I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0319

Optimal designs for inverse prediction in nonlinear calibration models

Nancy François, Bernadette Govaerts and Bruno Boulanger

http://www.stat.ucl.ac.be

Optimal designs for inverse prediction in nonlinear calibration models

Nancy François [1,*], **Bernadette Govaerts** [1], **Bruno Boulanger** [2]

[1]Université catholique de Louvain, Institut de Statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

[2]Eli Lilly, rue Granbonpré 11, B-1348 Mont-Saint-Guibert, Belgium.

July 31, 2003

SUMMARY

Calibration models are intended to link a quantity of interest X (e.g. the concentration of a chemical compound) to a value Y obtained from a measurement device. In this context, a major concern is to build calibration models that are able to provide precise (inverse) predictions for X from measured responses Y.

This paper aims at answering the following question : which experiments should be run to set up a (linear or nonlinear) calibration curve that maximises the inverse prediction precisions ? The well known class of optimal designs is presented as a possible solution.

The calibration model setup is first reviewed in the linear case and extended to the heteroscedastic nonlinear one. In this general case, asymptotic variance and confidence interval formulae are derived for inverse predictions.

Two optimality criteria are then introduced to quantify *a priori* the quality of inverse predictions for a given experimental design. The V_I criterion is based on the integral of the inverse prediction variance over the calibration domain and the G_I criterion on its maximum value. Algorithmic aspects of the optimal design generation are discussed.

In a last section, the methodology is applied to 4 possible calibration models (linear, quadratic, exponential and four parameters logistic). V_I and G_I optimal designs are compared to classical D, V and G optimal designs. Their predictive quality is also compared to the one of simple traditional equidistant designs and it is shown that, even if these last designs have very different shapes, their predictive quality are not far from the optimal design ones. Finally, some simulations evaluate small sample properties of asymptotic inverse prediction confidence intervals.

Keywords: Calibration, inverse prediction, nonlinear models, optimal designs

1. INTRODUCTION

In many scientific and industrial fields (e.g. biology, chemistry, pharmacy, etc), calibration models are developed to link a quantity of interest X (e.g. the concentration of a chemical compound in a product sample) to its measured value Y obtained from an analytical device (e.g. an optical density value).

*Correspondence to: Nancy François, Université catholique de Louvain, Institut de Statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve (Belgium). E-mail: francois@stat.ucl.ac.be Phone: +32-10-47.43.14 Fax: +32-10-47.30.32

There are two steps in the calibration process. First, for a serie of samples for which the value of X is known, a response Y is measured and a statistical model, the *calibration curve*, is fitted to the data. Then, at a second stage, samples with unknown X values are analysed with the test method and the values of X's are predicted from the responses Y's by inverting the calibration curve equation.

In many applications, the functional relating X and Y is linear over the calibration domain of interest and a simple linear statistical model is chosen as calibration curve. For more complicated analytical systems, the linear model may not be appropriate and nonlinear models are needed to approximate the phenomenon.

Calibration has received a wide interest in the literature and the reader may refer to the general review of Osborne (1991).

The main concern of the analyst in this context is to get "good" inverse predictions at the second step of the calibration process and to be able to quantify their precision. If the model equation is supposed to be correct, the quality of the inverse prediction may be quantified by building a confidence interval around the predicted value. In the linear case, exact confidence intervals exist and different approaches are discussed in the literature (see e.g. Graybill (1976), Chow and Liu (1995)). For nonlinear models, asymptotic confidence intervals may be built using the maximum likelihood theory (see Schwenke and Milliken (1991)) or derived by bootstrap simulations (Bonate (1993)).

A "good" calibration model should ideally give narrow confidence intervals. This depends directly on the quality of the measurement system and its related measurement repeatability. Another important factor is the form of the design of experiment used to set up the calibration curve. In calibration, a traditional approach consists of choosing k support design points $X_1, X_2, ..., X_k$ evenly distributed over the calibration domain $[X_{min}, X_{max}]$ and repeating the measurement 3 to 5 times at each point.

The statistical literature shows that this approach is not optimal in the sense that it does not provide minimum variance estimators for the model parameters and, in consequence, the best predictions or inverse predictions. The problem of finding a "good" design to fit a given statistical model is treated in the optimal design literature (see e.g. Atkinson and Donev (1992)) and has already been adressed in the linear calibration case by Buonaccorsi (1986) and Buonaccorsi and Iyer (1986).

The present paper is mainly concerned with the derivation of optimal designs when the calibration curve is nonlinear. It first reviews the linear case and shows that in this context the design problem is quite straigthforward. It discusses then how to obtain inverse confidence intervals in the heteroscedastic nonlinear case and propose two optimality criteria to measure the quality of an experimental design in inverse prediction. Algorithmic aspects of the generation of optimal designs are also discussed.

The last section illustrates the methodology on four common models : linear, quadratic, exponential and 4 parameters logistic models, and compares the optimal designs to traditionnal equidistant designs.

2. INVERSE PREDICTION IN THE LINEAR CALIBRATION MODEL

In the linear case, calibration is classically treated as follows. Let X and Y denote respectively the variable of interest and the measured response. We suppose that X and Y are related by the linear model :

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$$

where the ϵ_i's are i.i.d. with 0 mean and constant variance σ^2. In most cases, ϵ_i is supposed to be normally distributed for inferential purpose. Estimates b_1 and b_2 of

β_1 and β_2 are obtained from the calibration data sample $(X_i, Y_i), i = 1, ..., n$ by the usual least squares formula.

A new product sample with unknown value X_0 of X is then measured one or several (say m) times and X_0 is estimated from the observed responses $Y_{01}, Y_{02}, ..., Y_{0m}$ by the inverse prediction formula :

$$\hat{X}_0 = \frac{\bar{Y}_0 - b_1}{b_2} \text{ with } \bar{Y}_0 = \frac{1}{m}\sum_{i=1}^{m} Y_{0i}$$

The most common definition of confidence interval for X_0 is the set of X's such that $\bar{Y}_0$ belongs to the prediction interval of $\bar{Y}$ given X. This is illustrated in Figure 1.

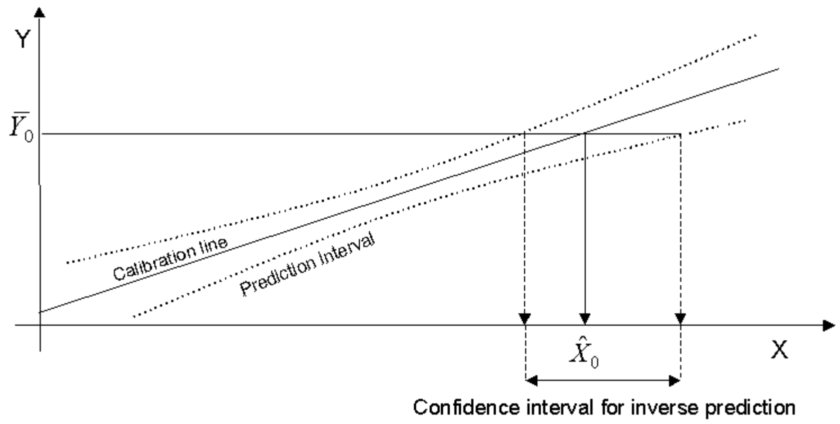

Figure 1: Confidence interval for inverse prediction in the linear case.

This confidence interval is obtained by inverting the traditional exact prediction interval formula. It takes the following form :

$$\hat{X}_{Up(Low)} = \frac{\hat{X}_0}{1-g^2} - \frac{\bar{X}g^2}{1-g^2} \pm \frac{t_{n-2;1-\alpha/2}S}{b_2^2(1-g^2)}\sqrt{b_2^2(1-g^2)\left(\frac{1}{m}+\frac{1}{n}\right)+\frac{(\bar{Y}_0-\bar{Y})^2}{S_{XX}}}$$

where

$$g^2 = \frac{t^2_{n-2;1-\alpha/2}S^2}{b_2^2 S_{XX}}, S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2, \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - b_1 - b_2 X_i)^2$$

and $t_{n-2;1-\alpha/2}$ is the quantile of level $1-\alpha/2$ of the t-distribution with $(n-2)$ degrees of freedom. A finite confidence interval occurs if and only if $g^2 < 1$, that is if β_2 is significant at level α.

Other versions of this interval are provided in the literature. Graybill (1976) gives a very similar interval by including the Y_{0i}'s in the estimation S^2 of σ^2. Chow and Liu (1995) discuss the possibility to estimate the reverse model $X_i = b_1^* + b_2^* Y_i + \epsilon_i^*$ and use the traditional prediction interval to derive a confidence interval for X_0. Finally, Neter *et al* (1996) base their interval on the asymptotic properties of maximum likelihood estimators. Their method is generalised in the nonlinear case in section 3.

In all these approaches, one can adress the question of what is the calibration experimental design which will give the "best" (i.e. the shorter) confidence intervals.

This is not straigthforward, since the confidence interval are not of the same length over the calibration domain of interest. One must then define a way to summarize the confidence interval quality over the whole domain. This will be discussed for the general case in section 4. We will show that, in the linear case, the optimal solution consists of splitting evenly the design points at the two ends of the calibration interval $[X_{min}, X_{max}]$. This result may be derived analytically from the interval above and is also true for intervals derived by other authors. This design corresponds to well known results available in the optimal design literature about D and G-optimal designs in the simple linear model. It is not necessarily a very realistic design in the calibration context because it supposes implicitely that the analyst is sure of the calibration domain and the linear shape of the calibration curve.

Discussion on design aspects in linear calibration may be found in Buanaccorsi (1986), Buonaccorsi and Iyer (1986).

3. INVERSE PREDICTION IN THE NONLINEAR CALIBRATION MODEL

3.1. The model

Let's give the following form to the general nonlinear calibration model :

$$Y_i = f(X_i; \beta) + \epsilon_i$$

where f is a strictly (usually increasing) monotonic function on the calibration domain $[X_{min}, X_{max}]$ ($X_{min} \geq 0$) and β, a vector of p unknown parameters $(\beta_1, \beta_2, ..., \beta_p)'$. Here, the error terms ϵ_i's are independent and supposed to follow a Normal distribution with mean 0 and variances σ_i^2 that are not necessarily constant over the calibration domain. Davidian and Giltinan (1995) propose to take the following form for the variance function :

$$\sigma_i^2 = \sigma^2 g^2(f(X_i; \beta), \gamma)$$

where g and γ are supposed to be known.

When the measurement variance is constant over the domain, g is simply the identity function. For other cases, a common choice for g is the power function : $f^\gamma(X_i; \beta)$ with $0 \leq \gamma \leq 1$. Davidian and Haaland (1990) present a wide range of possible forms for the error variance function.

3.2. Estimation of the calibration model parameters

As in the linear case, the calibration model is derived from a calibration data set (X_i, Y_i) with $i = 1, ..., n$ and parameters estimated by least squares. In this more general case, weighted least square (WLS) is required to take into account the heteroscedasticity of the residuals. The least squares function is then defined as :

$$\sum_{i=1}^{n} w_i (Y_i - f(X_i; \beta))^2 \text{ with } w_i = 1/g^2(f(X_i; \beta), \gamma)$$

Under normality of ϵ_i's, WLS estimators are equivalent to maximum likelihood ones.

3.3. Inverse prediction and related confidence interval

In a daily use of the calibration model, let's $Y_{01}, Y_{02}, ..., Y_{0m}$ be m measures of the same product sample with unknown concentration X_0. The inverse prediction of X_0 is given by :

$$\hat{X}_0 = f^{-1}(\bar{Y}_0; \hat{\beta}) \triangleq h(\bar{Y}_0; \hat{\beta}) \text{ where } \bar{Y}_0 = \frac{1}{m} \sum_{i=1}^{m} Y_{0i}$$

We suppose the existence of an analytical form for h and that h is differentiable over the domain of interest.

Since $\hat{\beta}$ and $\bar{Y}_0$ are maximum likelihood estimators of the unknown parameters β and $\mu_{Y_0} = f(X_0;\beta)$, $\hat{X}_0$ is also a maximum likelihood estimator and an asymptotic variance formula may then be derived for $\hat{X}_0$ by the delta method :

$$V(\hat{X}_0) = \frac{\partial h(\mu_{Y_0};\beta)}{\partial(\mu_{Y_0},\beta)'} V(\hat{\beta},\bar{Y}_0) \frac{\partial h(\mu_{Y_0};\beta)}{\partial(\mu_{Y_0},\beta)}$$

Where $V(\hat{\beta},\bar{Y}_0)$ is the variance covariance matrix of $(\hat{\beta},\bar{Y}_0)$. Since $\hat{\beta}$ and $\bar{Y}_0$ are independent, this becomes :

$$V(\hat{X}_0) = \frac{\partial h}{\partial \beta'} V(\hat{\beta}) \frac{\partial h}{\partial \beta} + \left(\frac{\partial h}{\partial \mu_{Y_0}}\right)^2 V(\bar{Y}_0)$$

These two formulae may also be found in Davidian and Giltinan [1995] or in Schwenke and Milliken (1991) but the later neglects the variance of $\bar{Y}_0$.

An asymptotic confidence interval for X_0 may then be derived as :

$$\hat{X}_0 \pm z_{1-\alpha/2}\sqrt{V(\hat{X}_0)}$$

This interval has the advantage to be quite simple but has the drawback to be symmetric around $\hat{X}_0$ while in reality it should not. This is especially problematic for functions which tend to asymptotes at their domain borders as the exponential or logistic models.

It is interesting to study more deeply the components of $V(\hat{X}_0)$:

- $V(\hat{\beta})$ is the $p \times p$ asymptotic variance-covariance matrix of $\hat{\beta}$ defined in the weighted case as :

  $$V(\hat{\beta}) = \sigma^2 \left[\frac{\partial f}{\partial \beta'} \Omega \frac{\partial f'}{\partial \beta}\right]^{-1} \triangleq \sigma^2 I(\beta,\chi)^{-1}$$

  where Ω is the diagonal matrix of weigths $w_i's$, $\frac{\partial f}{\partial \beta'}$ the $n \times p$ matrix of partial derivatives $\frac{\partial f(X_i;\beta)}{\partial \beta_j}$. The second part of this formula is in fact the inverse of I, the model information matrix which depends on β and on the design structure χ.

  An estimator of $V(\hat{\beta})$ is obtained by replacing β by $\hat{\beta}$ and σ^2 by

  $$S^2 = \frac{1}{n-p}\sum_{i=1}^{n} w_i (Y_i - f(X_i;\hat{\beta}))^2$$

  .

- $V(\bar{Y}_0)$, the second component of $V(\hat{X}_0)$ may be estimated as :

  $$V(\bar{Y}_0) = \frac{1}{m} S^2 g^2(f(\hat{X}_0;\hat{\beta}),\gamma)$$

- The two partial derivatives $\frac{\partial h}{\partial \beta}$ and $\frac{\partial h}{\partial \mu_{Y_0}}$ can be estimated by deriving first $h(\mu_{Y_0};\beta)$ with respect to β and μ_{Y_0} and then replacing β and μ_{Y_0} by their estimators $\hat{\beta}$ and $\bar{Y}_0$.

Note that, in $V(\hat{X}_0)$, $V(\hat{\beta})$ is the only part which directly depends on the design structure. This property will take all its importance in the practical derivation of optimal designs in section 4.

In the particular case of linear calibration models, Neter *et al* (1996) suggest to take this asymptotic formula approach to get a confidence interval for X_0. They show that $V(\hat{X}_0)$ is given by :

$$V(\hat{X}_0) = \frac{\sigma^2}{b_2}\left(\frac{1}{m} + \frac{1}{n} + \frac{(\hat{X}_0 - \bar{X})^2}{S_{XX}}\right)$$

A comparison of this expression and the exact formula of section 2 shows that this approximation holds when $g^2 = \frac{t^2_{n-2;1-\alpha/2}S^2}{b_2 S_{XX}}$ is small, i.e. when β_2 is highly significant.

4. OPTIMAL DESIGNS FOR NONLINEAR CALIBRATION MODELS

This section adresses the question of finding a calibration experimental design $X_i, i = 1, ..., n$ that guarantees good qualities for the inverse predictions. As for the linear model, the equidistant design is a first possible choice in the nonlinear context but it is unclear whether it is optimal or not.

In front of such an open question, optimal designs (Atkinson and Donev (1992)) are certainly a possible solution. The general principle of optimal designs is simple: for a given problem, the idea is to search into the space of possible designs the one which optimizes a given criterion of quality, usually a function of the variance covariance matrix of $\hat{\beta}$: $\sigma^2 I(\beta, \chi)^{-1}$.

The derivation of such designs requires two components. First, an optimality criterion able to quantify the quality of a given design with respect to the goal of the study. Second, an algorithm able to find a design in the design space that optimizes the criterion of interest.

4.1. Classical optimality criteria

In the context of linear and nonlinear models, three classical optimality criteria may certainly be emphasized. First, the *D*-criterion is a global measure of the precision of the estimated model parameters and is defined as the determinant of the inverse of the information matrix of the model. It should be minimized and can be written in our notations as:

$$D - \text{criterion} = |I(\beta, \chi)|^{-1} = \left|\frac{\partial f}{\partial \beta'}\Omega\frac{\partial f'}{\partial \beta}\right|^{-1}$$

The G and V optimality criteria measure the capacity of the model to provide precise predictions $\hat{Y} = f(X; \hat{\beta})$ over a domain of interest $\mathcal{D} = [X_{min}, X_{max}]$.

They are respectivelly defined as the maximum and the average prediction variance on $\mathcal{D}$ and should be minimized :

$$\begin{aligned}
&G - \text{criterion} = \max_{x \in \mathcal{D}} \frac{\partial f(x;\beta)}{\partial \beta'}|I(\beta,\chi)|^{-1}\frac{\partial f(x;\beta)}{\partial \beta} \\
&V - \text{criterion} = \frac{1}{\Delta_{\mathcal{D}}}\int_{X_{min}}^{X_{max}} \frac{\partial f(x;\beta)}{\partial \beta'}|I(\beta,\chi)|^{-1}\frac{\partial f(x;\beta)}{\partial \beta}dx \\
&\text{where} \quad \Delta_{\mathcal{D}} = X_{max} - X_{min}
\end{aligned}$$

The evaluation of these criteria for a given design in the nonlinear case is not direct since it requires the knowledge of the parameters' vector β which is in principle unknown when the design is set up. Two approaches are proposed in the literature. The easiest one consists of selecting an *a priori* value β_p for β on the basis of the context knowledge and then deriving the design conditionally on β_p. The resulting designs are called *locally* optimal designs. The second method consists of choosing a prior (multivariate) distribution for β and selecting the design which minimizes the average optimality criterion with respect to this distribution. The resulting designs, called *bayesian* optimal designs, are much more complicated to generate since they require numerical integration at each step of the design search.

In the calibration context, the first approach may be quite realistic since calibration is an iterative process where the researcher has often some good knowledge of the expected model parameters. In the next section, we suppose that *a priori* parameters β_p are available but the discussion may easily be generalized to bayesian optimal designs.

4.2. Inverse prediction optimality criteria

Two optimality criteria are proposed here to summarize the quality of an inverse prediction over the calibration domain $\mathcal{D}$. As this quality may be quantified by the length of the inverse prediction confidence interval : $2z_{1-\alpha/2}\sqrt{V(\hat{X}_0)}$, we suggest to base the criteria on $V(\hat{X}_0)$.

More precisely, we have seen in section 3.3 that the first part of $V(\hat{X}_0)$:

$$\sigma^2 \frac{\partial h(\mu_{Y_0}, \beta)}{\partial \beta'} I(\beta, \chi)^{-1} \frac{\partial h(\mu_{Y_0}, \beta)}{\partial \beta}$$

is the only part directly depending on the design while the second part is *a priori* the same for all possible designs. Moreover, in this expression :

- σ^2 is usually unknown but can be dropped because it is independent of the design,
- $I(\beta, \chi)$ depends on the design χ and on β set *a priori* to β_p, and
- $\frac{\partial h(\mu_{Y_0};\beta)}{\partial \beta}$ is evaluated at a point X_0 in $\mathcal{D}$ by replacing, after derivation, β by β_p and μ_{Y_0} by it's expected value $f(X_0, \beta_p)$. Let's denote by $H(x;\beta_p)$ this $p \times 1$ vector of derivatives evaluated at $x \in \mathcal{D}$.

The two optimality criteria proposed to summarize $V(\hat{X}_0)$ over the calibration domain for a given design χ are defined as :

$$G_I - \text{criterion} = \max_{x \in \mathcal{D}} H(x;\beta_p)' I(\beta_p, \chi)^{-1} H(x;\beta_p)$$

$$V_I - \text{criterion} = \frac{1}{\Delta_{\mathcal{D}}} \int_{X_{min}}^{X_{max}} H(x;\beta_p)' I(\beta_p, \chi)^{-1} H(x;\beta_p) dx$$

Both criteria should be minimized and follow the same principle than the G and V ones in being proportional to the maximum and mean inverse prediction variance over the domain $\mathcal{D}$. Similar criteria have already been proposed in the linear case by Buonaccorsi (1986) and by Buonaccorsi and Iyer (1986) under the name of AV and M optimality.

The calculation of these criteria for a given design is not straigthforward since they require the evaluation of $H(x;\beta_p)$ over the continuous domain $\mathcal{D}$. In practice, the best way to handle this is to evaluate H on a grid in $\mathcal{D}$. Note however that H does not depend on the design χ and can then be evaluated once before starting up the search algorithm. Moreover, the V_I criterion may be rewritten to make the criterion calculation more efficient. It can be shown to be equivalent to :

$$V_I - \text{criterion} = \text{trace}\left[I(\beta,\chi)^{-1}\left(\frac{1}{\Delta_{\mathcal{D}}}\int_{X_{min}}^{X_{max}} H(x;\beta_p)H(x;\beta_p)'dx\right)\right]$$

The second part of the formula is the integral (element by element) of a $p \times p$ matrix and should only be evaluated once before searching the optimal design. The efficiency of computation of V_I in the search algorithm will then only rely on the partial product of two $p \times p$ matrices and is not influenced by the grid fineness in the integral evaluation.

4.3. Optimal design generation

Several algorithms have been developed for the generation of exact optimal designs especially in the context of D-Optimal designs in the multiple linear regression model. Most of them are working on the exchange principle : a starting design is chosen in the space of all possible designs and is then improved using different styles of exchange procedures until an optimum is reached. The main disadvantage of these algorithms is that they do not guarantee to reach global optimal designs and often need to be run several times to get a satisfactory solution. The reader can find a review of classical exchange algorithms in Nguyen and Miller (1992).

In the context of nonlinear models, few literature is devoted to exact design generation and most people develop their own algorithmic strategy. Two algorithms have been tested on several models in this univariate calibration context.

The first one follows the principle of the modified Federov exchange algorithm (Federov (1992), Miller and Nguyen (1994)). It starts with a random design of n points X_i's in the experimental domain $\mathcal{D}$ and then tries in turn to exchange each design point X_i with the best possible X in $\mathcal{D}$. If the exchange improved the criterion, X_i is replaced by X in the design and the next design point is investigated. The algorithm stops when no design point X_i can improve the criterion value.

The second algorithm starts also from a random design but uses a global n dimensional optimization program to search the optimum design (*nlminb* function in S-Plus). This second procedure has revealed to provide much better solutions than the first one. However, the algorithm needs also to be run several times with different starting designs to get a good solution.

5. APPLICATION TO SOME LINEAR AND NONLINEAR MODELS

This section applies the techniques discussed in the previous section on four different possible calibration models having quite similar shapes but different mathematical forms. For each model, the optimal designs obtained with the five optimality criteria (D, V, G, V_I and G_I) are compared to each other and to the designs derived for the other models. The predictive quality of the V_I and G_I optimal designs are then compared to the one of equidistant designs which are very popular in calibration applications.

5.1. Models definition

The four models of interest are the linear, quadratic, exponential and 4 parameters logistic (4pl) models. Their equations are given by :

$$
\begin{aligned}
M_1 &: \quad Y = \beta_1 + \beta_2 X + \epsilon \\
M_2 &: \quad Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \epsilon \\
M_3 &: \quad Y = \beta_1 \times \exp\left(\beta_2 X\right) + \epsilon \\
M_4 &: \quad Y = \beta_1 + \frac{\beta_2 - \beta_1}{1 + \exp(\beta_4(\beta_3 - X))} + \epsilon
\end{aligned}
$$

Error terms are supposed of constant variance σ^2. The quadratic model is not monotonic on $\mathbb{R}$ and will be restricted to the domain where it is strictly increasing.

The inverse prediction formulae can be directly obtained from the model definitions:

$$
\begin{aligned}
M_1 &: \quad X = \frac{Y - \beta_1}{\beta_2} \\
M_2 &: \quad X = \frac{1}{2\beta_3}\left(-\beta_2 + \sqrt{\beta_2^2 - 4\beta_3(\beta_1 - Y)}\right) \\
M_3 &: \quad X = \frac{1}{\beta_2}(\log(Y) - \log(\beta_1)) \\
M_4 &: \quad X = \beta_3 - \frac{1}{\beta_4}\log\left(\frac{\beta_2 - Y}{Y - \beta_1}\right)
\end{aligned}
$$

The quadratic model has two possible roots on $\mathbb{R}$ but the solution given here is the root corresponding to the increasing part of the function. The root formula is the same for convex or concave quadratic functions.

For the derivation of optimal designs, we suppose that the calibration domain of interest is $[2, 10]$ and choose prior values β_p for the parameters β which ensure that the 4 models pass through the same points $(X, Y) = (2, 1)$ and $(10, 5)$. These parameters vectors are :

$$
\begin{aligned}
&\text{Linear model : } \beta_p = (0, 0.5)' \\
&\text{Quadratic model : } \beta_p = (5/8, 1/8, 1/32)' \\
&\text{Exponential model : } \beta_p = (5^{-0.25}, \log(5)/8)' \\
&\text{4pl model : } \beta_p = (0.75, 4.5, 6, \log(17)/4)'
\end{aligned}
$$

These four models are illustrated in Figure 2.

5.2. Optimality criteria calculation

The calculation of optimality criteria differs for these four models by the form of the partial derivatives $\partial f(x;\beta)/\partial\beta$ and $H(x;\beta)$. They are given in the appendix for the 4 models and have *in fine* all quite simple forms.

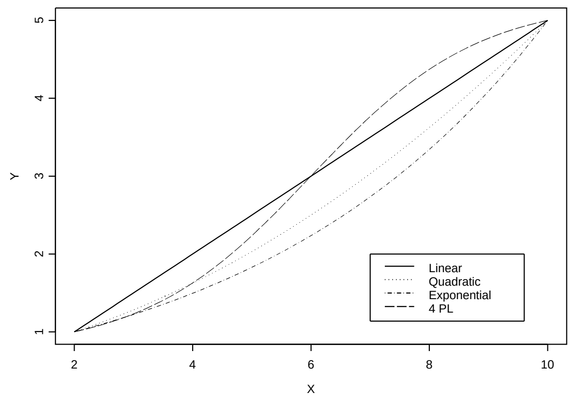

Figure 2: Linear, quadratic, exponential and 4pl models under test

5.3. Optimal designs

This section compares optimal designs for the four models of interest. For each model, D, V, G, V_I and G_I designs with 24 design points have been generated and are presented in Figure 3 and 4. For each design, the support points and corresponding number of replicates are given and design points drawn with a size proportionnal to the number of replicates. Designs of 24 runs have been chosen to give the algorithms the chance to distribute evenly the replicates over the optimal design support points. It is also a reasonable design size to set up a calibration model.

Table 1 and 2 give the exact values of the optimal designs support points, the number of replicates and the values of the corresponding optimal criterion. The right part of Figure 3 and 4 presents also the shape of the inverse prediction variance over the calibration domain for the V_I and G_I optimal designs. The drawn variance function is the part of $V(\hat{X}_0)$ which is "design-dependent" or equivalently $V(\hat{X}_0)$ with $\sigma^2 = 1$ and $m = \infty$.

This section compares also the (inverse) prediction quality of V_I and G_I optimal designs to the one of traditional equidistant designs. For this purpose, a 24 runs design with 6 equidistant support points in 2, 3.6, 5.2, 6.8, 8.4, 10 and 4 replicates at each point has been chosen. Table 1 and 2 give the value of the optimality criteria for this design and Figure 3 and 4 (right part) its inverse prediction variance.

From all these results, the following remarks can be drawn :

- All optimal designs are saturated in the sense that the number of support points is equal to the number of model parameters. This is an usual property of optimal designs which emphasizes one of their important characteristic: they are very good to fit a given model but provide few degrees of freedom to test a model lack of fit. In calibration this property is not necessarily a default since the model shape is often well known.

- For all models, the D and G optimal designs are identical. This is not surprising, while not automatic for exact optimal designs, on the basis of the general equivalence theorem which states the equivalence between these two criteria when generating continuous optimal designs (Atkinson and Donev (1992)).

- For the V_I and G_I optimal designs, more replicates are always observed in the

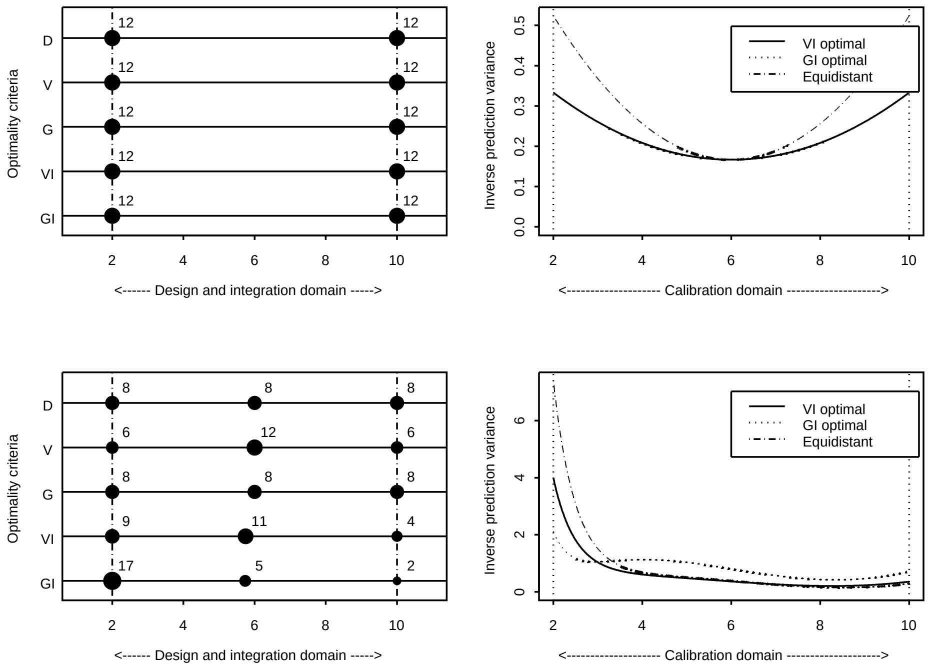

Figure 3: Comparison of D, V, G, V_I and G_I optimal designs and variance fonctions for the linear model (above) and quadratic model (below)

part of the calibration domain where the calibration curve slope is smaller. This can be explained probably by the fact that, in this area, the inverse prediction is less precise or equivalently the inverse prediction variance is higher and the replicates are then aimed at decreasing the lack of predictive ability of the model in that area. This property is mainly observed for the G_I optimal designs.

- For the linear model, all optimal designs evenly put the design points at the two domain bounds : 2 and 10. This result is well known in the optimal design literature for D, V and G optimal designs and this shows that it applies also to V_I and G_I designs. Moreover, this result may be easily proven analytically in the linear case even when the error variance is not constant over the domain.

- For the quadratic model, the D, V, G optimal designs are symmetrical while V_I and G_I ones contain more replicates in the left part of the domain for the reason explained above.

- The optimal designs obtained for the exponential model are much less intuitive. They all have two support points but, strangely, the left one is not at the left bound of the experimental domain. This is difficult to explain intuitively especially for V_I and G_I optimal designs were the higher inverse prediction variance is located in x=2. Note that this behaviour is not observed for all possible prior parameter vectors β_p.

- The 4PL optimal designs are all quite intuitive. They are symmetrical around 6 because the model curve is symmetrical and are nearly equidistant designs.

Linear Model					
	Optimal design			Equidistant design	Criteria
Criteria	Design points	Replicates	Criteria value	Criteria value	ratios
D	2 10	12 12	0.0001	0.0002	2.3
V	2 10	12 12	0.0556	0.0714	1.3
G	2 10	12 12	0.0833	0.1309	1.6
V_I	2 10	12 12	0.2222	0.2857	1.3
G_I	2 10	12 12	0.3333	0.5238	1.6
Quadratic Model					
	Optimal design			Equidistant design	Criteria
Criteria	Design points	Replicates	Criteria value	Criteria value	ratios
D	2 6 10	8 8 8	1.1910^{-7}	2.3810^{-7}	2
V	2 6 10	6 12 6	0.089	0.099	1.1
G	2 6 10	8 8 8	0.125	0.205	1.6
V_I	2 5.746 10	19 11 4	0.603	0.807	1.3
G_I	2 5.735 10	17 5 2	2.12	7.39	3.5

Table 1: D, V, G, V_I and G_I optimal designs for the linear and the quadratic models

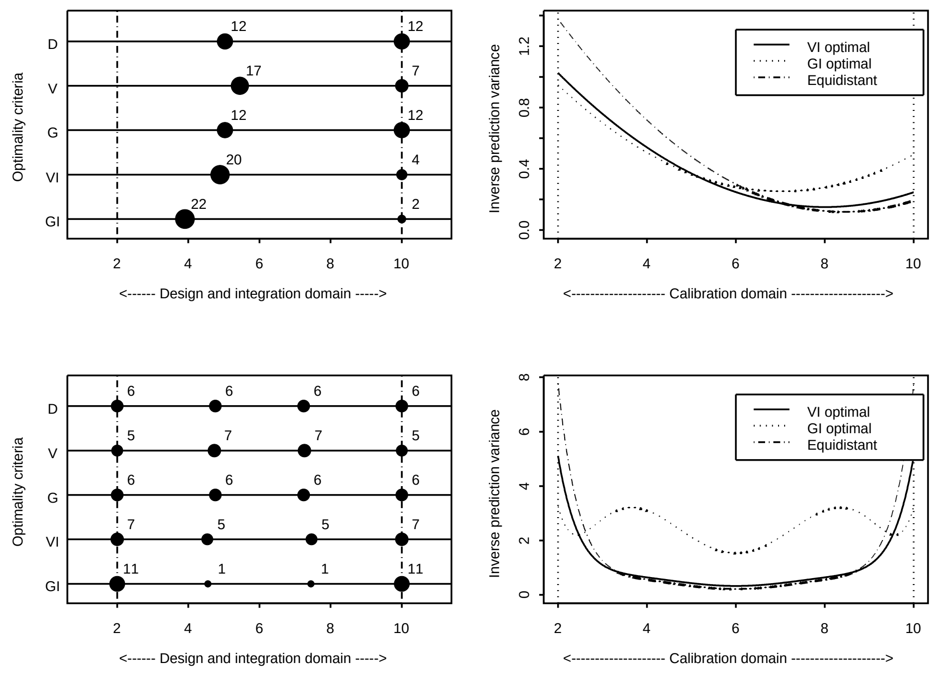

Figure 4: Comparison of D, V, G, V_I and G_I optimal designs and variance fonctions for the exponential model (above) and 4pl model (below)

As expected, the V_I and G_I designs contain more replicates in 2 and 10 where the model slope is lower and closer to the asymptotes.

- All G_I optimal designs obtained here tend to show that this optimality criterion does not necessarily generate interesting designs. As a matter of fact,

Exponential Model					
	Optimal design			Equidistant design	Criteria
Criteria	Design points	Replicates	Criteria value	Criteria value	ratios
D	5.029 10	12 12	1.4910^{-6}	2.6610^{-6}	1.8
V	5.453 10	17 7	0.0613	0.0704	1.2
G	5.029 10	12 12	0.0833	0.196	2.3
V_I	4.894 10	20 4	0.378	0.462	1.2
G_I	2.907 10	22 2	0.947	1.377	1.5
Four parameters logistic Model					
	Optimal design			Equidistant design	Criteria
Criteria	Design points	Replicates	Criteria value	Criteria value	ratios
D	2 4.759 7.241 10	6 6 6 6	0.002	0.003	1.7
V	2 4.733 7.267 10	5 7 7 5	0.136	0.148	1.1
G	2 4.759 7.241 10	6 6 6 6	0.167	0.217	1.3
V_I	2 4.537 7.463 10	7 5 5 7	1.035	1.23	1.2
G_I	2 4.550 7.450 10	11 1 1 11	3.25	7.76	2.4

Table 2: D, V, G, V_I and G_I optimal designs for the quadratic and 4PL models

even if these designs provide lower maximum prediction variance in the worst domain part, this happens to the detriment of the level of the prediction variance in other domain parts. This is clear from the right graphics of Figure 3 and 4 and is especially clear for the 4PL model.

- The comparison of optimal and equidistant designs is also instructive. The prediction variance curve in Figure 3 and 4 and the criteria values in Table 1 and 2 show that the loss of using an equidistant design is always present but not necessarily big. In practice, such a comparison should guide the experimenter to choose between the traditional robust equidistant designs and the more precise but less intuitive optimal ones.

5.4. Validity of asymptotic variance formulae

All the developments to generate inverse prediction confidence intervals and optimal designs presented in the previous sections are based on an asymptotic inverse prediction variance formula for $V(\hat{X}_0)$. Since, for nonlinear models, there is no simple alternative to this approximate solution, one may ask the following question: is this variance reasonably valid in small samples ?

We have approached this problem through a Monte Carlo study and results are quite encouraging. More precisely, for the models and designs described in section 5.1 and 5.3, we have simulated the two-steps calibration process. First, response datasets have been randomly generated for the given 24 run designs to set up the calibration curve. Then, for each estimated curve, data have been generated in order to mimic the inverse prediction scheme and to evaluate the quality and variability of inverse predictions over the calibration domain. Finally, the inverse prediction variance $V(\hat{X}_0)$ observed in these simulations have been compared to the associated asymptotic ones. A graphical study of the results has shown no systematic bias between both variances.

Figure 5 and 6 give two examples of such simulations. Figure 5 compares the two variances for the exponential model using the 24 run V_I optimal design. Figure 6 is derived from similar simulations for the 4 parameters logistic model.

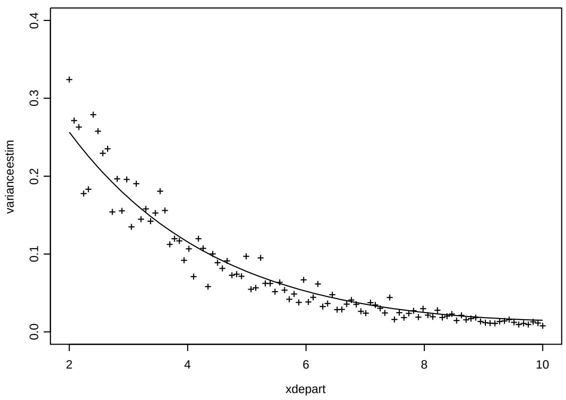

Figure 5: Comparison of asymptotic and small sample inverse prediction variances for the exponential model

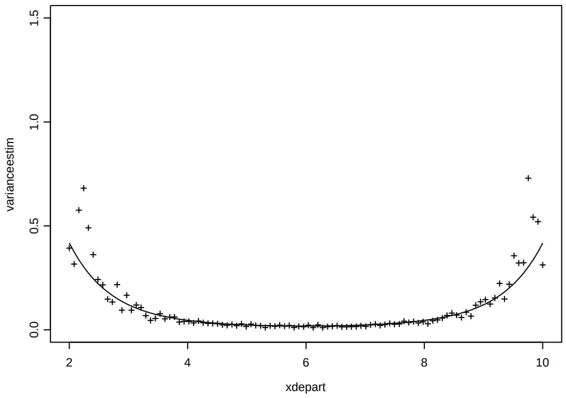

Figure 6: Comparison of asymptotic and small sample inverse prediction variances for the 4PL model

6. CONCLUSION AND FURTHER DEVELOPMENTS

This paper proposes a methodology to generate optimal designs to set up nonlinear calibration models. These designs are based on two possible optimality criteria, the V_I and the G_I criteria aimed at summarizing the precision of the inverse prediction over the calibration domain. The methodology has been applied to different possible nonlinear models and the designs were compared to classical D, V and G optimal designs and equidistant designs.

The examples show that, among the two optimality criteria, the V_I one is certainly the best choice as the G_I one tends to overweight the design in the domain area where the inverse prediction variance is high. They also show that classical equidistant designs are not always so worst compared to the optimal ones. In such cases, the practitioner may then take the optimal design as a reference to decide whether an equidistant design is a reasonable choice or not.

All generated designs in the examples are saturated. This is often observed for locally optimal designs in nonlinear models. It would be of great interest to compare them to bayesian optimal designs. In principle, this would provide designs with a number of design points greater than the number of model parameters. This is not conceptually difficult but computer intensive because it requires a multidimensional integration over the parameters space at each cycle of the design search.

7. APPENDIX : PARTIAL DERIVATIVES OF CALIBRATION MODELS

This appendix gives the analytical form of partial derivatives $\partial f(X;\beta)/\partial\beta$ and $\partial h(\mu_Y;\beta)/\partial\beta$ for the four models presented in section 5.1. Note that $\partial h/\partial\beta$ has been calculated in two steps : first $h(\mu_Y,\beta)$ has been derived with respect to β and second, μ_Y replaced by $f(X;\beta)$ in the expression. This second step leads to great simplifications of the analytical expressions.

For the linear model we have :

$$\partial f/\partial\beta = (1, X)' \qquad \partial h/\partial\beta = -\frac{1}{\beta_2}(1, X)'$$

For the quadratic model :

$$\partial f/\partial\beta = (1, X, X^2)' \qquad \partial h/\partial\beta = -\frac{1}{\beta_2 + 2\beta_3 X}(1, X, X^2)'$$

For the exponential model :

$$\partial f/\partial\beta = \exp(\beta_2 X)(1, \beta_1 X)' \qquad \partial h/\partial\beta = -\frac{1}{\beta_2}(\frac{1}{\beta_1}, X)'$$

For the four parameters logistic model :

$$\partial f/\partial\beta = \frac{1}{1+A}\left(A, 1, \frac{-\beta_4(\beta_2-\beta_1)A}{1+A}, \frac{-(\beta_3-X)(\beta_2-\beta_1)A}{1+A}\right)'$$

$$\partial h/\partial\beta = \left(-\frac{1+A}{\beta_4(\beta_2-\beta_1)}, -\frac{1+A}{\beta_4(\beta_2-\beta_1)A}, 1, \frac{\beta_3-X}{\beta_4}\right)'$$

where $A = \exp(\beta_4(\beta_3 - X))$

ACKNOWLEDGEMENTS

The authors are grateful to Eli Lilly (Mont-Saint-Guibert - Belgium) for having inspired this research and for fruitful discussion and collaboration.

REFERENCES

1. Atkinson A.C. and Donev A.N. (1992), *Optimal Experimental Designs*, Clarendon Presse, Oxford
2. Bonate P. L. (1993), Approximate confidence intervals in calibration using the bootstrap, *Analytical chemistry*, **65**,1367-1372.
3. Buonaccorsi J.P. (1986), Design considerations for calibration, *Technometrics*, **28**, 149-155.
4. Buonaccorsi J.P. and Iyer, H. K. (1986), "Optimal Designs for Ratios of Linear Combinations in the General linear Model", *Journal of Statistical Planning and Inference*, **13**, 345-356.
5. Chow S. C. and Liu J. P. (1995), *Statistical design and analysis in pharmaceutical science: validation, process controls and stability*, Ed. Marcel Dekker, New York.
6. Davidian M. and Haaland P. D. (1990), Regression and calibration with non constant error variance, *Chemometrics and Intelligent Laboratory Systems*, **9**, 231-248.
7. Davidian M. and Giltinan D.M. (1995), *Nonlinear models for repeated measurement data*, Monographs on statistics and applied probability, Chapman and Hall.
8. Dette H. and Wong W. K. (1999), Optimal design when the variance is a function of the mean, *Biometrics*, **55**, 925-929.
9. Federov V.V. (1972), *Theory of optimal experiments*, Academic Press, New York.
10. Graybill F. A. (1976), *Theory and application of the linear model*, Ed. Duxbury Press, Boston.
11. Miller A. J. and Nguyen, N-K. (1994). A Fedorov exchange algorithm for D-optimal designs, *Applied Statistics*, **4**3, 669-678.
12. Neter J., Kutner M. H., Nachtsheim C. J., and Wasserman W. (1996), *Applied Linear Statistical Models*, Irwin
13. Nguyen, N-K. and Miller A. J. (1992). A review of some exchange algorithms for constructing discrete D-optimal designs, *Computational Statististics and Data Analysis*, **14**, 489-498.
14. Osborne C. (1991), Statistical calibration: a review, *International Statistical Review*, **59 (3)**, 309-336.
15. Schwenke J. R. and Milliken G.A. (1991), On the calibration problem extended to nonlinear models, *Biometrics*, **47**, 563-574.