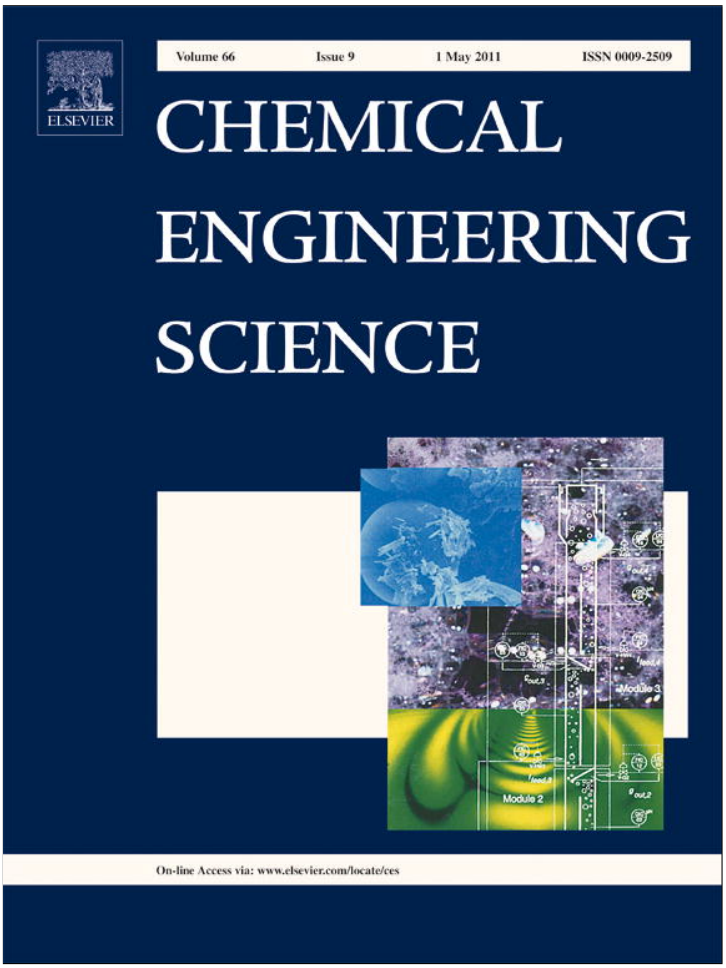

Chemical Engineering Science 66 (2011) 1940–1952

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Experimental design for the joint model discrimination and precise parameter estimation through information measures

André L. Alberton [a], Marcio Schwaab [b], Marcos Wandir Nery Lobão [c], José Carlos Pinto [a,*]

[a] Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária, CP:68502, Rio de Janeiro 21941-972 RJ, Brasil
[b] Departamento de Engenharia Química, Universidade Federal de Santa Maria, Av. Roraima, 1000, Cidade Universitária, Santa Maria 97105-900 RS, Brasil
[c] Instituto de Tecnologia e Pesquisa, Universidade Tiradentes, Avenida Murilo Dantas, 300, Farolândia, Aracaju 49032-490 SE, Brasil

ARTICLE INFO

Article history:
Received 10 August 2010
Received in revised form
21 November 2010
Accepted 20 January 2011
Available online 27 January 2011

Keywords:
Parameter identification
Mathematical modeling
Kinetics
Optimal design
Sequential experimental design
Model discrimination

ABSTRACT

Experimental design procedures for model discrimination and for estimation of precise model parameters are usually treated as independent techniques. In order to conciliate the objectives of both experimental design procedures, the present paper proposes the use of experimental design criteria that are based on measures of the information gain when new experiments are carried out. The proposed criterion depends on the volumes of the confidence regions of the model parameters and presents a number of advantageous aspects, such as the conciliation of the usual experimental design objectives and the fact that the obtained criterion values can be easily interpreted in terms of the information eliminated after carrying out additional experiments. Besides, the proposed design criterion can easily accommodate multiobjective experimental design approaches, as shown in the examples.

1. Introduction

Modeling of chemical processes comprises the development of mathematical expressions to describe the correlations that exist among the observed process variables. Model development can be based on theoretical reasoning or on empirical procedures, but certain unknown model parameters must always be estimated from available experimental data. Several models can be proposed simultaneously and successfully to explain the available data after suitable estimation of model parameters and within the experimental precision; therefore, the *true state of nature* is uncertain and not known. The main objective pursued by experimental design procedures is to approach the *true state of nature* through reduction of the uncertainty content of the modeling problem, described in terms of the number of candidate models and the uncertainties of the respective model parameters (Chernoff, 1972; Lauter, 1974; Dette and O'Brien, 1999).

1.1. Information measures

The concept of information is associated with the number of distinct states that can be regarded as the *true state of nature* (DeGroot, 1962); the higher the number of plausible states, the lower the information content of the analyzed problem. Entropy is an information measure that was originally developed by Shannon (1948) to represent the information content in discrete problems. If the discrete variable Θ_d can assume the values $\{\Theta_d^{(1)}, \Theta_d^{(2)}, \ldots, \Theta_d^{(n)}\}$ with probabilities $\{p_d^{(1)}, p_d^{(2)}, \ldots, pd^{(n)}\}$, then the entropy H of the analyzed discrete problem can be written as

$$H(\Theta_d) = -\sum_{i=1}^{n} p_d(\Theta_d^{(i)})\log(p_d(\Theta_d^{(i)})) \quad (1)$$

For continuous random variables described by the probability distribution function (pdf) $\pi(\Theta)$, the entropy can be written in the form (Cover and Thomas, 1991):

$$H(\Theta) = E_\pi(-\log\pi(\Theta)) = -\int \pi(\Theta)\log(\pi(\Theta))\,d\Theta \quad (2)$$

As the number of plausible states that can be used to represent the analyzed problem decreases, entropy also decreases (desired scenario). For this reason, entropy functions have been used to formulate experimental design procedures (Lindley, 1956; Bard, 1974; Soofi, 1994), although Eq. (2) can be negative and sensitive to reparameterization, which can cause some application problems (Soofi, 1994).

In order to compare two distinct pdfs $\pi^{(1)}(\Theta)$ and $\pi^{(2)}(\Theta)$ and characterize the information gain in a stochastic problem, the Kullback–Leibler divergence criterion (*DKL*, also known as relative entropy) has been widely used (Kullback and Leibler, 1951;

* Corresponding author. Tel.: +55 21 25628337; fax: +55 21 25628300.
E-mail address: pinto@peq.coppe.ufrj.br (J.C. Pinto).

0009-2509/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2011.01.036

Kullback, 1959) in the form:

$$DKL(\pi^{(1)}\|\pi^{(2)}) = E_{\pi^{(1)}(\Theta)}\left(\log\left(\frac{\pi^{(1)}(\Theta)}{\pi^{(2)}(\Theta)}\right)\right) \tag{3a}$$

$$DKL(\pi^{(1)}\|\pi^{(2)}) = \int \pi^{(1)}(\Theta)\ln\left(\frac{\pi^{(1)}(\Theta)}{\pi^{(2)}(\Theta)}\right)d\Theta \tag{3b}$$

where $\pi^{(1)}(\Theta)$ is usually regarded as the true or best representation of the pdf. Eq. (3) is strictly positive, is invariant to reparameterization and can be interpreted in terms of the information contained in $\pi^{(1)}(\Theta)$ that can lead to discrimination between $\pi^{(1)}(\Theta)$ and $\pi^{(2)}(\Theta)$ (Cover and Thomas, 1991). Despite that, Eq. (3) is not commutative, cannot be used for comparison of multiple pdfs and is not a true information measure (Cover and Thomas, 1991), making its use difficult when one does not know which pdf is the true or the best one. In order to overcome some of these drawbacks, the *mutual Kullback–Leilber divergence criterion* (*MDKL*) has been proposed as (Soofi, 1994)

$$MDKL(\pi^{(1)},\pi^{(2)}) = DKL(\pi^{(1)}\|\pi^{(2)}) + DKL(\pi^{(2)}\|\pi^{(1)}) \tag{4}$$

Nevertheless, it is not possible to use and evaluate the *MDKL* when more than two pdfs are considered simultaneously. Besides, as the computation of $DKL(\pi^{(1)}\|\pi^{(2)})$ can lead to spurious numerical values, some additional works proposed the renormalization of Eq. (4) (López-Fidalgo et al., 2007, 2008).

Another classical information measure is the Lindley information (Lindley, 1956). Let us denote ξ^0 as the experimental results already obtained and ξ as the set of plausible experimental conditions, containing one or more experiments that can be performed by the analyst. The Lindley information measures the information gain of the experimental design ξ as the change in entropy of the analyzed variable Θ as (Lindley, 1956; Parmigiani and Inoue, 2009)

$$I(\xi) = E_{\pi_{|\xi}(\Theta)}(-\log\pi(\Theta)) - E_{\pi_{|\xi}(\Theta)}(-\log\pi_{|\xi}(\Theta)) \tag{5a}$$

$$I(\xi) = \int \pi_{|\xi}(\Theta)\ln\left(\frac{\pi_{|\xi}(\Theta)}{\pi(\Theta)}\right)d\Theta \tag{5b}$$

where $\pi_{|\xi}(\Theta)$ represents the posterior pdf of Θ, given the experimental design ξ; and $\pi(\Theta)$ represents the prior pdf of Θ. Therefore, the Lindley information as presented in (5b) is the *DKL* when the posterior and prior pdfs of variable Θ are considered. As a consequence, all drawbacks described below for *DKL*-based information measures also apply to Eq. (5).

In order to formulate a more general measure of the information content of a particular problem, it is important to emphasize that the reduction of uncertainty is related to the elimination of possible *states of nature*, allowing for presentation of the following postulate (DeGroot, 1962):

Postulate I: "The information gain when a set of experiments ξ is performed can be described as the percentage of the states of the nature that were regarded as possible states of nature before performing ξ and that were discarded after performing ξ."

Elimination of 100% of the possible *states of nature* after performing ξ indicates that an absolutely true solution was found by the researcher (which is indeed impossible, given the unavoidable experimental errors of variable measurements) or that preliminary propositions were proved to be incorrect after observation of ξ (meaning that additional propositions are required). It must be observed that the number of possible *states of nature* can increase after performing ξ, indicating that the researcher knew less about the analyzed phenomenon than initially imagined (DeGroot, 1962; Parmigiani and Inoue, 2009). As the information gain must always be positive (Ginebra, 2007), the gain of information must consider the knowledge that the researcher had *before* performing ξ, as evaluated *after* performing ξ (DeGroot; 1962, Ginebra, 2007). In more concrete mathematical terms, let us consider ξ^0 as the set of preliminary experimental data and ξ as the set of experimental results for which one wants to evaluate the information gain. Therefore, if $\psi^{\Theta}(\xi)$ represents the information gain after performing ξ, then (DeGroot, 1962)

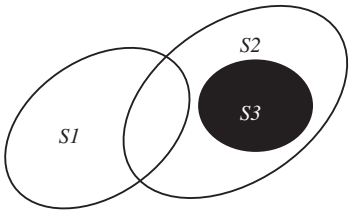

Fig. 1. Information gain in terms of confidence regions of the *states of nature*.

$$\psi^{\Theta}(\xi) = 1 - \frac{\left(\text{Uncertainty after } (\xi^0+\xi)\right)}{\left(\text{Uncertainty after } \xi^0, \text{ assuming that } \xi \text{ is known}\right)} \tag{6}$$

One can assume that the term in the denominator of Eq. (6) (the uncertainty after ξ^0, assuming that ξ was known) can be represented by the union of the uncertainty that the researcher imagined to have at ξ^0 (S1) and the uncertainty that the researcher really had at ξ^0 (S2), as illustrated in Fig. 1 (the uncertainty can be computed as the total number of possible parameter values). S1 represents the set of possible parameter values after ξ^0; however, after performing ξ, the researcher can re-evaluate the set of possible parameter values at ξ^0 as the ones contained in region S2. Finally, after performing the design $\xi^0+\xi$, the researcher concludes that the possible parameters values are the ones contained in region S3. Thus, Eq. (6) can be rewritten as

$$\psi^{\Theta}(\xi) = 1 - \frac{S3}{S1 \cup S2} \tag{7}$$

which can be applied as illustrated in the following illustrative example.

1.2. Illustrative Example 1

Let us assume that an analyst wants to design the number of stages for a distillation column. According to available experimental data (ξ^0), the analyst estimates some model parameters and concludes that the number of stages θ should be in the range 19–21. New experiments are then performed (ξ), allowing the analyst to re-estimate model parameters. When this new set of parameters are used to analyze the data available at ξ^0 (for instance, as initial guesses for a new round of parameter estimation), the analyst concludes that the number of stages θ should be in the range 20–23 after ξ^0. Taking into account the new experimental design ξ, the analyst concludes that the number of stages θ should be in the range 22–23. In this case, assuming that the size of the uncertainty region can be calculated as the range of plausible number of stages ($\Theta=\{\theta\}$), $S1=\{19,20,21\}$, $S2=\{20,21,22,23\}$, $S1\cup S2=\{19,20,21,22,23\}$ and $S3=\{22,23\}$. Thus, according to Eq. (7), it is possible to obtain:

$$\psi^{\Theta}(\xi) = 1 - \frac{2}{5}, \quad \psi^{\Theta}(\xi) = 0.6$$

meaning that 60% of the possible states of nature were eliminated after ξ, which corresponds to the information gain after ξ.

As shown below, Eqs. (6) and (7) can be extended to allow for analysis of problems involving continuous variables, multiple pdfs and multiple candidate models.

1.3. Parameter estimation and the Fisher information matrix

Given some experimental data Z, which are subject to normal fluctuations, and assuming that the proposed model is correct, the maximum likelihood principle states that the parameters can be obtained through minimization of the following objective function (Bard, 1974; Schwaab and Pinto, 2007):

$$F = (Z^{exp} - Z^{calc})^T V_Z^{-1} (Z^{exp} - Z^{calc}) \tag{8}$$

where V_Z represents the covariance matrix of experimental uncertainties and Z^{exp} and Z^{calc} represent the experimental and calculated variables, respectively. During the estimation of model parameters, the experimental uncertainties are transferred to parametric uncertainties, which are usually described by the covariance matrix of parameter estimates V_Θ, as (Bard, 1974; Schwaab and Pinto, 2007)

$$V_\Theta = \left(\left(\frac{\partial Z}{\partial \Theta} \right)^T V_Z^{-1} \left(\frac{\partial Z}{\partial \Theta} \right) \right)^{-1} \tag{9}$$

where the sensitivity matrix $(\partial Z/\partial \Theta)$ must be calculated with the estimated model parameters. V_Θ^{-1} is also known as the Fisher information matrix, as it can be used to describe the information content of the estimated states Θ.

Parameter uncertainties can be described in terms of the *confidence regions of model parameters*, which can be understood as the set of plausible parameter values that can describe the available experimental data in accordance with the distributions of measurement errors and proper statistical tests, after establishing a desired confidence level. Detailed definition and computation of confidence regions of model parameters are presented elsewhere (Koch, 2007; Buzzi-Ferraris and Manenti, 2009).

If normal fluctuation of parameter uncertainties is assumed, which is rigorously valid only when the model is linear in the parameters and/or when experimental errors are small (Chernoff, 1972; Schwaab and Pinto, 2007), the confidence region of parameter estimates has the well-known ellipsoidal shape and V_Θ contains information about the confidence region of model parameters, as illustrated in Fig. 2 for a model containing two parameters θ_1 and θ_2. When the model is nonlinear, the confidence region of parameter estimates may present very complex shapes, as shown by Schwaab et al. (2008a). Even so, for practical reasons the ellipsoidal shape of the confidence region has been always assumed for formulation of experimental design criteria for estimation of precise parameters (Bard, 1974; Atkinson et al., 2007; Schwaab and Pinto, 2007; Franceschini and Macchietto, 2008).

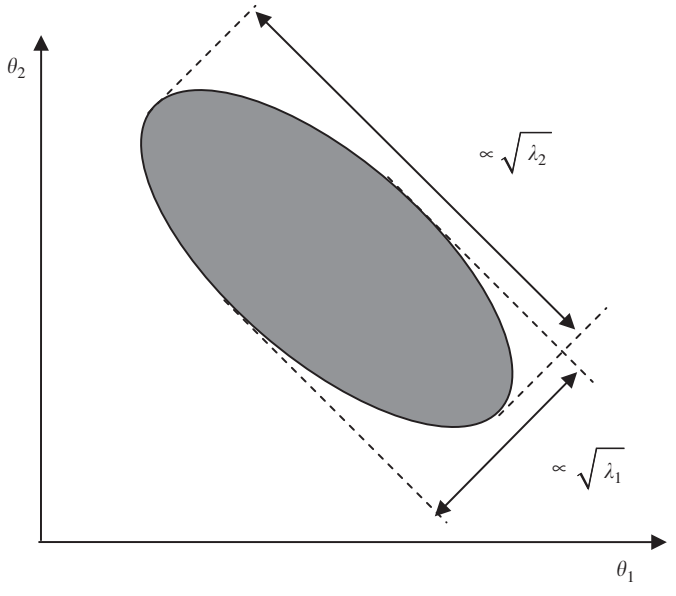

Fig. 2. Confidence region of parameters θ_1 and θ_2 for model $y = x\theta_1 + \theta_2$. The radii of the ellipsis are proportional to the square roots of eigenvalues of the matrix V_Θ.

The adequacy of the model fit is usually evaluated in terms of the chi-square distribution, as (Bard, 1974; Schwaab and Pinto, 2007)

$$\chi^2_{N-Np,(1-\alpha)/2} \le F \le \chi^2_{N-Np,(1+\alpha)/2} \tag{10}$$

where χ^2 represents the chi-square function calculated with N–Np degrees of freedom and confidence level α, N represents the number of experimental points and Np represents the number of parameters of the model. Based on Eq. (10), Schwaab et al. (2006) proposed the definition of model probabilities in the form

$$P = 1 - \Pr(\chi^2_v = F) \tag{11}$$

where P represents the model probability and $\Pr(\chi^2_v = F)$ is the cumulative pdf of the chi-square function evaluated at $\chi^2_v = F$. The analyzed model can be regarded as good if the model probability is inside the range:

$$\frac{(1-\alpha)}{2} \le P \le \frac{(1+\alpha)}{2} \tag{12}$$

1.4. Experimental design

Criteria for design of experiments for estimation of precise parameters are normally associated with the minimization of some metric of the posterior covariance matrix of parameter uncertainties $\hat{V}_\Theta$. The matrix $\hat{V}_\Theta$ differs from matrix $\hat{V}_\Theta$ because $\hat{V}_\Theta$ is estimated with the help of available models and respective parameter values at the analyzed experimental condition, which means that some prior knowledge of the model and respective parameters is assumed. The matrix $\hat{V}_\Theta$ can be calculated as (Bard, 1974; Atkinson et al., 2007)

$$\hat{V}_\Theta = \left(\left(\frac{\partial Z(\xi)}{\partial \Theta} \right)^T_{Z^{calc}} V_Z(\xi)^{-1} \left(\frac{\partial Z(\xi)}{\partial \Theta} \right)_{Z^{calc}} + {V_\Theta}^{-1} \right)^{-1} \tag{13}$$

where the sensitivity matrix $(\partial Z(\xi)/\partial \Theta)$ must be computed at ξ. Experimental design criteria for estimation of precise parameters are generally defined as the minimization of some function of the eigenvalues of matrix $\hat{V}_\Theta$: summation of the eigenvalues (*A-optimal*, trace criterion), multiplication of the eigenvalues (*D-optimal*, volume criterion), ratio between the maximum and minimum eigenvalues (*E-optimal*, shape criterion), among others. Some other criteria are defined in terms of the eigenvalues of the covariance matrix of model prediction uncertainties, although application of different criteria frequently lead to similar results, as stated by the General Equivalence Theorem (Kiefer and Wolfowitz, 1960; White, 1973; Bard, 1974; Atkinson et al., 2007).

In order to avoid the inversion of the covariance matrix of model parameters, the Fisher information matrix (J) is frequently used and defined as

$$J = (\hat{V}_\Theta)^{-1} = \left(\frac{\partial Z(\xi)}{\partial \Theta} \right)^T_{Z^{calc}} V_Z(\xi)^{-1} \left(\frac{\partial Z(\xi)}{\partial \Theta} \right)_{Z^{calc}} + V_\Theta^{-1} \tag{14}$$

The extension of the previously described criteria to Eq. (14) is straightforward, as the eigenvalues of J are the inverse of the eigenvalues of $\hat{V}_\Theta$.

When poor prior knowledge of the models and respective parameters values is available, the efficiency of experimental designs based on Eqs. (13)–(14) can be significantly reduced (Asprey and Macchietto, 2002). In these cases, robust experimental designs may be required, assuming that model parameters are known in a range $\Theta_{min} \le \Theta \le \Theta_{max}$ (Dette et al., 2005) and

defining the efficiency of estimation as (Atkinson et al., 2007)

$$eff_E(\xi,\Theta)=\left(\frac{|(J(\xi,\Theta))|}{\max_\xi|(J(\xi,\Theta))|}\right)^{1/Np} \tag{15}$$

where $|.|$ represents a measure (usually the determinant) of the Fisher information matrix and $1/Np$ is a normalization factor (Atkinson et al., 2007). The experimental design is carried out with the help of the Maximin, introduced by Dette (1997), or Bayesian criteria, in the form (Dette et al., 2005, 2007):

$$\xi_{otm}=\max_\xi\min_\Theta eff_E(\xi,\Theta) \tag{16}$$

$$\xi_{otm}=\text{argmax}_\xi\int\pi(\Theta)eff_E(\xi,\Theta)d\Theta \tag{17}$$

where $\pi(\Theta)$ represent the pdf of Θ. Eqs. (16), (17) pose a very hard optimization task, as multiple Fisher information matrices, determinants and integrals must be computed. Nevertheless, the use of Eqs. (16) and (17) can lead to significant reduction of the experimental efforts required to estimate the model parameters with good precision (Dette et al., 2005). The above criteria have been extensively studied for classes of polynomial models (Dette, 1993, 1994; Dette and Studden, 1995).

Experimental design procedures for model discrimination usually seek for experimental conditions where divergence (D) among model predictions is highest. Published criteria propose that model predictions should be performed with estimated model parameters (Hunter and Reiner, 1965; Buzzi-Ferraris and Forzatti, 1983; Schwaab et al., 2006; Atkinson and Fedorov, 1975a, 1975b), and the efficiency of a discrimination function can be obtained as

$$eff_D(\xi,\Theta^{(n)})=\frac{D(\xi,\Theta^{(n)})}{\max_\xi\left(D(\xi,\Theta^{(n)})\right)} \tag{18}$$

Based on principles of the Information Theory and following a Bayesian approach, Box and Hill (1967) defined a set of model probabilities and took into account model uncertainties in order to formulate a Kullback–Leibler divergence measure of the model responses. Buzzi-Ferraris and Forzatti (1983) showed that the divergence measure defined by Box and Hill presented many undesired mathematical properties, such as the dependence on experimental ordering; for this reason, Buzzi-Ferraris and co-workers introduced new design criteria based on the divergence among model responses, taking into account model uncertainties, but without use of model probabilities (Buzzi-Ferraris and Forzatti, 1983; Buzzi-Ferraris et al., 1984, 1990). Recent propositions (Otsu, 2008; López-Fidalgo et al., 2008; Tommasi, 2007; Tommasi and Lopez-Fidalgo, 2010) also make use of the Kullback–Leibler divergence criterion. In some cases, a normalization of the Kullback–Leibler divergence criterion is necessary to avoid computation of spurious numerical values (López-Fidalgo et al., 2007, 2008), as already described. More recently, Alberton et al. (submitted for publication) proposed that each model could be regarded as a possible *true state of nature*, so that the discriminant D should contain two distinct terms: the number of models discriminated after performing the experimental design ξ and the relative discrimination efficiency, as described in Eq. (18).

Seemingly, Hill et al. (1968) were the first to propose the conciliation of experimental design criteria for precise parameter estimation and model discrimination. They proposed the maximization of a weighted sum of efficiencies for discrimination and precise parameters estimation, although weighting values were defined heuristically as a function of the number of experiments. Atkinson (2008) proposed a similar approach, based on the multiplication of the design efficiencies, including the weighting values in the optimization scheme. Tommasi (2009) followed a similar approach, based on the Kullback–Leibler divergence criterion. Biswas and Chaudhuri (2002) presented a procedure for the simultaneous experimental design for model discrimination and precise parameter estimation for a class of nested models.

Lauter (1974) proposed a criterion for robust experimental design that can be used when multiple models are considered simultaneously, as

$$\xi_{otm}=\text{argmax}_\xi\sum_{n=1}^{M}\log\left|\left(J(\xi,\Theta^{(n)})\right)\right| \tag{19}$$

Such criterion has been extensively explored for polynomial models in the works of Dette (1990, 1992). Similar criteria were proposed by different authors (Atkinson and Donev, 1992; O'Brien and Rawlings, 1996), after modification of the weighting factors as

$$\xi_{otm}=\text{argmax}_\xi\sum_{n=1}^{M}\log\left|\left(J(\xi,\Theta^{(n)})\right)\right|^{w^{(n)}/Np^{(n)}} \tag{20}$$

where $w^{(n)}$ represents the weight (importance) given to model n and $Np^{(n)}$ represents the number of parameters of model n.

Atkinson and Fedorov (1975a, 1975b) presented a robust optimal design for model discrimination, where one model is considered to be the true one and re-estimation of the parameters values is performed for all the remaining models. This procedure was analyzed in detail by Dette and Titoff (2009) in a series of practical problems. In other approaches that do not consider one model as the true one and do not re-estimate the parameters of the models, Schwaab et al. (2008b) and Donckels et al. (2009) showed that model discrimination can be improved when the covariance matrix of parameter estimates is updated during the experimental design procedure. This was also supported by Alberton et al. (2010), who proposed a multiobjective formulation of the simultaneous experimental design problems for parameter estimation and model discrimination.

In all previously analyzed procedures, a great number of objective functions can be formulated for estimation of precise model parameters of multiple models, discrimination of multiple model candidates, simultaneous estimation of model parameters and discrimination of rival models, etc. As a consequence, multiobjective formulations of the experimental design problem have been proposed by Cook and Wong (1994) and Clyde and Chaloner (1996). Particularly, Donckels et al. (2010) considered a multiobjective formulation when more than one model is analyzed simultaneously and concluded that it may be preferable to perform the design for estimation of precise model parameters than for model discrimination when the initial uncertainty of model parameters is high. Alberton et al. (2010) also analyzed the simultaneous experimental design problems for parameter estimation and model discrimination and concluded that both objectives can be frequently conciliated.

In a complex scenario, when multiple rival models are considered simultaneously and multiple objectives are pursued, one can wonder whether the analyst can indeed understand the meaning of the proposed optimization problem and of the selected experimental condition. Therefore, it seems convenient to develop a design criterion that can be easily interpreted by the analyst and hopefully conciliate the different pursued objectives.

2. Conciliating model discrimination and precise parameter estimation through information measures

Let us assume that M rival models are considered, that some experimental data Z^{exp} from experimental design ξ^0 are available and that model parameters can be estimated.

The volume of the confidence region of parameter estimates obtained for model n at ξ^0 is $VCR^{(n)}(\xi^0)$. As a new experimental design ξ may reveal that the parameter uncertainty was different from imagined (see Fig. 1), $VCR^{(n)}_{|Z(\xi)}(\xi^0)$ represents the volume of the confidence region of parameter estimates for model n at ξ^0, but with knowledge of ξ. Thus, let us denote $VCRT^{(n)}(\xi^0)$ as the total uncertainty at ξ^0, which can be represented in the form:

$$VCRT^{(n)}(\xi^0) = Volume\left(CR^{(n)}(\xi^0) \cup CR^{(n)}_{|Z(\xi)}(\xi^0)\right) \tag{21}$$

where $CR^{(n)}(\xi^0)$ and $CR^{(n)}_{|Z(\xi)}(\xi^0)$ represent the confidence regions of the parameters of model n at ξ^0, not considering and considering the data at ξ, respectively.

The final volumes of the confidence regions at $\xi^0+\xi$ are $VCR^{(n)}(\xi^0+\xi)$. In Fig. 1, the volume $VCR^{(n)}(\xi^0)$ is related to the number of elements of $S1$, the volume $VCR^{(n)}_{|Z(\xi)}(\xi^0)$ is related to the number of elements of $S2$, the volume $VCRT^{(n)}(\xi^0)$ is related to the number of elements of $S1 \cup S2$ and the volume $VCR^{(n)}(\xi^0+\xi)$ is related to the number of elements of $S3$. Although the number of elements may be infinite in the case of continuous variables, the ratio between the number of elements in $S3$ and the number of elements in $(S1 \cup S2)$ for model n corresponds to the ratio between the volumes $VCR^{(n)}(\xi^0+\xi)/VCRT^{(n)}(\xi^0)$.

Based on Postulate I (and as shown in Appendix A), it is possible to derive the following criterion for determination of the information gain after design ξ:

$$\psi^{\Theta}(\xi) = \sum_{n=1}^{M} w^{(n)}\left(1 - \frac{VCR^{(n)}(\xi^0+\xi)}{VCRT^{(n)}(\xi^0)}\right) \tag{22}$$

where $w^{(n)}$ are a weighting factors given to model n.

One must observe that Eq. (22) leads to conciliation of the two pursued objectives: model discrimination and precise parameters estimation. If one model is eliminated after performing ξ, the volume of the confidence region of the respective parameter estimates vanishes, meaning that model discrimination leads to high information gain. On the other hand, improvement of the precision of the parameter estimates of any of the analyzed models also leads to information gain. Besides, the information function presented in the Eq. (22) is very informative, as the information gain values lie between 0 and 1 and provide information about the relative amount of possible states of nature that was eliminated after performing ξ. For this reason, the use of Eq. (22) may be advantageous during the experimental design for model discrimination and estimation of precise model parameters.

2.1. Illustrative Example 2

Let us assume that the researcher is analyzing four models ($n=1,2,3,4$), each of them containing two parameters $\theta_1^{(n)}$ and $\theta_2^{(n)}$. After execution of the initial experimental design ξ^0, the volumes of the confidence regions of the parameter estimates $VCR^{(n)}(\xi^0)$ are presented in Fig. 3. After the design of the additional set of experiments ξ, the evaluation of the confidence regions with experimental data at ξ^0 but for parameter estimated at $\xi^0+\xi$ lead to the volume of confidence regions $VCR^{(n)}_{|Z(\xi)}(\xi^0)$ and $VCRT^{(n)}(\xi^0)$, as illustrated in Fig. 4. After re-estimation of model parameters and model probabilities, model 4 can be eliminated (P is lower than the minimum limit in Eq. (12)), so that the confidence regions of parameter estimates become as illustrated in Fig. 4C.

2.2. Weighting factors

$w^{(n)}$ weighs the importance of model n and can be defined by following distinct lines of thought. For instance, in terms of the plausible *states of nature*, $w^{(n)}$ can be written as

$$w^{(n)} = \frac{VCRT^{(n)}(\xi^0)}{\sum_{k=1}^{M} VCRT^{(k)}(\xi^0)} \tag{23a}$$

Alternatively, if similar importance is assigned to all models, then:

$$w^{(n)} = \frac{1}{M} \tag{23b}$$

One can also assume that $w^{(n)}$ should be related to the model probability $P^{(n)}$, as

$$w^{(n)} = P^{(n)} \tag{23c}$$

Other weighting factors can be adopted; however, for the sake of simplicity, the use the weighting factors presented in Eq. (23b) or (23c) can be recommended (Alberton, 2010).

2.3. Calculating the volumes of the confidence regions

The confidence regions of parameter estimates are generally assumed to be ellipsoidal, as illustrated in Figs. 3 and 4. The determination of $VCRT^{(n)}(\xi^0)$ may pose a very hard numerical task, as the ellipses are likely to present intersections (or the confidence regions are not ellipsoidal at all, as shown by Schwaab et al., 2008a), so that

$$VCRT^{(n)}(\xi^0) = VCR^{(n)}(\xi^0) + VCR^{(n)}_{|Z(\xi)}(\xi^0) - Volume\left(CR^{(n)}(\xi^0) \cap CR^{(n)}_{|Z(\xi)}(\xi^0)\right) \tag{24}$$

If there is no intersection between the confidence regions $CR^{(n)}(\xi^0)$ and $CR^{(n)}_{|Z(\xi)}(\xi^0)$, then the total volume is given as the sum of the individual volumes. On the other hand, if one region is contained by another, than the total volume of the confidence region is equal to the maximum value of $VCR^{(n)}(\xi^0)$ or $VCR^{(n)}_{|Z(\xi)}(\xi^0)$. Then, it is possible to write:

$$\max\left(VCR^{(n)}(\xi^0), VCR^{(n)}_{|Z(\xi)}(\xi^0)\right) \le VCRT^{(n)}(\xi^0) \le VCR^{(n)}(\xi^0) + VCR^{(n)}_{|Z(\xi)}(\xi^0) \tag{25}$$

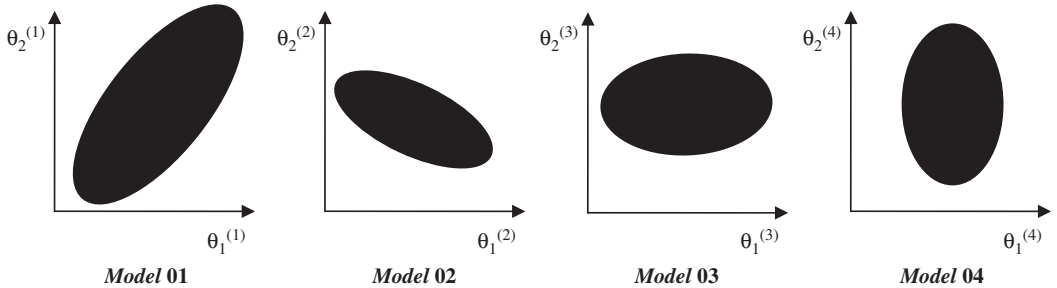

Fig. 3. Initial confidence regions of model parameters in the illustrative example.

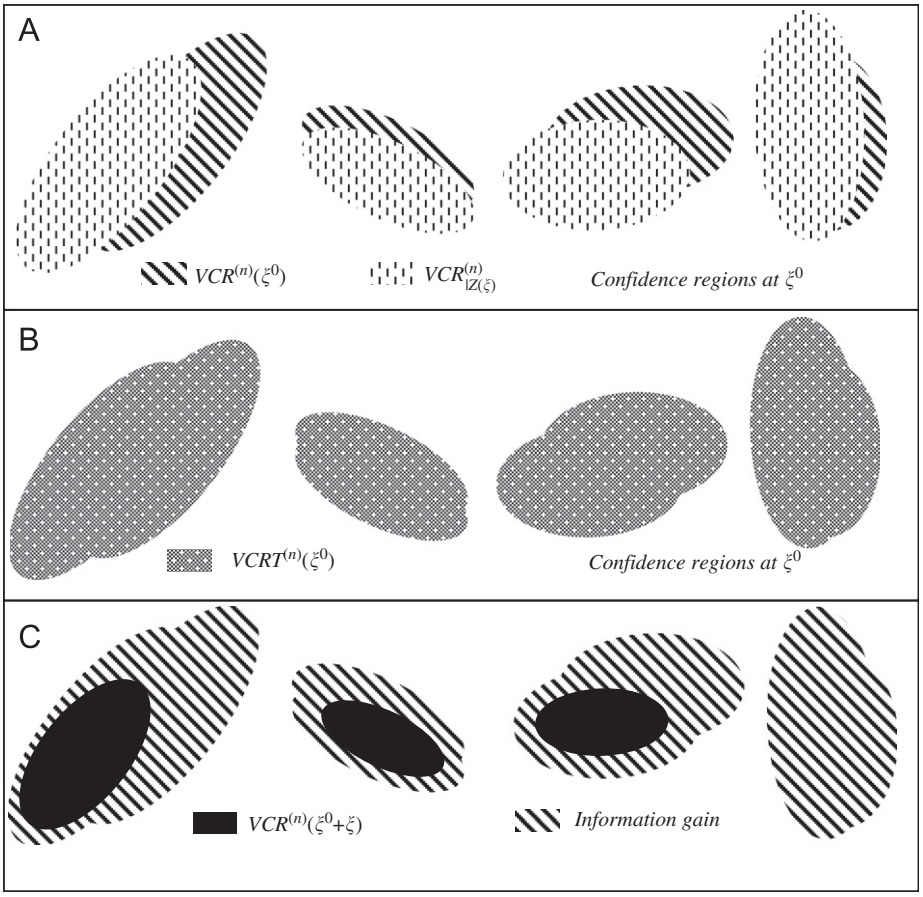

Fig. 4. Confidence regions of model parameters and information gains in the illustrative example, after performing the additional set of experiments.

If the parameter estimates do not change very significantly during the analysis, as usually assumed, it is reasonable to write:

$$VCRT^{(n)}(\xi^0) \approx \max\left(VCR^{(n)}(\xi^0), VCR^{(n)}_{|Z(\xi)}(\xi^0)\right) \tag{26}$$

Therefore,

$$\frac{VCR^{(n)}(\xi^0+\xi)}{VCRT^{(n)}(\xi^0)} = \frac{\sqrt{|V^{(n)}_{\Theta}(\xi^0+\xi)|}}{\max\left(\sqrt{|V^{(n)}_{\Theta}(\xi^0)|}, \sqrt{|V^{(n)}_{\Theta|Z(\xi)}(\xi^0)|}\right)} \tag{27}$$

and

$$\psi^{\Theta}(\xi) = \sum_{n=1}^{M} w^{(n)} \left(1 - \frac{\min\left(\sqrt{|J^{(n)}(\xi^0)|}, \sqrt{|J^{(n)}_{|Z(\xi)}(\xi^0)|}\right)}{\sqrt{|J^{(n)}(\xi^0+\xi)|}}\right) \tag{28}$$

Eqs. (27) and (28) seem more appropriate for implementation of numerical procedures.

2.4. Predicting experimental responses

In order to evaluate the volumes of confidence regions after performing ξ for a given mathematical model, it is necessary to know the experimental response and re-estimate the model parameters. In order to predict the experimental responses, one may rely on the available models or not. For instance, it seems reasonable to assume that the response variables for a given experimental design ξ can be anywhere inside the experimental range $\Delta Z(\xi)$. If one does not want to rely on model predictions, it is possible to maximize the information gain for the worst possible response value in $\Delta Z(\xi)$, as

$$\xi = \text{argmax}_{\xi} \min_{Z(\xi) \subset \Delta Z(\xi)} \left(\sum_{n=1}^{M} w^{(n)} \left(1 - \frac{VCR^{(n)}(\xi^0+\xi)}{VCRT^{(n)}(\xi^0)}\right)\right) \tag{29}$$

$$\text{Subject to}: \sum_{n=1}^{M} \varphi(\xi,n) < M$$

where $\varphi(\xi,n)$ is a binary function, which is equal to 1 if model n is eliminated and 0 otherwise. One should observe that the optimization problem proposed in Eq. (29) may become untreatable by numerical reasons.

It is certainly preferable to believe that experimental responses can be predicted by the models as usual criterions in the literature (Atkinson and Fedorov, 1975a, 1975b; Buzzi-Ferraris and Forzatti, 1983; Schwaab et al., 2006). However, one might also consider a multiobjective problem, where each model can be regarded as the true model for computation of experimental responses. As a consequence, the number of objective functions becomes equal to M and the set of objective functions OF can be given as

$$OF = \left\{\psi^{\Theta}_{|Z^{(1)}(\xi)}, \psi^{\Theta}_{|Z^{(2)}(\xi)}, \psi^{\Theta}_{|Z^{(3)}(\xi)}, \ldots, \psi^{\Theta}_{|Z^{(M-1)}(\xi)}, \psi^{\Theta}_{|Z^{(M)}(\xi)}\right\} \tag{30}$$

where the volumes of confidence regions at $\xi^0+\xi$ and at ξ^0 can be evaluated with outputs of model m as

$$VCR^{(n)}(\xi^0+\xi) = VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0+\xi) \tag{31a}$$

$$VCRT^{(n)}(\xi^0) = \max\left(VCR^{(n)}(\xi^0), VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)\right) \tag{31b}$$

As originally suggested by Atkinson and Fedorov (1975a, 1975b), the parameters of the rival models can be re-estimated when one of the models is assumed to be the true one. Since

From experimental data at ξ^0, estimate the parameters for all models $n=1,M$ and calculate the volume $VCR^{(n)}(\xi^0)=\left(\left|J^{(n)}(\xi^0)\right|\right)^{-0,5}$, where $\left|J^{(n)}(\xi^0)\right|$ is the determinant of the Fisher information matrix for model n
Specify the weight $w^{(n)}$ for all models $n=1,M$; for example, $w^{(n)}=1/M$

For all models m ($m=1,M$), do:

Generate experimental data with model m for experimental design ξ, considering these simulated data as experimental data
For all models n ($n=1,M$; $n\neq m$), do:

Re-estimate the parameters of the model n and evaluate the model adequacy with a statistic test (chi-square, for example)

If model n became inadequate, then:

$$VCR^{(n)}(\xi^0+\xi)=0$$
$$VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)=0$$

Else

$$VCR^{(n)}(\xi^0+\xi)=\left(\left|J^{(n)}_{|Z^{(m)}(\xi)}(\xi^0+\xi)\right|\right)^{-0,5}$$, where $J^{(n)}_{|Z^{(m)}(\xi)}(\xi^0+\xi)$ is the Fisher information matrix for model n calculated with experimental data at ξ^0 and data at ξ simulated with model m

$$VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)=\left(\left|J^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)\right|\right)^{-0,5}$$, where $J^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)$ is the Fisher information matrix for the model n calculated with data at ξ^0, but with parameters re-estimated considering the data $Z^{(m)}(\xi)$

End if

If ($VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)>VCR^{(n)}(\xi^0)$), then

$$VCRT^{(n)}(\xi^0)=VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)$$

Else

$$VCRT^{(n)}(\xi^0)=VCR^{(n)}(\xi^0)$$

End IF

End do

Calculate $\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi)$ with Equation (27)

Return model probabilities, experimental data and parameter values for all models n ($n=1,M$) to their original values

End do

Fig. 5. Algorithm used for computation of $\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi)$ for all models m ($m=1,M$) for an experimental design candidate ξ.

several objectives can be formulated, the number of possible solutions can be infinite and the selection of a particular experimental design in the Pareto front must follow some multiobjective decision criterion, such as the Maximin, Bayesian or weighted sum of criteria, as (Alberton et al., submitted for publication; Chernoff and Moses, 1959; Donckels et al., 2010)

$$\xi_{otm}=\text{argmax}_{\xi}\text{min}_{m}\left[\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi)\right] \tag{32a}$$

$$\xi_{otm}=\text{argmax}_{\xi}\sum_{m=1}^{M}P^{(m)}\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi) \tag{32b}$$

$$\xi_{otm}=\text{argmax}_{\xi}\frac{1}{M}\sum_{m=1}^{M}\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi) \tag{32c}$$

3. Numerical procedure

The algorithm used for evaluation of the experimental design criteria at ξ is presented in Fig. 5. For simplicity and if necessary, the re-estimation step can be skipped and parameters can be assumed to be the same, while the covariance matrix of model parameters can be updated with the help of Eq. (13). In this case, the adequacy of the model fit can also be updated in the form:

$$F^{(n)}_{|Z^{(m)}(\xi)}=F^{(n)}+\left(Z^{(n)}(\xi)-Z^{(m)}(\xi)\right)^{T}V_{Z}(\xi)^{-1}\left(Z^{(n)}(\xi)-Z^{(m)}(\xi)\right) \tag{33}$$

where $F^{(n)}_{|Z^{(m)}(\xi)}$ is the objective function for model n calculated with experimental data at ξ^0 and data at ξ simulated with model m, $F^{(n)}$ is the objective function for model n calculated with data at ξ^0, $V_Z(\xi)$ is the covariance matrix of experimental uncertainties at ξ, $Z^{(n)}(\xi)$ is the response of models n at ξ and $Z^{(m)}(\xi)$ is the response of models m at ξ. With this simplification, the algorithm becomes simpler and more robust, as iterative model computations required by the parameter estimation procedure are avoided. Despite that, it becomes impossible to evaluate how flexible the proposed models are to accommodate new experimental data and the evaluation becomes certainly poorer (Alberton et al., submitted for publication).

In order to obtain the optimal experimental design, an optimization algorithm must be employed. In this work, optimization was performed through grid search. This choice does not affect any of the conclusions presented here, as other alternative techniques might certainly be used to perform the optimization task proposed here. Besides, in most practical experimental design problems, the candidate experiments are indeed presented in the discrete form, as a set of discrete experimental conditions that can be performed at the experimental plant.

4. Examples

Two numerical examples are presented below to illustrate the applicability of the new proposed procedure. In both examples, preliminary experiments were carried out to allow for initial estimation of model parameters. In both examples, the confidence level was considered to be equal to 95%. Models probabilities were calculated as recommended by Schwaab et al. (2006), according to Eq. (11).

Parameter estimation was performed with a hybrid numerical procedure, as described by Schwaab et al. (2008a, 2008b). Initially, parameter search was carried out with the Particle Swarm Optmization (PSO) procedure (Kennedy and Eberhart, 1995); afterwards, the best parameter values were used as initial guesses in a Gauss–Newton procedure (Schwaab et al., 2010), as described in the classical literature (Bard, 1974).

For selection of the optimal design, a simple grid search technique was implemented. In this case, the experimental range was divided into a finite number of distinct characteristic experimental conditions (discretization of the original experimental region). Afterwards, the procedure described in Fig. 5 was applied to each individual experimental condition. Re-estimation of model parameters was performed as described previously (Schwaab and Pinto, 2007; Schwaab et al., 2010). The obtained criterion values, obtained when each of the rival models was assumed to be the correct one, were stored for all experimental conditions. Finally, the Maximin, Bayesian and Equal Model Weights criteria were computed, as presented in Eq. (32). All numerical procedures were implemented in Fortran.

4.1. Example 1—a simple linear problem

The main objective of Example 1 is providing a simple test case, which allows for simple computations and easy understanding of the obtained results. Let us consider two models:

Model 01: $y=\theta^{(1)}x$
Model 02: $y=\theta^{(2)}x^{1,5}$

The experimental variance is assumed to be constant and equal to $\sigma_y^2=4\times10^{-4}$. The true model used for generation of pseudo-experimental data was the second model, with $\theta^{(2)}=1$. Three initial experiments were simulated with model 2, as shown in Table 1. The experimental data were corrupted with a random noise signal generated in accordance with a normal distribution, with zero mean and variance σ_y^2.

The sets of model parameters contain a single parameter for each model: $\Theta^{(1)}=\{\theta^{(1)}\}$ and $\Theta^{(2)}=\{\theta^{(2)}\}$. Parameters $\theta^{(1)}$ and $\theta^{(2)}$ can be estimated, as

$$\theta^{(1)est}=\frac{\sum_{i=1}^{3}y_i^{exp}x_i}{\sum_{i=1}^{3}x_i^2}=0.6676 \tag{34a}$$

$$VCR^{(1)}(\xi^0)=\sqrt{\sigma^2_{\theta^{(1)}}}=\sqrt{\frac{\sigma_y^2}{\sum_{i=1}^{3}x_i^2}}=\sqrt{0.00133} \tag{34b}$$

Table 1
Initial experiments for Example 1.

x	y^{exp}
0.1	0.0405
0.2	0.1010
0.5	0.3520

$$\theta^{(2)est}=\frac{\sum_{i=1}^{3}y_i^{exp}x_i^{1,5}}{\sum_{i=1}^{3}x_i^3}=1.0059 \tag{34c}$$

$$VCR^{(2)}(\xi^0)=\sqrt{\sigma^2_{\theta^{(2)}}}=\sqrt{\frac{\sigma_y^2}{\sum_{i=1}^{3}x_i^3}}=\sqrt{0.00298} \tag{34d}$$

Final objective functions for models 1 and 2 are $F^{(1)}=5.1738$ and $F^{(2)}=0.5341$. The upper limit of the chi-square function with two degrees of freedom and confidence level of 97.5% is equal to $F_{\chi2}=7.3777$. Therefore, both objective functions can be regarded as good model candidates. The initial experimental design contains three experiments $\xi^0=\{x_1, x_2, x_3\}$ and a fourth experiment $\xi=\{x_4\}$ must be performed to improve the parameter quality and allow for model discrimination. Given the previous results, both models should be regarded as true models. Taking model 1 as the true model, the parameters values $\theta^{(1)}$ and $\theta^{(2)}$ and their respective variances can be calculated as

$$\theta^{(1)est}{}_{|Z^{(1)}(\xi)}=\theta^{(1)est} \tag{35a}$$

$$VCR^{(1)}_{|Z^{(1)}(\xi)}(\xi^0+\xi)=\sqrt{\sigma^2_{\theta^{(1)}|Z^{(1)}(\xi)}}=\sqrt{\frac{\sigma_y^2}{x_4^2+\sum_{i=1}^{3}x_i^2}} \tag{35b}$$

$$\theta^{(2)est}_{|Z^{(1)}(\xi)}=\frac{\overbrace{\left(\theta_1{}^{est}x_4\right)}^{y_4{}^{exp}}x_4^{1,5}+\sum_{i=1}^{3}y_i^{exp}x_i^{1,5}}{x_4^3+\sum_{i=1}^{3}x_i^3} \tag{35c}$$

$$VCR^{(2)}_{|Z^{(1)}(\xi)}(\xi^0+\xi)=\sqrt{\sigma^2_{\theta^{(2)}|Z^{(1)}(\xi)}}=\sqrt{\frac{\sigma_y^2}{x_4{}^3+\sum_{i=1}^{3}x_i^3}} \tag{35d}$$

Calculation of the total uncertainty at ξ^0 requires the calculation of the determinant at ξ^0 with parameters estimated at $\xi^0+\xi$. As the analyzed models are linear with respect to model parameters, the parameter uncertainty does not depend on the parameters values. Thus, in this case one can observe that the uncertainty at ξ^0 with parameters estimated at $\xi^0+\xi$ is equal to the uncertainty at ξ^0 with parameters estimated at ξ^0. Therefore:

$$VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)=VCR^{(n)}(\xi^0) \tag{36}$$

$$VCRT^{(n)}(\xi^0)=\max\left(VCR^{(n)}(\xi^0),VCR^{(n)}_{|Z^{(m)}(\xi)}(\xi^0)\right)=VCR^{(n)}(\xi^0) \tag{37}$$

The experimental result obtained at x_4 can lead to elimination of model 2 if

$$\varphi_2=\begin{cases}1 & \text{if } F^{(2)}>F_{\chi2}\\ 0 & \text{if } F^{(2)}\le F_{\chi2}\end{cases} \tag{38}$$

where $F_{\chi2}=9.3484$. Then,

$$\psi^{\Theta}_{|Z^{(1)}(\xi)}(\xi)=w^{(1)}\left(1-\sqrt{\frac{\sigma^2_{\theta^{(1)}|Z_1(\xi)}}{\sigma^2_{\theta^{(1)}}}}\right)+w^{(2)}\left(1-\sqrt{\frac{\sigma^2_{\theta^{(2)}|Z_1(\xi)}}{\sigma^2_{\theta^{(2)}}}}(1-\varphi_2)\right) \tag{39}$$

If $w^{(1)}$ and $w^{(2)}$ are assumed to be equal:

$$\psi^{\Theta}_{|Z^{(1)}(\xi)}(\xi)=\frac{1}{2}\left(1-\sqrt{\frac{\sigma^2_{\theta^{(1)}|Z_1(\xi)}}{\sigma^2_{\theta^{(1)}}}}\right)+\frac{1}{2}\left(1-\sqrt{\frac{\sigma^2_{\theta^{(2)}|Z_1(\xi)}}{\sigma^2_{\theta^{(2)}}}}(1-\varphi_2)\right) \tag{40}$$

which can be promptly computed with standard spreadsheets.

Similar computations can be performed, by assuming that model 2 is the correct one, leading to

$$\theta^{(2)est}_{|Z^{(2)}(\xi)} = \frac{\overbrace{\left(\theta_2^{est} x_4^{1,5}\right)}^{y_4^{exp}} x_4 + \sum_{i=1}^{3} y_i^{exp} x_i}{x_4^2 + \sum_{i=1}^{3} x_i^2} \tag{41a}$$

$$VCR^{(1)}_{|Z^{(2)}(\xi)}(\xi^0+\xi) = \sqrt{\sigma^2_{\theta^{(1)}|Z^{(2)}(\xi)}} = \sqrt{\frac{\sigma_y^2}{x_4{}^2 + \sum_{i=1}^{3} x_i^2}} \tag{41b}$$

$$\theta^{(2)est}_{|Z^{(2)}(\xi)} = \theta_2^{est} \tag{41c}$$

$$VCR^{(2)}_{|Z^{(2)}(\xi)}(\xi^0+\xi) = \sqrt{\sigma^2_{\theta^{(2)}|Z^{(2)}(\xi)}} = \sqrt{\frac{\sigma_y^2}{x_4^3 + \sum_{i=1}^{3} x_i^3}} \tag{41d}$$

$$\varphi_1 = \begin{cases} 1 & se\ F^{(1)} > F_{\chi 2} \\ 0 & se\ F^{(1)} \le F_{\chi 2} \end{cases} \tag{42}$$

$$\psi^{\Theta}_{|Z^{(2)}(\xi)}(\xi) = \frac{1}{2}\left(1 - \sqrt{\frac{\sigma^2_{\theta^{(1)}|Z^{(2)}(\xi)}}{\sigma^2_{\theta^{(1)}}}}(1-\varphi_1)\right) + \frac{1}{2}\left(1 - \sqrt{\frac{\sigma^2_{\theta^{(2)}|Z^{(2)}(\xi)}}{\sigma^2_{\theta^{(2)}}}}\right) \tag{43}$$

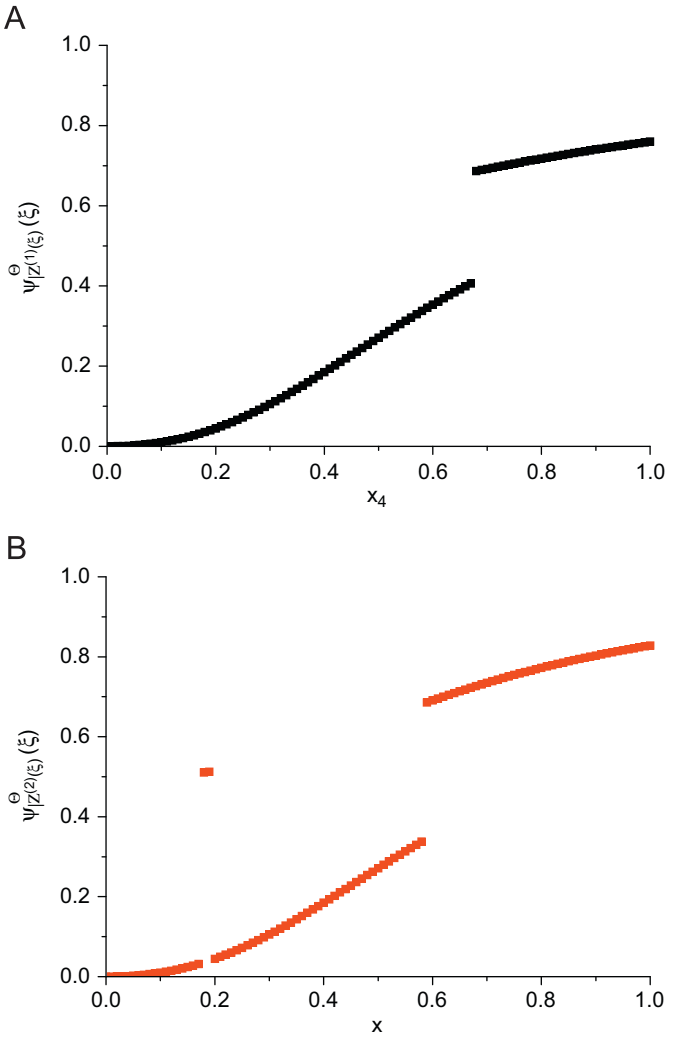

Fig. 6. Experimental design criteria when models (A) 1 and (B) 2 are considered as the true *states of nature* in Example 1.

The values of the analyzed experimental design criteria are presented in Fig. 6. Considering model 1 as the true model, one should not expect to eliminate model 2 when $x_4 < 0.68$. Nevertheless, for $x_4 \ge 0.68$ model 2 can be discriminated from model 1, which is indicated by the discontinuity in the value of the information criterion in Fig. 6A. Considering model 2 as the true one, a narrow discriminating region is observed for x_4 near to 0.2, as indicated by the discontinuity in Fig. 6B. However, despite the existence of this narrow region, model 1 is sufficiently flexible to accommodate the additional experimental point when $x_4 < 0.59$. For $x_4 \ge 0.59$, model 1 is expected to be discriminated from model 2, as indicated by the discontinuity in Fig. 6B. In this simple problem, the optimum experimental design is $x_4 = 1.00$ where the information gain is maximum for both models 1 and 2. Considering model 1 as the true model, 76% of the original uncertainty can be eliminated at $x_4 = 1.00$; if model 2 is the *true state of nature*, 83% of the original uncertainty can be eliminated at $x_4 = 1.00$. In both cases, one can expect to discriminate model 1 from model 2.

4.2. Example 2—a kinetic problem

This problem has been studied previously (Buzzi-Ferraris et al., 1984; Schwaab et al., 2006; Alberton et al., submitted for publication) and constitutes a more complex benchmark for comparative analyses, when multiple-input multiple-output models are considered. In the proposed kinetic problem, four rival models are proposed to describe the behavior of a system containing two response variables, two design variables and four parameters. The models are presented below:

$$\text{Model 1: } y_1 = \frac{\theta_1^{(1)} x_1 x_2}{1 + \theta_3^{(1)} x_1 + \theta_4^{(1)} x_2}, \quad y_2 = \frac{\theta_2^{(1)} x_1 x_2}{1 + \theta_3^{(1)} x_1 + \theta_4^{(1)} x_2}$$

$$\text{Model 2: } y_1 = \frac{\theta_1^{(2)} x_1 x_2}{\left(1 + \theta_3^{(2)} x_1 + \theta_4^{(2)} x_2\right)^2}, \quad y_2 = \frac{\theta_2^{(2)} x_1 x_2}{\left(1 + \theta_3^{(2)} x_1\right)^2}$$

$$\text{Model 3: } y_1 = \frac{\theta_1^{(3)} x_1 x_2}{\left(1 + \theta_4^{(3)} x_2\right)^2}, \quad y_2 = \frac{\theta_2^{(3)} x_1 x_2}{\left(1 + \theta_3^{(3)} x_1\right)^2}$$

$$\text{Model 4: } y_1 = \frac{\theta_1^{(4)} x_1 x_2}{1 + \theta_3^{(4)} x_1 + \theta_4^{(4)} x_2}, \quad y_2 = \frac{\theta_2^{(4)} x_1 x_2}{1 + \theta_3^{(4)} x_1}$$

Each model presents four parameters: $\Theta^{(1)} = \{\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \theta_4^{(1)}\}$; $\Theta^{(2)} = \{\theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)}, \theta_4^{(2)}\}$; $\Theta^{(3)} = \{\theta_1^{(3)}, \theta_2^{(3)}, \theta_3^{(3)}, \theta_4^{(3)}\}$; and $\Theta^{(4)} = \{\theta_1^{(4)}, \theta_2^{(4)}, \theta_3^{(4)}, \theta_4^{(4)}\}$. In this case, it is assumed that model 1 is the true model, with parameter values equal to: $\theta_1^{(1)} = 0.1$; $\theta_2^{(1)} = 0.01$; $\theta_3^{(1)} = 0.1$; and $\theta_4^{(1)} = 0.01$. Experimental variances were considered constant and equal to 0.35 and 2.3×10^{-2} for variables y_1 and y_2, respectively. Five initial experiments were generated with the true model, as presented in Table 2. The estimation of the model parameters was performed for the four models with the preliminary experiments presented in Table 2, as shown in Table 3.

Model 3 if found to be inadequate after the preliminary experiments it is rejected. Models 1, 2 and 4 are still valid after the initial preliminary experiments, justifying the implementation of the discrimination procedure. The grid search is applied in the experimental ranges $5.0 \le x_1 \le 55.0$ and $5.0 \le x_2 \le 55.0$. In this example, 20 equally spaced nodal points were defined for each variable, resulting in 400 possible trials inside the experimental range. Values of the experimental design criteria and of the number of discriminated models are presented in Fig. 7 for all analyzed experimental conditions. Discontinuities correspond to experimental conditions where the number of discriminated

models changes. According to the Maxmin criterion, the next experiment should be performed at $x_1=12.9$ and $x_2=55.0$. At this point, the number of discriminated models is not maximum, but the precise estimation of model parameters leads to elimination of almost 90% of the possible *states of nature*. This indicates the high sensitivity of the confidence regions to variations of the experimental conditions. When the experiment is performed, model 4 becomes inadequate and models 1 and 2 are still probable. The procedure can be repeated again, as indicated in Table 4, leading to selection of the true model 1 and improvement of the parameter estimates. Compared to the results presented by Schwaab et al. (2006) and Alberton et al. (submitted for publication), one additional experiment was required for discrimination of the true model; however, model parameters were obtained with higher precision. Therefore, from this point of view, the overall performance can be regarded as better in the present case, as one might already expect because of the conciliation of the distinct experimental criteria in terms of the information content of the problem.

Table 2
Initial experiments in Example 2.

Run	x_1	x_2	y_1	y_2
01	20.0	20.0	13.443	1.299
02	30.0	20.0	13.817	1.433
03	20.0	30.0	17.809	1.885
04	30.0	30.0	21.139	2.118
05	25.0	25.0	16.039	1.635

Table 3
Parameters estimates and adequacy of model fits in Example 2.

Model	$\theta_1^{(m)}$	$\theta_2^{(m)}$	$\theta_3^{(m)}$	$\theta_4^{(m)}$	Model is good?
01	0.1311	0.0134	0.1431	0.0145	Yes
02	0.0743	0.0068	0.0233	0.0034	Yes
03	0.0281	0.0067	0.0017	0.0232	**No**
04	0.1162	0.0107	0.1187	0.0162	Yes

Fig. 8 presents the information gain, $\psi^{\Theta}(\xi)$, and the logarithm of the Fisher information matrix for model 1 as functions of the additional experiments that are carried out. The information gain is very high in the initial experiments, as uncertainties of model parameters can be reduced significantly, and afterwards approaches the value of 20%. At the end, as observed with the logarithm of the Fisher information matrix, the parameter uncertainties do not change very significantly and the analyst might wonder whether the experimental plan should be interrupted. It is interesting to observe how the proposed design procedure conciliates the model discrimination and the precise parameter estimation procedures: if model parameters are known with low precision, it may be important to estimate the parameters more precisely in order to guarantee more accurate model predictions, as observed by Donckels et al. (2010); on the other hand, model discrimination leads to high information gain and is convenient for reduction of the design criterion values.

5. Conclusions

In order to conciliate the objectives of experimental design procedures for both model discrimination and estimation of model parameters, an experimental design criterion based on measures of the information gain was proposed. The proposed criterion depends on the volumes of the confidence regions of the model parameters and presents a number of advantageous

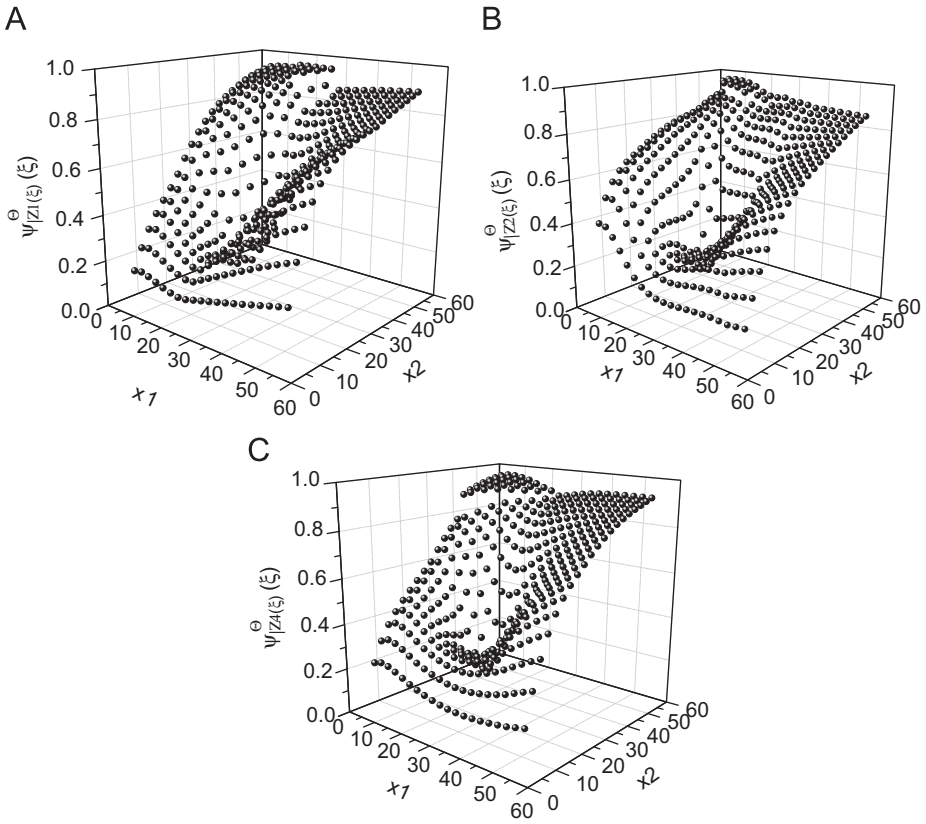

Fig. 7. Values of the experimental design criteria (A) $\psi^{\Theta}_{|Z^{(1)}(\xi)}(\xi)$, (B) $\psi^{\Theta}_{|Z^{(2)}(\xi)}(\xi)$, and (C) $\psi^{\Theta}_{|Z^{(4)}(\xi)}(\xi)$ when models $m=1,2,4$ are the true models in Example 2.

Table 4
Design criteria, designed experimental conditions, parameter estimates and adequacy of model fits for three additional sequential experiments in Example 2.

Design criterion (*Maximin*)	x_1	x_2	$\psi^{\Theta}_{\|Z^{(1)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(2)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(4)}(\xi)}(\xi)$	$min_m(\psi^{\Theta}_{\|Z^{(m)}(\xi)}(\xi))$
	12.9	55.0	0.95	0.97	0.96	0.95
Experiment	*Run*	x_1	x_2	y_1	y_2	$\psi^{\Theta}(\xi)$
	06	12.9	55.0	24.859	2.488	0.93
Parameter estimates	*Model*	$\theta_1^{(m)}$	$\theta_2^{(m)}$	$\theta_3^{(m)}$	$\theta_4^{(m)}$	*Model is good?*
	01	0.1413	0.0143	0.1523	0.0199	Yes
	02	0.0537	0.0053	0.0164	0.0005	Yes
	03	–	–	–	–	**No**
	04	0.0592	0.0059	0.0486	0.0011	**No**
Design criterion (*Maximin*)	x_1	x_2	$\psi^{\Theta}_{\|Z^{(1)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(2)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(4)}(\xi)}(\xi)$	$min_m(\psi^{\Theta}_{\|Z^{(m)}(\xi)}(\xi))$
	55.0	55.0	0.92	0.90	–	0.92
Experiment	*Run*	x_1	x_2	y_1	y_2	$\psi^{\Theta}(\xi)$
	07	55.0	55.0	43.523	4.341	0.89
Parameter estimates	*Model*	$\theta_1^{(m)}$	$\theta_2^{(m)}$	$\theta_3^{(m)}$	$\theta_4^{(m)}$	*Model is good?*
	01	0.1074	0.0108	0.1084	0.0122	Yes
	02	0.0561	0.0054	0.0173	0.0007	Yes
	03	–	–	–	–	**No**
	04	–	–	–	–	**No**
Design criterion (*Maximin*)	x_1	x_2	$\psi^{\Theta}_{\|Z^{(1)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(2)}(\xi)}(\xi)$	$\psi^{\Theta}_{\|Z^{(4)}(\xi)}(\xi)$	$min_m(\psi^{\Theta}_{\|Z^{(m)}(\xi)}(\xi))$
	39.2	55.0	0.67	0.68	0.0	0.67
Experiment	*Run*	x_1	x_2	y_1	y_2	$\psi^{\Theta}(\xi)$
	08	39.2	55.0	39.970	3.987	0.74
Parameter estimates	*Model*	$\theta_1^{(m)}$	$\theta_2^{(m)}$	$\theta_3^{(m)}$	$\theta_4^{(m)}$	*Model is good?*
	01	0.1033	0.0104	0.1033	0.0109	Yes
	02	0.0570	0.0054	0.0176	0.0010	**No**
	03	–	–	–	–	**No**
	04	–	–	–	–	**No**

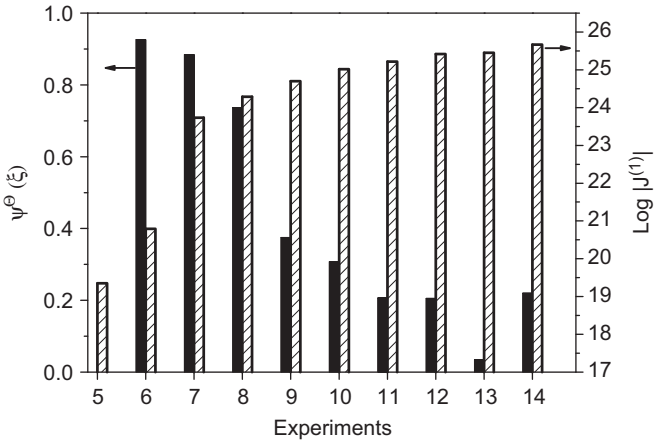

Fig. 8. Information gain $\psi^{\Theta}(\xi)$ and $\log|J^{(1)}|$ as additional experiments are carried out in Example 2.

aspects, such as the conciliation of the usual experimental design objectives and the fact that the obtained criterion values can be easily interpreted in terms of the information eliminated after carrying out additional experiments. Besides, the proposed design criterion can easily accommodate multiobjective experimental design approaches, as shown in the examples. Compared to results presented previously in the literature, it can be said that the overall performance obtained with the present design criterion was better, as similar number of experiments were performed to discriminate among candidate models, although the final obtained parameter estimates were more precise because of the conciliation of the distinct experimental criteria in terms of the information content of the problem.

Nomenclature

$CR^{(n)}(\xi^0)$ confidence region at ξ^0
$CR^{(n)}_{|Z(\xi)}(\xi^0)$ confidence region at ξ^0, after performing ξ
D and D_{max} discriminant function and maximum value of the discriminant function
DKL Kullback–Leibler divergence criterion
$eff_D(\xi,\Theta^{(n)})$ and $eff_E(\xi,\Theta^{(n)})$ efficiency of discrimination and precise parameter estimation, when the parameters values are $\Theta^{(n)}$ and experiments in ξ are performed
F likelihood function
H entropy
$I(\xi)$ Lindley information of an experimental design ξ
$J^{(n)}$ Fisher information matrix for model n
OF set of objective functions
p_d probability of the state of nature $\Theta_d^{(n)}$
$P^{(n)}$ probability of model n
$S1$, $S1$, $S3$ sets of possible parameter values

$VCR^{(n)}(\xi^0)$ volume of the confidence region of parameter uncertainties for model n, calculated with data obtained at ξ^0

$VCR^{(n)}_{|Z(\xi)}(\xi^0)$ volume of the confidence region of parameter uncertainties for model n, calculated with data obtained at ξ^0 with parameters obtained at $\xi^0+\xi$

$VCRT^{(n)}$ total volume of the confidence region ξ^0

$VCR^{(n)}(\xi^0+\xi)$ volume of the confidence region of parameter uncertainties for model n, calculated with data obtained at $\xi^0+\xi$

V_Z covariance matrix of experimental uncertainties

V_Θ covariance matrix of parameter estimates

w weighting values

Z^{exp}, Z^{calc} experimental and calculated variables

Greek symbols

ξ^0 initial set of experiments

ξ additional set of experiments

θ model parameter

Θ set of model parameters

Θ_d discrete variables

$\Theta_d^{(n)}$ possible states of nature of discrete variable Θ_d

$\varphi(\xi,n)$ or $\varphi(\xi,\Theta_i^{(G)})$ binary function, equal to 1 if model n or possible parameter values $\Theta_i^{(G)}$ are eliminated after performing experiments in ξ and equal to 0 otherwise

$\pi(\Theta)$ probability distribution function of the parameters Θ

$\psi^{\Theta}(\xi)$ information gain at ξ

$\psi^{\Theta}_{|Z^{(m)}(\xi)}(\xi)$ information gain at ξ given model predictions obtained with model m

$\chi^2_{N-Np,\alpha}$ chi-square function with N–Np degrees of freedom and confidence interval α

Acknowledgments

The authors thank CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico – and FAPERJ—Fundação Carlos Chagas Filho de Apoio à Pesquisa do Estado do Rio de Janeiro – for supporting this research and providing scholarships.

Appendix A. Information gain derived from Postulate I

Let us assume that M rival models are considered, that some experimental data Z^{exp} from experimental design ξ^0 are available and that model parameters can be estimated. Therefore, it is possible to evaluate the confidence regions for all model parameters of all analyzed rival models. $\Theta^{(n)}$ represents all possible parameter values for model n and $\Theta^{(G)}$ represents the total set of parameter values for all models, as illustrated in Fig. A1.

According to Postulate I, the following function can be proposed as a measure of the information gain when the experiments in ξ are performed:

$$\psi^{\Theta}(\xi)=\sum_i w^{(G)}(\Theta_i^{(G)})\frac{\varphi\left(\xi,\Theta_i^{(G)}\right)}{N\Theta^{(G)}} \tag{A.1}$$

where $w^{(G)}(\Theta_i^{(G)})$ is a weighting value given to element i of $\Theta^{(G)}$, $N\Theta^{(G)}$ represents the number of elements in $\Theta^{(G)}$ and $\varphi(\xi,\Theta_i^{(G)})$ is a binary function, which assumes the value 1 when the element i is eliminated after performing ξ and 0 otherwise. Assuming that all parameters that belong to the same model structure can be weighted equally:

$$\psi^{\Theta}(\xi)=\sum_{n=1}^{M}\left(w^{(G,n)}\frac{\sum_{j=1}^{N\Theta^{(n)}}\varphi\left(\xi,\Theta_j^{(n)}\right)}{N\Theta^{(G)}}\right) \tag{A.2}$$

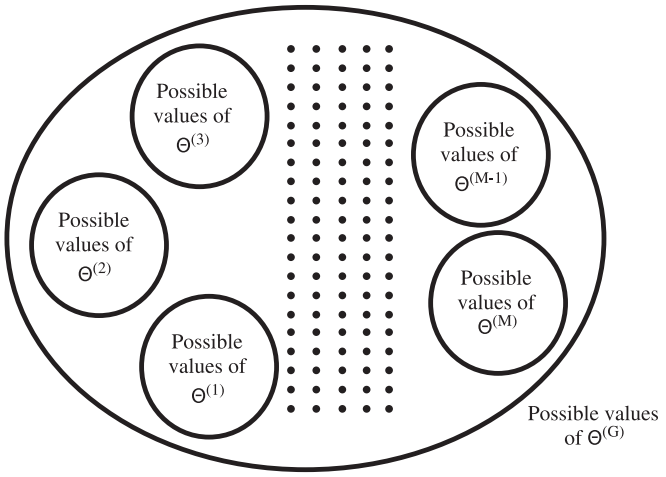

Fig. A1. Possible parameter values $\Theta^{(n)}$ for all models n (n=1,M).

where $w^{(G,n)}$ is the weight given to parameters of model n and $N\Theta^{(n)}$ represents the number of possible parameter values for model n. A simple weighting value for model n can be defined as

$$w^{(G,n)}=\frac{w^{(n)}}{\left(N\Theta^{(n)}/N^{(G)}\right)} \tag{A.3}$$

where $w^{(n)}$ weights the relative importance of model n and $(N\Theta^{(n)}/N\Theta^{(G)})$ weights the precision of model n, as $N\Theta^{(n)}$ is smaller for more accurate model candidates. In this case:

$$\psi^{\Theta}(\xi)=\sum_{n=1}^{M}w^{(n)}\left(\frac{\sum_j\varphi\left(\xi,\Theta_j^{(n)}\right)}{N\Theta^{(n)}}\right) \tag{A.4}$$

The term $\sum_j\varphi(\xi,\Theta_j^{(n)})/N\Theta^{(n)}$ represents the percentage of possible parameter values for model n that are eliminated after performing experiments at ξ. Therefore, the information gain can be written as

$$\psi^{\Theta}(\xi)=\sum_{n=1}^{M}w^{(n)}\left(\frac{VCRT^{(n)}(\xi^0)-VCR^{(n)}\left(\xi^0+\xi\right)}{VCRT^{(n)}(\xi^0)}\right) \tag{A.5}$$

$$\psi^{\Theta}(\xi)=\sum_{n=1}^{M}w^{(n)}\left(1-\frac{VCR^{(n)}(\xi^0+\xi)}{VCRT^{(n)}(\xi^0)}\right) \tag{A.6}$$

References

Alberton, 2010. Estimação de parâmetros e planejamento de experimentos: estudo de incertezas e funções de informação. D.Sc. Thesis, Universidade Federal do Rio de Janeiro, Brasil (in Portuguese).

Alberton, A.L., Schwaab, M., Biscaia, E.C., Pinto, J.C., 2010. Sequential experimental design through a multiobjective optimization approach. Chemical Engineering Science 65, 5482–5494.

Alberton, A.L., Schwaab, M., Lobão, M.W., Pinto, J.C., Design of experiments for discrimination of rival models based on multiobjective optimization procedures. Chemical Engineering Science, submitted for publication.

Asprey, S.P., Macchietto, S., 2002. Designing robust optimal dynamic experiments. Journal of Process Control 12, 545–556.

Atkinson, A.C., Donev, A.N., 1992. Optimal Experimental Design. Clarendon Press, Oxford.

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs, with SAS. Clarendon Press, Oxford.

Atkinson, A.C., 2008. DT-optimum designs for model discrimination and parameter estimation. Journal of Statistical Planning and Inference 138, 56–64.

Atkinson, A.C., Fedorov, V.V., 1975a. The design of experiments for discriminating between two rival models. Biometrika 62, 57–70.

Atkinson, A.C., Fedorov, V.V., 1975b. Optimal design: experiments for discriminating between several models. Biometrika 62, 289–303.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York.

Biswas, A., Chaudhuri, P., 2002. An efficient design for model discrimination and precise parameters estimation in linear models. Biometrika 89, 709–718.

Box, G.E.P., Hill, W.J., 1967. Discrimination among mechanistic models. Technometrics 9, 57–71.
Buzzi-Ferraris, G., Forzatti, P., 1983. A new sequential experimental design procedure for discriminating among rival models. Chemical Engineering Science 38, 225–232.
Buzzi-Ferraris, G., Manenti, F., 2009. Kinetic models analysis. Chemical Engineering Science 64, 1061–1074.
Buzzi-Ferraris, G., Forzatti, P., Emig, G., Hofmann, H., 1984. Sequential experimental design for model discriminating in the case of multiresponse models. Chemical Engineering Science 39, 81–85.
Buzzi-Ferraris, G., Forzatti, P., Canu, P., 1990. An improved version of a sequential design criterion for discriminating among rival multiresponse models. Chemical Engineering Science 45, 477–481.
Chernoff, H., 1972. Sequential Analysis and Optimal Design. SIAM, Philadelphia.
Chernoff, H., Moses, L.E., 1959. Elementary Decision Theory. John and Wiley Sons, New York.
Clyde, M., Chaloner, K., 1996. The equivalence of weighted designs in multiobjective design problems. Journal of American Statistical Association 91, 1236–1244.
Cook, R.D., Wong, W.K., 1994. On the equivalence of constrained and compounded optimal design. Journal of American Statistical Association 89, 687–692.
Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.
DeGroot, M.H., 1962. Uncertainty, information and sequential experiments. Annals of Mathematical Statistics 33, 404–419.
Dette, H., 1990. A generalization of D- and D1-optimal designs in polynomial regression. Annals of Statistics 18, 1784–1805.
Dette, H., 1992. Optimal designs for a class of polynomials of odd or even degree. Annals of Statistics 20, 238–259.
Dette, H., 1993. On a mixture of the D- and D1-optimality criterion in polynomial regression. Journal of Statistical Planning and Inference 35, 233–249.
Dette, H., 1994. Discrimination designs for polynomial regression on compact intervals. Annals of Statistics 22, 890–904.
Dette, H., 1997. Designing experiments with respect to ''standardized'' optimality criteria. Journal of the Royal Statistical Society B 59, 97–110.
Dette, H., O'Brien, T.E., 1999. Optimality criteria for regression models based on predicted variance. Biometrika 86, 93–106.
Dette, H., Studden, W.J., 1995. Optimal designs for polynomial regression when the degree is not known. Statistica Sinica 5, 459–473.
Dette, H., Melas, V.B., Pepelyshev, A., Strigul, N., 2005. Robust and efficient design of experiments for the Monod model. Journal of Theoretical Biology 234, 537–550.
Dette, H., Haines, L.M., Imhof, L.A., 2007. Maximin and Bayesian optimal designs for regression models. Statistica Sinica 17, 463–480.
Dette, H., Titoff, S., 2009. Optimal discrimination designs. Annals of Statistics 37, 2056–2082.
Donckels, B.M.R., De Pauw, D.J.W., Baets, B.D., Maertens, J., Vanrolleghem, P.A., 2009. An anticipatory approach to optimal experimental design for model discrimination. Chemometrics and Intelligent Laboratory Systems 95, 53–63.
Donckels, B.M.R., De Pauw, D.J.W., Vanrolleghem, P.A., Baets, B.D., 2010. An ideal point method for the design of compromise experiments to simultaneously estimate the parameters of rival mathematical models. Chemical Engineering Science 65, 1705–1719.
Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: state of the art. Chemical Engineering Science 63, 4846–4872.
Ginebra, J., 2007. On the measure of the information in a statistical experiment. Bayesian Analysis 2 (1), 167–212.
Hill, W.J., Hunter, W.G., Wichern, D.W., 1968. A joint design criterion for dual problem of model discrimination and parameter estimation. Technometrics 10, 145–160.
Hunter, W.G., Reiner, A.M., 1965. Designs for discriminating between two rival models. Technometrics 7, 307–323.
Kennedy J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, Perth, Australia, pp. 1942–1948.
Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. Canadian Journal of Mathematics 12, 363–366.
Koch, K.R., 2007. Introduction to Bayesian Statistics 2nd Ed. Springer-Verlag, Berlin.
Kullback, S., 1959. Information Theory and Statistics. John Wiley and Sons, New York.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–86.
Lauter, E., 1974. Experimental design in a class of models. Mathematische Operationsforschung und Statistik, Serie Statistik 5, 379–398.
Lindley, D.V., 1956. On a measure of the information provided by an experiment. Annals of Mathematical Statistics 27, 986–1005.
López-Fidalgo, J., Tommasi, C., Trandafir, P.C., 2007. An optimal experimental design criterion for discriminating between non-normal models. Journal of the Royal Statistical Society: Series B—Methodological 69, 231–242.
López-Fidalgo, J., Tommasi, C., Trandafir, P.C., 2008. Optimal designs for discriminating between some extensions of the Michaelis–Menten model. Journal of Statistical Planning and Inference 138, 3797–3804.
O'Brien, T.E., Rawlings, J.O., 1996. A nonsequential design procedure for parameter estimation and model discrimination in nonlinear regression models. Journal of Statistical and Planning Inference 55, 77–93.
Otsu, T., 2008. Optimal experimental design criterion for discriminating semiparametric models. Journal of Statistical Planning and Infererence 138, 4141–4150.
Parmigiani, G., Inoue, L., 2009. Decision Theory: Principles and Approaches. John and Willey Sons, Chichester, West Sussex.
Schwaab, M., Queipo, C.Q., Silva, F.M., Barreto Jr., A.G., Nele, M., Pinto, J.C., 2006. A new approach for sequential experimental design for model discrimination. Chemical Engineering Science 61, 5791–5806.
Schwaab, M., Pinto, J.C., 2007. Análise de Dados Experimentais I: Fundamentos de Estatística e Estimação de Parâmetros, E-Papers, Rio de Janeiro (in Portuguese).
Schwaab, M., Biscaia, E.C., Monteiro, J.L., Pinto, J.C., 2008a. Nonlinear parameter estimation through particle swarm optimization. Chemical Engineering Science 63, 1542–1552.
Schwaab, M., Monteiro, J.L., Pinto, J.C., 2008b. Sequential experimental design for model discrimination: taking into account the posterior covariance matrix of differences between model predictions. Chemical Engineering Science 63, 2408–2419.
Schwaab, M., Alberton, A.L., Pinto, J.C., 2010. ESTIMA&PLANEJA: Pacote computacional para estimação de parâmetros de planejamento de experimentos, Relatório Técnico, PEQ/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.
Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27, 623–656.
Soofi, E.S., 1994. Capturing the intangible concept of information. Journal of the American Statistical Association 89, 1243–1254.
Tommasi, C., 2007. Optimal designs for discriminating among several non-normal models. In: López-Fidalgo, J., Rodríguez-Díaz, J.M., Torsney, B. (Eds.), Advances in Model-Oriented Design and Analysis. Physica-Verlag, New York, pp. 213–220.
Tommasi, C., 2009. Optimal designs for both model discrimination and parameter estimation. Journal of Statistical Planning and Inference 139, 4123–4132.
Tommasi, C., Lopez-Fidalgo, J., 2010. Bayesian optimum designs for discriminating between models with any distribution. Computational Statistics, & Data Analysis 54, 143–150.
White, L.V., 1973. An extension of the general equivalence theorem to nonlinear models. Biometrika 60, 345–348.