using the design criterion related to the MAP estimator. By large uncertainty in C_i, we mean that the prior variances of the parameters are given by

$$\begin{aligned} \sigma_{R_i}^2 &= 10^{-4} & \sigma_{R_a}^2 &= 10^{-4} \\ \sigma_{C_i}^2 &= 10^{-1} & \sigma_{C_m}^2 &= 10^{-6} \end{aligned} \tag{13}$$

and by large uncertainty in C_m, they are given by

$$\begin{aligned} \sigma_{R_i}^2 &= 10^{-4} & \sigma_{R_a}^2 &= 10^{-4} \\ \sigma_{C_i}^2 &= 10^{-6} & \sigma_{C_m}^2 &= 10^{-1} \end{aligned} \tag{14}$$

Considering J_2 and the prior variances given by (13), the optimal power distribution is shown in Fig. 2.
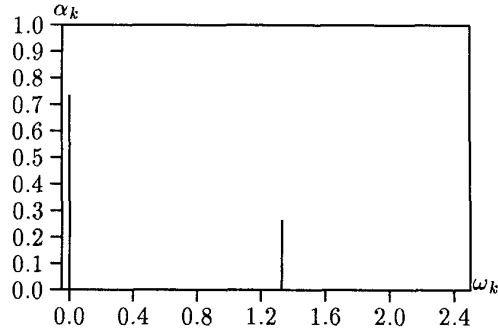

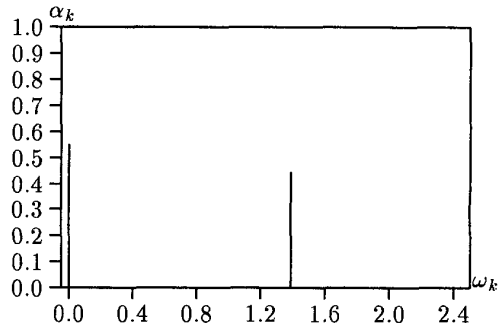**Figure 1:** Optimal power distribution for building modeling, criterion J_1.

Figure 2: Optimal power distribution for building modeling, criterion J_2, large uncertainty on C_i.

It is interesting to note that the optimal input is informative about the most uncertain capacity, since it shifts power to higher frequencies. When the uncertain parameter is the larger capacity, optimizing J_2 will lead to a design which solely consists of a DC-component, i.e. a drastic shift of power towards lower frequencies. In the assessment that follows later, it is illustrated how the designs obtained by optimizing J_2, where an extra amount of prior information is added, together with using the MAP estimator result in a reduction of the experiment time relative to the design obtained by minimizing J_1 and using the ML estimator. However, notice that misguiding results could be caused by wrongly or ill-specified prior information. For instance, if the smaller capacity is quite different from our prior guess, very little information about its value is provided by a DC-excitation, and the posterior estimate of the capacity will shift only slightly from the initial wrong value. This flaw can not be

assigned only to the experimental design, but more generally to a wrong choice of the *whole* experimental strategy, which among others covers the choice of an unsuitable estimator and its corresponding experiment.

Denoting the optimal design by ξ^*, we define PAV by

$$PAV = (det(\Sigma_{post}(\beta, \xi^*)))|_{\beta=E\{\beta\}})^{(\frac{1}{n_p})}$$

as a measure by which it is possible to compare the posterior precisions of different sets of estimates for a specific experiment time. This is equivalent to comparing the mean volumes of the posterior confidence ellipsoids for the parameters, where the posterior knowledge has been gained under different conditions. Note that, prior to the experiment, this mean volume is approximated by setting the parameters equal to their prior mean values. The experiments can be compared either by comparing the experiment lengths for a predefined PAV or by comparing the PAVs for a predefined experiment time. A study of the results in Table 2 and Table 3 show that an optimal experiment corresponding to the criterion J_2 together with the MAP estimation shorten N, the assessed length of experiment, considerably.

Table 2: Experimental design results for a building, ML estimator. PAVs are scaled by 10^6.

MLE		
Prior variances	$N = 1540h$	$PAV = 2.8$
(13)	$PAV = 18$	$N = 10000h$
(14)	$PAV = 18$	$N = 10000h$

Table 3: Experimental design results for a building, MAP estimator. PAVs are scaled by 10^6.

MAP		
Prior variances	$N = 1540h$	$PAV = 2.8$
(13)	$PAV = 2.8$	$N = 1540h$
(14)	$PAV = 3.6$	Not computed

6. CONCLUSION

In the paper, different levels of including prior information about the system, for experiment design, have been discussed. The different levels, which are summarized in Table 1, include the use of the MAP and ML estimators as well as Bayesian and non-Bayesian criteria for optimality of the experiment. Besides, it is shown that the physical knowledge of the system enables us to design experiments which are informative about the physical parameters of interest. The

application of experimental design for grey-box models to the heat dynamics of a building illustrates the role of the physical knowledge and shows how the choice of estimators affects the optimal design. By considering the MAP estimator the optimal design is dependent upon the prior uncertainty of the parameters, and in general the total experiment time is significantly shortened. The application demonstrates the importance of using grey-box methods for experiment design.

References

[1] T. Bohlin. Computer-aided grey-box validation. Technical Report TRITA-REG-8403, Dept of Automatic Control, Royal Institute of Technology, Stockholm, Sweden, 1984.

[2] S. Graebe. *Theory and Implementation of Gray Box Identification.* PhD thesis, Dept of Automatic Control, Royal Institute of Technology, Stockholm, Sweden, 1990.

[3] J. Holst, U. Holst, H. Madsen, and H. Melgaard. Validation of grey box models. In *IFAC Symposium on Adaptive Control and Signal Processing 92,* pages 407–414, 1992.

[4] Herbert J. A. F. Tulleken. Grey-box modelling and identification using physical knowledge and bayesian techniques. *Automatica,* 29(2):285–308, 1993.

[5] Payman Sadegh. Experiment design for grey box models. Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark, 1993.

[6] P. Sadegh, J. Holst, H. Madsen, and H. Melgaard. Experiment design and estimation for greybox models. Technical Report EE9401, Department of Electrical Engineering and Computer Science, The University of Newcastle, Newcastle, Australia, 1994.

[7] P. Sadegh, H. Melgaard, H. Madsen, and J. Holst. On the usefulness of grey-box information when doing experiment design for system modeling and identification. In *10th IFAC Symposium on System Identification and Parameter Estimation,* 1994.

[8] H. Melgaard, P. Sadegh, H. Madsen, and J. Holst. Experiment design for grey-box models. In *Preprints of the 12th IFAC World Congress, Sydney,* volume 2, pages 145–148, 1993.

[9] C. R. Rao. *Linear Statistical Inference and Its Applications.* 2nd edition, Wiley, 1973.

[10] Graham C. Goodwin and Robert L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, 1977.

[11] Henrik Madsen, Henrik Melgaard, and Jan Holst. Identification of building performance parameters. In J.J. Bloem, editor, *Workshop on Advanced Identification Tools in Solar Energy Research,* Non Nuclear Energy, pages 37–60. Commission of the European Communities, DG XII, 1990.