



Available online at www.sciencedirect.com



**INTERNATIONAL JOURNAL OF
Food Microbiology**

International Journal of Food Microbiology 100 (2005) 153–165

www.elsevier.com/locate/ijfoodmicro

Optimal experiment design for cardinal values estimation: guidelines for data collection

K. Bernaerts, K.P.M. Gysemans, T. Nhan Minh, J.F. Van Impe*

*BioTeC-Bioprocess Technology and Control, Department of Chemical Engineering, Katholieke Universiteit Leuven, W. de Crooylaan 46,
B-3001 Leuven, Belgium*

Received 27 September 2004; accepted 6 October 2004

Abstract

Optimal experiment design for parameter estimation (OED/PE) is an interesting technique for modelling practices when aiming for maximum parameter estimation accuracy. Nowadays, experimental designs for secondary modelling within the field of predictive microbiology are mostly arbitrary or based on factorial design. The latter type of design is common practice in response surface modelling approaches. A number of levels of the factor(s) under study are selected and all possible treatment combinations are performed. It is however not always clear which levels and treatment combinations are most relevant. An answer to this question can be obtained from optimal experiment design for—in this particular case—parameter estimation. This technique is based on the extremisation of a scalar function of the Fisher information matrix. The type of scalar function determines the final focus of the optimised design.

In this paper, optimal experiment designs are computed for the cardinal temperature model with inflection point (CTMI) and the cardinal pH model (CPM). A model output sensitivity analysis (depicting the sensitivity of the model output to a small change in the model parameters) yields a first indication of relevant temperature or pH treatments. Performed designs are: *D*-optimal design aiming for a maximum global parameter estimation accuracy (by minimising the determinant of the Fisher information matrix), and *E*-optimal design improving the confidence in the most uncertain model parameter (by maximising the smallest eigenvalue of the Fisher information matrix). Although lowering the information content of a set of experiments, boundary values on the design region need to be imposed during optimisation to exclude unworkable experiments and partly account for incorrect nominal parameter values.

Opposed to the frequently applied equidistant or arbitrary treatment placement, optimal design results show that typically four informative temperature or pH levels are selected and replicate experiments are to be performed at these points. Informative experiments are typically placed at points with an extreme model output sensitivity.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Optimal experiment design; Parameter estimation; Cardinal values; Cardinal temperature model with inflection point; Cardinal pH model; Predictive microbiology

* Corresponding author.

E-mail address: jan.vanimpe@cit.kuleuven.ac.be (J.F. Van Impe).

1. Introduction

In the context of secondary modelling (mathematical models describing effects of environmental factors on microbial kinetics), factorial experimental designs are often common practice. Full factorial designs refer to experimental plans which encompass all possible combinations of the levels of the factor(s) (i.e., independent variables) under study (Anderson and McLean, 1974). Factorial designs have the advantage of studying all factors simultaneously as to obtain information on both main effects and interactions among the various factors. It is, however, not always evident to identify levels and treatment combinations which are most relevant.

If prior knowledge on the microbial dynamics is lacking and (thus) no model structure is available, it is wise to choose values of the independent variable(s) equally distributed within the region of interest (Box and Draper, 1971; Davies, 1993). Suffice to say that extrapolation, i.e., making model predictions outside the studied region, is out of the question. The scale of the design should thus include the region within which predictions are to be made (Davies, 1993). If a priori information on the kinetics is available, it is recommendable to space more treatment levels at regions where rapid changes of the dependent variable(s) are expected (Davies, 1993; Walker and Jones, 1993). For example, treatment levels should be more concentrated near the growth boundaries to identify the growth/no growth transition zone (where the growth rates are usually largely affected by small environmental changes) (Walker and Jones, 1993; Whiting, 1997).

Complete factorial designs typically require extensive experimental studies. Sometimes, the number of experiments (or treatments) can be reduced by intelligent confounding of (high-order) interactions which are of minor importance (i.e., fractional factorial designs). Basically, factorial designs assume a polynomial relation (i.e., a response surface model) between the independent and dependent variables which needs to be determined and the coefficients of which are to be estimated. In most statistical handbooks, two- and three-leveled factorial designs are well-explained and indicated as sufficient to investigate main effects and interactions between influencing variables in polynomial relations. Anderson and

McLean (1974), however, note that much more levels are necessary to investigate nonlinear trends correctly.

In case a secondary model structure has been properly characterised, the remaining problem is the estimation of the model parameters on the basis of experimental data. The model structure often complies with kinetics of a large set of food-related microorganisms but model parameters need to be re-estimated for each microorganism (in different medium conditions). In many experimental studies, (full) factorial designs in which treatment levels (or combinations) are selected arbitrary, e.g., evenly distributed within the region of interest, are performed (e.g., Uljas et al., 2001; Bharathi et al., 2001). This experimental approach may be suitable for the determination of the model structure (see above) but is not always the best choice when unique and accurate parameter estimation for a known (*non-linear*) model structure is aimed at.

Given the model structure is correct, an optimal experimental plan can be computed by using the methodology of optimal experiment design for parameter estimation (OED/PE) (e.g., Walter and Pronzato, 1997; Bernaerts et al., 2002). When static experiments are considered, issue involved in optimal experiment design is the optimal treatment placement. Work on optimal *dynamic* experiment design for two-parameter estimation problems in predictive microbiology has been presented in Bernaerts et al. (2002) and Versyck et al. (1999).

This contribution presents OED/PE results for the Cardinal Temperature Model with Inflection point (CTMI) and the Cardinal pH Model (CPM). More precisely, the accurate estimation of cardinal temperatures and pHs is aimed at by optimal selection of temperatures and pH values applied in static growth experiments. Cardinal models were first developed to describe (predict) growth kinetics but are also applicable for the determination of growth/no growth boundaries (Le Marc et al., 2002).

Cardinal models are *nonlinear* with respect to the model parameters, implying that optimal experiment designs depend on the model parameters itself. Some initial guess for the parameters is required. This parameter dependence of the design imposes a sequential optimal design scheme to be followed (see text). Remark that the presented methodology

equally well applies to polynomial linear model structures. In such case, designs will be independent of the unknown parameters.

This paper includes and elaborates on the results reported in Bernaerts et al. (2003a,b) and is organised as follows. First, the model structures and the tech-

nique of optimal experiment design for parameter estimation are presented (Section 2). Preceded by an examination of the model output sensitivities, optimal design results are reported and discussed for both models in Section 3. Conclusions and critical comments are formulated in Section 4.

2. Material and methods

2.1. Mathematical models

The family of cardinal models (CM_n) is derived from the following general formulation (as presented by Rosso, 1995):

$$\mu_{\max}(x) = \mu_{\text{opt}}\gamma(x) \quad (1)$$

with

$$\gamma(x) = \begin{cases} x < p_{\min}, & \gamma(x) = 0 \\ p_{\min} \leq x \leq p_{\max}, & \\ & \frac{(x - p_{\min})^n(x - p_{\max})}{(p_{\text{opt}} - p_{\min})^{n-1}[(p_{\text{opt}} - p_{\min})(x - p_{\text{opt}}) - (p_{\text{opt}} - p_{\max})((n-1)p_{\text{opt}} + p_{\min} - nx)]} \\ x > p_{\max}, & \gamma(x) = 0 \end{cases}$$

with μ_{\max} the maximum specific growth rate [h^{-1}], x the independent variable (e.g., temperature, pH), p_{\min} , p_{opt} and p_{\max} the cardinal values associated with x , μ_{opt} the maximum specific growth rate corresponding with p_{opt} , and n a shape parameter. This shape parameter determines the final model structure, the structural model properties of which may differ reasonably between cardinal models (see below).

The Cardinal Temperature Model with Inflection point relates the maximum specific growth rate with temperature (T [$^{\circ}\text{C}$]) and can be obtained by selecting $n=2$ in Eq. (1) (Rosso et al., 1995):

$$\gamma(T) = \frac{(T - T_{\min})^2(T - T_{\max})}{(T_{\text{opt}} - T_{\min})[(T_{\text{opt}} - T_{\min})(T - T_{\text{opt}}) - (T_{\text{opt}} - T_{\max})(T_{\text{opt}} + T_{\min} - 2T)]}. \quad (2)$$

The biologically meaningful parameters in Eq. (2) are the *cardinal temperatures*, i.e., T_{\min} , T_{opt} and T_{\max} ($^{\circ}\text{C}$), the minimum, optimum and maximum temperature for growth, respectively.

Selecting $n=1$, Eq. (1) reduces to the Cardinal pH Model describing the effect of pH on the maximum specific growth rate, i.e.,

$$\gamma(pH) = \frac{(pH - pH_{\max})(pH - pH_{\min})}{(pH - pH_{\min})(pH - pH_{\max}) - (pH - pH_{\text{opt}})^2} \quad (3)$$

with pH_{\min} , pH_{opt} , and pH_{\max} the cardinal pH values.

Cardinal models incorporating multiple factors are constructed by multiplication of factors typical for individual effects, e.g., $\mu_{\max}=\mu_{\text{opt}}\gamma(T)\gamma(pH)$. Reported cardinal models include effects of temperature, pH, and/or organic

acids on microbial growth (Rosso, 1995; Rosso et al., 1995; Le Marc et al., 2002), and the influence of water activity on fungal growth (Rosso and Robinson, 2001; Sautour et al., 2001). Cardinal models are parsimonious, and the applicability of CMs for T and pH is tested on a wide range of data in Rosso (1995). The sole exception known to date are strains of *Listeria monocytogenes* (Bajard et al., 1996). The maximum specific growth rate of *L. monocytogenes* strains shows a biphasic behaviour within the suboptimal temperature range that can be described by a cardinal temperature model embedding two additional model parameters, as proposed by Le Marc et al. (2002).

2.2. Optimal experiment design for parameter estimation

Optimal experiment designs are obtained by the minimisation/maximisation of a scalar function of the Fisher information matrix \mathbf{F} (Walter and Pronzato, 1997). Under certain presumptions, the inverse of \mathbf{F} yields a lower bound for the parameter estimation variance–covariance matrix (see Walter and Pronzato, 1997), or in other words, \mathbf{F} yields a quantification of the parameter estimation accuracy related with a particular experimental plan. For the present case study, the Fisher information matrix \mathbf{F} can be expressed as follows:

$$\mathbf{F} = \frac{1}{\sigma_{\mu_{\max}}^2} \sum_{i=1}^{n_x} \left[\frac{\partial \mu_{\max}}{\partial \mathbf{p}}(x_i) \right]_{\mathbf{p}=\mathbf{p}^{\circ}} \left[\frac{\partial \mu_{\max}}{\partial \mathbf{p}}(x_i) \right]_{\mathbf{p}=\mathbf{p}^{\circ}}^T \quad (4)$$

with $\partial \mu_{\max}/\partial \mathbf{p}$ the vector of model output sensitivities, $\sigma_{\mu_{\max}}^2$ the measurement error variance on μ_{\max} , \mathbf{p} the vector of model parameters, \mathbf{p}° the vector of nominal model parameters, and n_x the number of model outputs (i.e., treatment levels included in the experimental design). Each model output sensitivity function ($\partial \mu_{\max}/\partial p_j$) expresses the effect of a small parameter deviation on the course of the model output, and plays a key role in accurate parameter estimation. Nominal parameters are an initial estimate for the unknown model parameters. Dealing with a nonlinear model structure, the choice of the nominal values does affect the final design because model output sensitivities on itself depend on one or more model parameters. An iterative design scheme should be followed to converge to the true process parameters (Bernaerts et al., 2001).

As an example, optimal experiment designs are computed for *Escherichia coli*, the nominal model parameters of which are listed in Fig. 1. Observe that μ_{\max} —which is here considered as the measured output—actually results from a complete growth curve described by a primary model. Hence, the variance on μ_{\max} reflects primary estimation errors as well as biological variability. Here, $\sigma_{\mu_{\max}(T)}^2$ is set equal to 2.229×10^{-2} and $\sigma_{\mu_{\max}(\text{pH})}^2$ to 2.834×10^{-2} (Rosso, 1995).

A remark needs to be made at this point. The above-defined Fisher information matrix (Eq. (4)) prerequisites a constant measurement error variance. On the contrary, it is often highlighted that the measurement error variance on μ_{\max} increases with increasing temperature, and a square root transformation stabilising the error variance is recommended before parameter estimation (Zwietering et al., 1994); One must however note that growth kinetics predictions are derived from the untransformed (secondary) model equation inserted into a (primary) growth model. Hence, parameter uncertainties related with the identification of the CTMI (or CPM) on μ_{\max} versus T (or pH) are most relevant. A square root transformation affects the model output sensitivities and, consequently, the experimental designs presented in this publication are not applicable to the square root transformation of the considered cardinal models (see text for further details).

Among existing criteria, D -optimal design and E -optimal design are selected for this study. D -optimal design aims at the minimisation of the global parameter estimation uncertainty by maximising the determinant of \mathbf{F} :

$$\max [\det(\mathbf{F}(\mathbf{x}))] \text{ with } x_i \in [x_{\text{low}}, x_{\text{high}}],$$

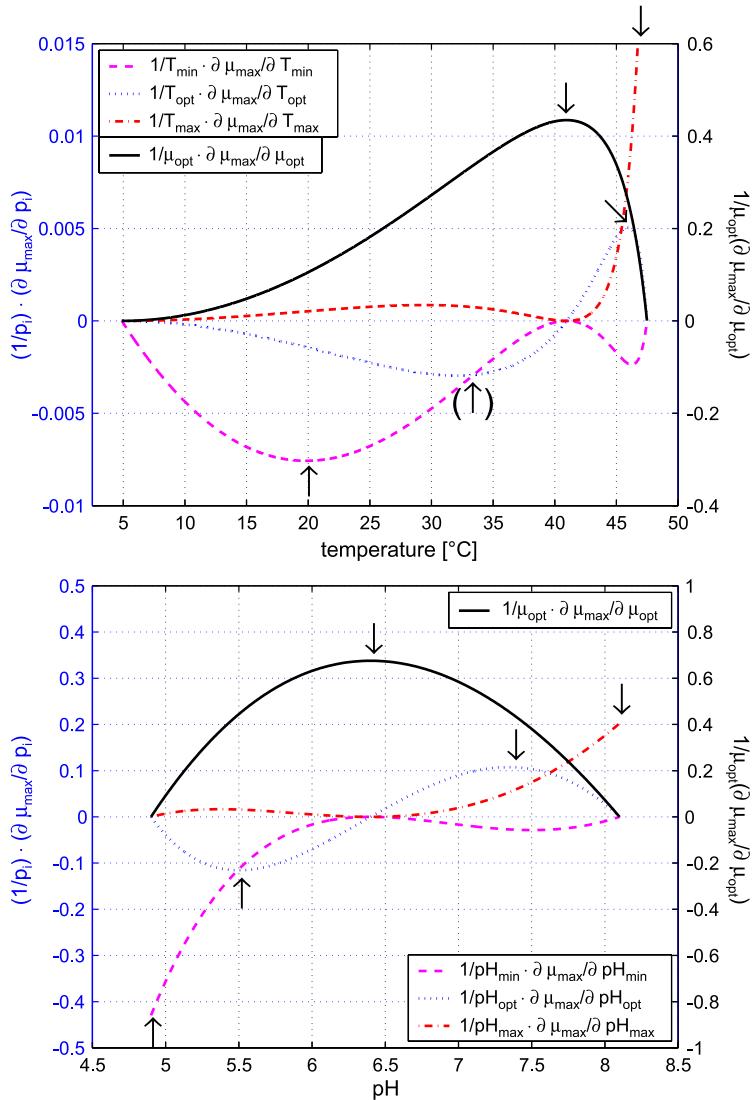


Fig. 1. Model output sensitivities corresponding with the CTMI (top) and the CPM (bottom). Nominal model parameters for *E. coli* are $T_{min} = 4.888^{\circ}\text{C}$, $T_{opt} = 41.28^{\circ}\text{C}$, $T_{max} = 47.48^{\circ}\text{C}$, $\mu_{opt} = 2.301 \text{ h}^{-1}$ and $pH_{min} = 4.9$, $pH_{opt} = 6.4$, $pH_{max} = 8.1$, $\mu_{opt} = 1.48 \text{ h}^{-1}$ (taken from Rosso, 1995). Arrows point at temperatures/pHs with the largest model output sensitivity for each model parameter.

while *E*-optimal design focuses on the minimisation of the *largest* parameter estimation uncertainty by maximising the smallest eigenvalue of \mathbf{F} :

$$\max [\lambda_{\min}(\mathbf{F}(\mathbf{x}))] \text{ with } x_i \in [x_{\text{low}}, x_{\text{high}}]$$

(see e.g., Walter and Pronzato, 1997). Optimisations are performed using the NAG routine E04UCF (The Numerical Algorithms Group Ltd) with Fortran. During optimisation, the number of treatments n_x is a priori fixed and control inputs are confined to the interval $[x_{\text{low}}, x_{\text{high}}]$. As the optimisation problem is

multimodal (i.e., multiple *local* optimal solutions can be found), the optimisation algorithm is sufficiently often initialised using the NAG pseudo-random generator G05FAF. Details on the choice of the lower and higher boundary, x_{low} and x_{high} , are given further in the text.

3. Results and discussion

3.1. Model output sensitivities

Looking at the formulation of the Fisher information matrix, one can derive that: when aiming for the highest parameter estimation accuracy, there is the (obvious) need to minimise the measurement error variance (here $\sigma_{\mu_{\max}}^2$) but, in addition, experimental inputs must be selected at positions with a high model output sensitivity. The former is the task of the laboratory technician, the latter is the task of the modeller. Extreme model output sensitivities indicate points at which small changes in one or more model parameters significantly affect the (simulated) model output, or vice versa, measured outputs at these extremes have the largest effect on the parameter estimates and the associated parameter estimation variance. Hence, a model output sensitivity analysis provides a *first* idea of optimal experimental treatments.

The model output sensitivities for both models are depicted in Fig. 1. Observe the difference in model output sensitivities between both models. The (a)symmetry of the model structures is reflected in the sensitivities with respect to the model parameters. In particular, the model structures behave differently at the level of the minimum cardinal value, namely, T_{\min} and pH_{\min} . This will become clearer within the optimised experiment designs.

On the basis of the presented sensitivity functions, the following *preliminary* direction with regard to experiment planning can be formulated: *select temperature/pH levels at/near the maxima of the model output sensitivities*. These extremes are indicated by arrows in Fig. 1. Measurements at these positions have the largest influence on the parameter values. Experimental uncertainties or errors at these points shall have major (adverse) effects on the parameter values during parameter estimation.

The following should be remarked (maybe somewhat opposed to intuition). (i) T_{\min} is most affected by

$\mu_{\max}(T)$ measurements at ambient growth temperatures (see $\max(\partial\mu_{\max}/\partial T_{\min})$). Moreover, based on the formula for prediction error variances (see, e.g., Van Impe et al., 2001), it can be shown that the prediction for $\mu_{\max}(T_{\min})$ is theoretically perfect which is not realistic (for more details, see Bernaerts et al., 2003a,b). (ii) Temperature and pH values at exactly T_{opt} or pH_{opt} yield no information on the cardinal values whatsoever (sensitivities are equal to zero). This means that T_{opt} and pH_{opt} cannot be accurately determined when placing experiments at exactly the optimum growth temperature or pH but experiments must be performed with T or pH values around T_{opt} or pH_{opt} (see $\max(\partial\mu_{\max}/\partial T_{\text{opt}})$ and $\max(\partial\mu_{\max}/\partial pH_{\text{opt}})$ in Fig. 1).

As stated previously, a square root transformation of the model affects the course of the output sensitivities. In particular for the CTMI, a square root transformation alters the model output sensitivities. This is depicted in Fig. 2. The major effect is associated with the model output sensitivity with respect to T_{\min} . In contrast to the results for the original model structure, $\partial\mu_{\max}/\partial T_{\min}$ reaches an extremum at T_{\min}° itself. This implies that temperature treatments close to T_{\min} are informative and contribute to the accurate estimation of the minimum temperature for growth. Optimal experiment design will encompass temperature levels at $T=T_{\min}$. Moreover, the prediction error on the estimated maximum specific growth rate at T_{\min} differs from zero.

3.2. Optimal experiment designs for CTMI

D- and *E*-optimal experiment designs have been computed. All results are summarised in Figs. 3 and 4.

3.2.1. Unconstrained design

Let us first consider the so-called unconstrained designs. Temperature treatments are allowed within the assumed growth temperature region, i.e., $[T_{\min}^\circ, T_{\max}^\circ]$, during optimisation. Minimum four T levels are required for parameter estimation (less outputs

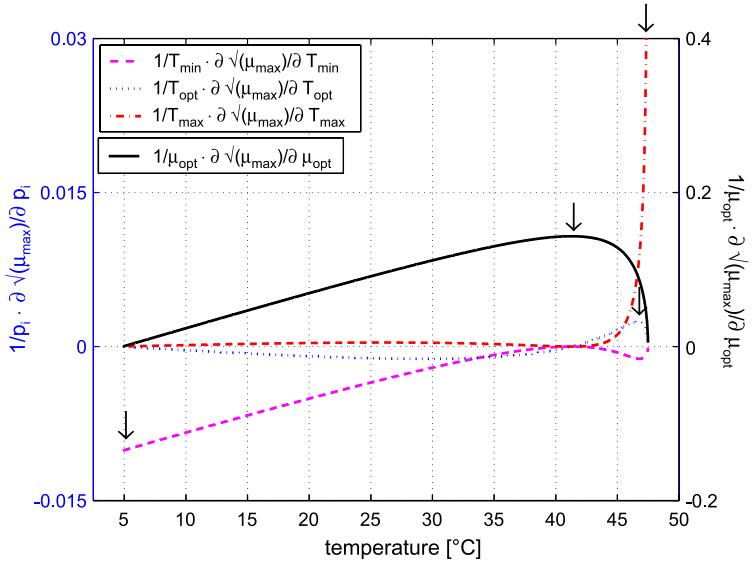


Fig. 2. Model output sensitivities corresponding with the square root transformation of the CTMI using nominal parameters as listed in Fig. 1. Arrows point at temperatures with the largest model output sensitivity for each model parameter.

yield a singular \mathbf{F}). When including more treatments, replicated experiments are preferable instead of spreading these experiments within the growth temperature range.

Note at this point that it is not possible to determine the measurement error variance on the basis of four $\mu_{\max}(T_i)$ data points. The measurement

error variance is commonly approximated by the mean sum of squared errors which is defined as the sum of squared errors divided by the difference between the number of data (n_T) and the number of model parameters (n_p), and thus becomes infinitely large given four data points and four unknown model parameters. Consequently, the derived para-

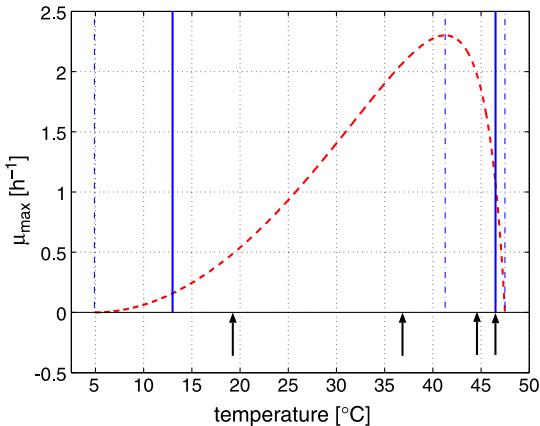


Fig. 3. D - and E -optimal design results for the CTMI with $T_i \in [T_{\min}, T_{\max}]$ (top plots), $T_i \in [13^{\circ}\text{C}, 46.5^{\circ}\text{C}]$ (middle plots, full lines), or $T_i \in [13^{\circ}\text{C}, 45^{\circ}\text{C}]$ (lower plots, full lines). Left column: informative temperature levels resulting from D -optimal design (indicated by arrows). In case $n_T > 4$, additional temperature treatments are evenly spread among these four temperature levels. Determinant values are not affected by the selected design. The dashed line depicts the nominal model output. Right column: optimum placement of temperature levels according to an E -optimal design. Indices indicate the number of repeated experiments. Dash-dotted lines mark the nominal parameters. Optimum objective function values corresponding with D - and E -optimal designs with increasing number of temperature treatments are shown in Fig. 4.

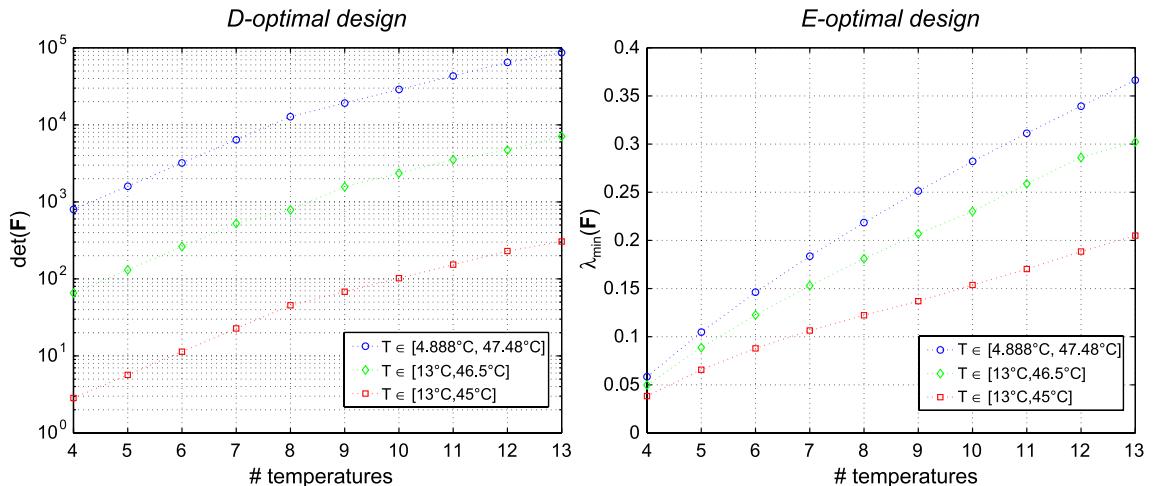


Fig. 4. Optimum objective function values corresponding with *D*- and *E*-optimal designs with increasing number of temperature treatments depicted in Fig. 3.

meter estimation variances will be infinitely large as well.

On the one hand, irrespective of the number of T -levels included, *D*-optimal design always yields the same four informative temperature levels (see Fig. 3, top left). As the *D*-criterion focuses on the global parameter estimation uncertainty, additional temperature treatments ($n_T > 4$) are evenly spread among these four informative temperature levels. For a given number of temperatures, the exact order does not matter as the determinant value of \mathbf{F} always takes the same value irrespective of the placement of replicates. The associated standard deviation on the parameter estimates, however, shall vary between different designs with the same number of temperature levels (for more details, see Bernaerts et al., 2003a,b; Gysemans et al., 2004).

On the other hand, the *E*-criterion yields designs with replications near $\max(\partial\mu_{\max}/\partial T_{\min})$ (Fig. 3, top right). This shows that the largest parameter estimation uncertainty can be attributed to T_{\min} .

(*D*- and *E*-)Optimal designs are largely in agreement with the aforementioned directive concerning experiment design based on the model output sensitivity functions. Yet, inputs are not exactly positioned at the maxima of the sensitivity functions due to the model structure nonlinearity.

Finally, including more temperature treatments yields obviously a higher information content (as expressed as either $\det(\mathbf{F})$ or $\lambda_{\min}(\mathbf{F})$) (see Fig. 4).

3.2.2. Constrained design

Irrespective of the design criterion (and the number of temperature treatments), experiments at the maximum temperature for growth are involved. In practical terms, these experiments at the growth boundary are not workable. One is balancing on the edge between growth and inactivation. Supposing an erroneous nominal value for T_{\max} , namely, T_{\max}° larger than the true maximum growth temperature, an inactivation experiment will be performed which is not useable for cardinal values estimation. Imposing an upper boundary on the temperature, $T_{\text{high}} < T_{\max}^{\circ}$, is advisable. The same reasoning can be formulated for experiments at $T = T_{\min}$. When the true minimum temperature for growth is (much) larger than the selected nominal parameter T_{\min}° , little information can be extracted from experiments at T_{\min}° (showing no growth). A lower boundary, $T_{\text{low}} > T_{\min}^{\circ}$, must ensure that experiments show growth.

Constrained optimal design, i.e., imposing an upper and lower temperature boundary during input optimisation, is thus needed. Fig. 3 shows two temperature-bounded optimal designs. For example, temperatures are confined within the interval [13 °C,

46.5 °C] and [13 °C, 45 °C]. T_{low} is chosen high enough to guarantee significant growth measurements with respect to experimental noise and T_{high} is taken low enough such that inactivation or no (significant) growth is unlikely. The lowered upper boundary mainly alters the design results. The treatments are *shifted* to the left as compared to the unconstrained designs (see Fig. 3). No effect of the lower boundary can be observed in this example because the lower boundary does not interfere with optimal temperature levels from the unconstrained design.

Imposing temperature constraints implies making sacrifices on the final information content as shown in Fig. 4 (and thus the parameter estimation accuracy). In this particular case, especially the uncertainty on T_{max} increases which has reasonable effects on the prediction error variance (for more details, see Gysemans et al., 2004).

3.3. Optimal experiment designs for CPM

In the same manner as for the CTMI, a summary of all CPM related designs is presented in Figs. 5 and 6. To some extent, results are comparable with the optimal design for the CTMI. Some differences are highlighted.

3.3.1. Unconstrained design

Resulting designs encompass both pH experiments at the lower and upper boundary of the assumed growth pH region, i.e., $[\text{pH}_{\text{min}}, \text{pH}_{\text{max}}]$. This can be easily understood because of the extremely high model output sensitivity for pH_{min} and pH_{max} at these points. D -optimal designs again reduce to four informative pH levels with replicates evenly distributed among these points. Like D -optimal design, E -design converges more or less to a set of four informative pH levels but with a preferred replication scheme with increasing number of pH experiments included (the single exception is $n_{\text{pH}}=5$, see figure).

3.3.2. Constrained design

Imposing both a lower and an upper constraint on pH is necessary. As an example, $[\text{pH}_{\text{low}}, \text{pH}_{\text{high}}]$ is set equal to [5.2, 7.8] and [5.4, 7.6], respectively. In the latter case, these bounds set rather large safety

margins and will have a great effect on the information content. As compared to above-presented designs for the CTMI, the constraints have a more pronounced effect on the information content (and thus the parameter estimation uncertainty) because both the lower and upper boundary exclude a region of high model output sensitivity. The information loss is especially remarkable at the level of $\lambda_{\min}(\mathbf{F})$.

3.4. Comparison with equidistant designs

A complete quantitative illustration of the *benefit* of optimal designs versus a (more classical) equidistant design in terms of parameter estimation quality (or prediction quality) is out of the scope of this publication; however, by means of example, the effect of the treatment placement on the parameter estimation uncertainty is illustrated for a constrained D -optimal, E -optimal and equidistant design including eight treatment levels. The standard deviation on the parameter estimates can be computed on the basis of the inverse of the Fisher information matrix (see Section 2.2). The measurement error variances are taken equal to the values given in Section 2.2 such that the positioning of the treatment levels is the sole factor determining the final parameter estimation quality.

Table 1 summarises the standard deviations on the parameter estimates corresponding to the different designs for both the CTMI and the CPM. For the CTMI, both optimised experiment designs yield the lowest standard deviations on the model parameters as compared to the equidistant design. Clearly, a D -optimal design focuses on the overall uncertainty while an E -optimal design puts the focus on the most uncertain model parameter, i.e., T_{min} . Note that the relatively large standard deviations associated with T_{min} are due to an inherent model structure property and, in this case also, the large distance between T_{min} and T_{opt} . For the CPM, mainly the parameter estimation uncertainty on pH_{min} and pH_{max} are reduced; hereby expressing that the highest uncertainty is associated with these parameters. The contribution of $s_{\mu_{\text{opt}}}$ to the overall parameter error is apparently redundant (i.e., not the focus of D -optimal nor E -optimal design). Distributing the pH levels equidistantly around pH_{opt} results into

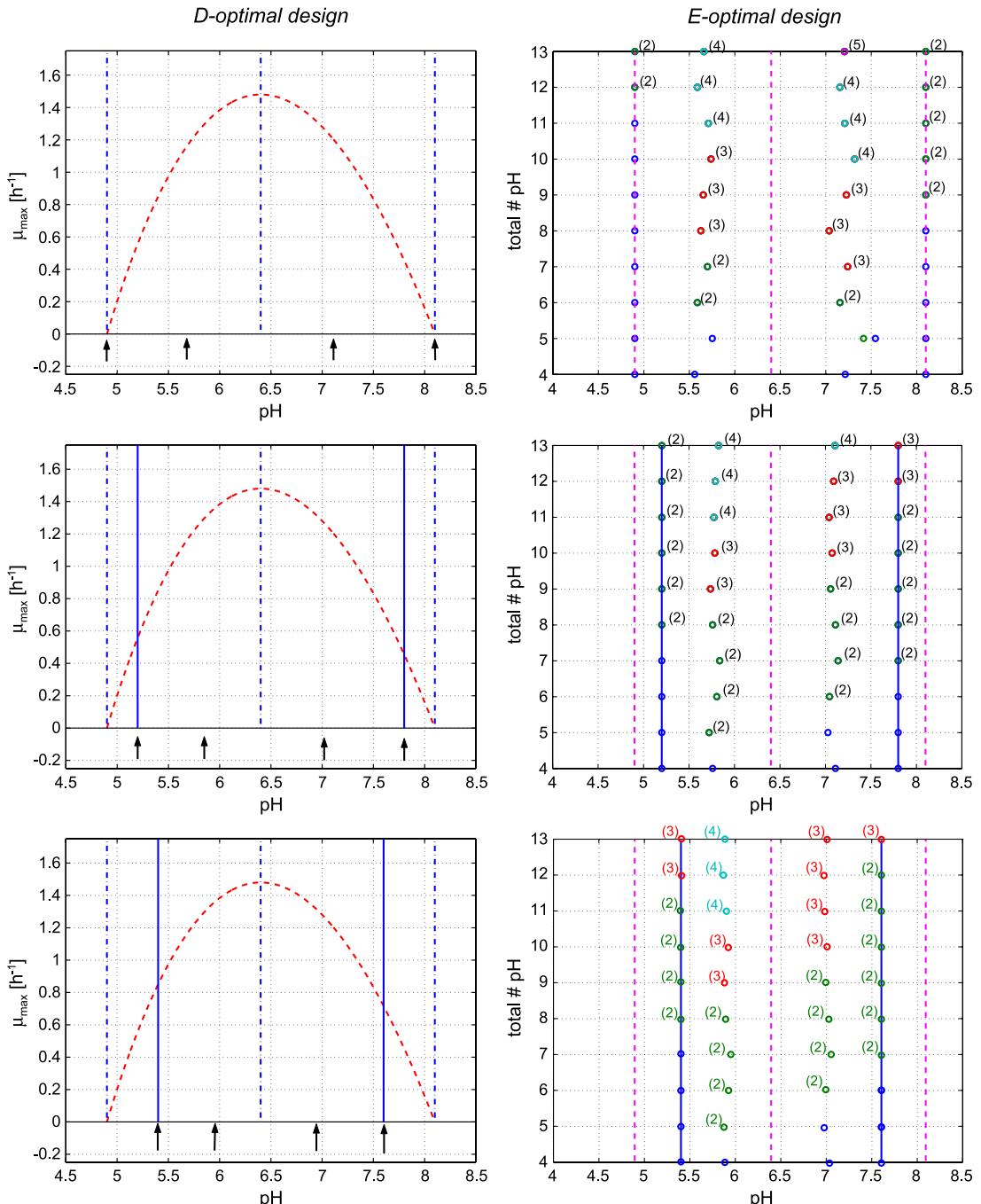


Fig. 5. *D*- and *E*-optimal design results for the CPM with $pH_i \in [pH_{\min}^\circ, pH_{\max}^\circ]$ (top plots), $pH_i \in [5.2, 7.8]$ (middle plots, full lines), or with $pH_i \in [5.4, 7.6]$ (lower plots, full lines). Left column: informative pH levels resulting from *D*-optimal design (indicated by arrows). In case $n_{pH} > 4$, additional pH treatments are evenly spread among these four pH levels. Determinant values are not affected by the selected design. The dashed line depicts the nominal model output. Right column: optimum placement of pH levels according to an *E*-optimal design. Indices indicate the number of repeated experiments. Dash-dotted lines mark the nominal parameters. Optimum objective function values corresponding with *D*- and *E*-optimal designs with increasing number of temperature treatments are shown in Fig. 6.

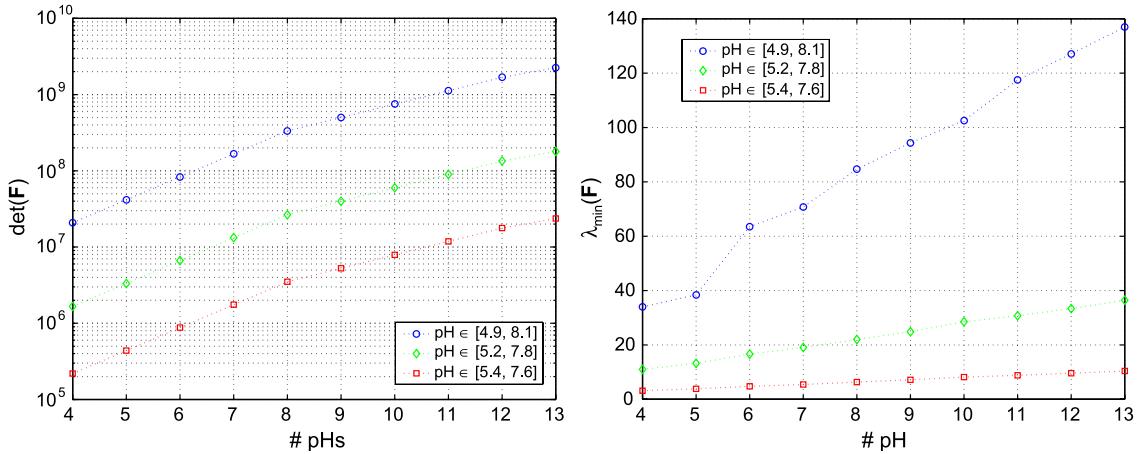


Fig. 6. Optimum objective function values corresponding with *D*- and *E*-optimal designs with increasing number of pH treatments depicted in Fig. 5.

the lowest $s_{\mu_{\text{opt}}}$ (as could be expected from the shape of $\partial \mu_{\text{max}}/\partial \mu_{\text{opt}}$).

4. Conclusions

When the secondary model structure is known, (constrained) optimal experiment design for parameter estimation provides experimental settings for accurate parameter estimation. Opposed to the frequently applied equidistant (or arbitrary) treatment placement, it is shown that replicated experiments at *four* informative inputs are most favourable for estimation of the cardinal model parameters. Considerations on the model output sensitivities give a first indication for optimal experiment design and may also unveil hidden model features. In this particular case, it could be inferred that the prediction error variance on μ_{max} estimates will be zero when using the cardinal temperature model with inflection point (without any transformation).

Let us repeat that optimal designs including *just* four temperature or pH levels as depicted in Figs. 3 and 5 are not recommended. In such case, the parameter uncertainty computed on the basis of the mean sum of squared errors (approximating the measurement error variance) will be infinitely large and erroneous parameter estimates almost inherently emerge from experimental errors. As commonly known, it is strongly advised to perform *sufficient*

experiments, i.e., $n_x \gg n_p$ (with n_p the number of model parameters) because the parameter estimation error variance not only depends on the positioning of the treatment levels but also on the number of experimental data points (as illustrated in Figs. 4 and 5). Furthermore, *measurements* of the maximum specific growth actually evolve as an estimate from an individual growth curve and proper experiment design (maximising the confidence on μ_{max} esti-

Table 1

Effect of the treatment placement on the parameter estimation quality (expressed by means of the expected standard deviation on the parameter estimation)

CTMI	Equidistant design	Constrained <i>D</i> -optimal	Constrained <i>E</i> -optimal
$s_{T_{\min}}$	3.447	3.3067	2.311
$s_{T_{\text{opt}}}$	5.831×10^{-1}	3.253×10^{-1}	4.883×10^{-1}
$s_{T_{\max}}$	1.226	6.368×10^{-1}	9.088×10^{-1}
$s_{\mu_{\text{opt}}}$	1.312×10^{-1}	9.598×10^{-2}	1.433×10^{-1}
CPM	Equidistant design	Constrained <i>D</i> -optimal	Constrained <i>E</i> -optimal
$s_{\mu_{\text{min}}}$	1.882×10^{-1}	1.466×10^{-1}	1.482×10^{-1}
$s_{\mu_{\text{opt}}}$	1.610×10^{-1}	1.572×10^{-1}	1.491×10^{-1}
$s_{\mu_{\text{max}}}$	1.923×10^{-1}	1.480×10^{-1}	1.500×10^{-1}
$s_{\mu_{\text{opt}}}$	9.075×10^{-2}	1.073×10^{-1}	1.195×10^{-1}

Each design encompasses eight temperature or pH levels which are confined within the interval [13 °C, 46.5 °C] and [5.2, 7.8], respectively. *D*- and *E*-optimal designs are shown in Figs. 3 and 5.

mates) at this level is therefore also required. By means of Monte Carlo simulations, Poschet et al. (2004) have shown that the optimal spacing of cell density measurements over time for estimation of primary parameters of the Baranyi and Roberts (1994) growth model encompasses frequent sampling at the start of the experiment (to determine the initial microbial load), within the transition zone between lag and exponential growth, within the transition zone between exponential growth and the stationary phase, and at the end of the experiment (to determine the asymptotic level).

Unresolved issues at this moment read as follows.

Present design results are under the condition that the nominal model parameters are exactly known, which is obviously not the case in practice. A first preventive first measure to account for this uncertainty—within a single design step—is to impose constraints on the admissible temperature/pH region during optimisation. Hereby, nongrowth and inactivation experiments falling outside the validity region (i.e., growth region) of the cardinal values model are avoided. However, the robustness of optimal designs with regard to erroneous (or uncertain) nominal cardinal parameters must be further investigated. Generally, the effect of wrongly selected nominal model parameters is resolved by applying an iterative design scheme (e.g., Bernaerts et al., 2001). Starting with an initial guess for the nominal model parameters, the parameter values are systematically improved in subsequent optimal experiment design steps. In the first iteration, optimal treatment levels are computed using \mathbf{p}° . This set of experiments yields a first (improved) estimate for the model parameters ($\hat{\mathbf{p}}_1$). In the second iteration, \mathbf{p}° is replaced by $\hat{\mathbf{p}}_1$ and treatment levels are again optimally selected. Hence, each iteration is based on the parameter estimates evolving from the preceding optimal design round. Rapid convergence of this iterative design scheme towards the true model parameters can be expected (results not shown) but sufficient experiments should be performed in order to obtain acceptable parameter estimation error variances. Complementary to this iterative design, replicates at exactly the same four informative temperature or pH levels could be replaced by some (modest) spreading of experiments around the main informative treatment levels.

As the interpretation of the information content (expressed by $\det(\mathbf{F})$ or $\lambda_{\min}(\mathbf{F})$) associated with the presented optimal designs in (more classical) terms of parameter estimation uncertainty is not directly obvious, results in Section 3.4 are added by means of example for the effective contribution of optimal experiment designs. A full exploration of the overall advantage of presented designs (both regarding parameter estimation quality and amount of experimental work, including the iterative design) is the subject of current research.

Finally, the issue on the square root transformation of the cardinal values models is summoned. A square root transformation is commonly applied prior to parameter estimation to stabilise the measurement error variance on $\mu_{\max}(x)$ (e.g., Zwietering et al., 1994; Rosso and Robinson, 2001). Presented results, however, consider the untransformed model structure, and are therefore not directly transferable to the transformed model (see infra). In this respect, one may note that growth kinetics predictions are derived from the untransformed (secondary) model equation inserted into a (primary) growth model. Hence, parameter uncertainties related with the identification of the CTMI (or CPM) on μ_{\max} versus T (or pH) are most relevant. Additionally, the benefit of a global identification step (deriving secondary model parameters immediately from cell density data) may be highlighted (see, e.g., Valdramidis et al., 2004). Within such global identification procedure, the uncertainty on μ_{\max} coupled to the primary model parameter estimation step (i.e., $s_{\mu_{\max}(x)}^2$) can be accounted for without the explicit need for replicates.

Acknowledgements

This research is supported by the Research Council of the Katholieke Universiteit Leuven as part of projects OT/99/24 and IDO/00/008, the Belgian Program on Interuniversity Poles of Attraction and the Second Multi-annual Scientific Support Plan for a Sustainable Development Policy, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. Kristel Bernaerts is a Postdoctoral Fellow with the Fund for Scientific Research Flunders (Belgium). The scientific responsibility is assumed by its authors.

References

- Anderson, V.L., McLean, R.A., 1974. Design of Experiments: A Realistic Approach. Marcel Dekker, New York.
- Bajard, S., Rosso, L., Fardel, G., Flandrois, J.P., 1996. The particular behaviour of *Listeria monocytogenes* under sub-optimal conditions. International Journal of Food Microbiology 29 (2–3), 201–211.
- Baranyi, J., Roberts, T.A., 1994. A dynamic approach to predicting bacterial growth in food. International Journal of Food Microbiology 23, 277–294.
- Bernaerts, K., Servaes, R.D., Kooyman, S., Van Impe, J.F., 2001. Iterative optimal experiment design for estimation of microbial growth kinetics as function of temperature. In: Dochain, D., Perrier, M. (Eds.), 8th International Conference on Computer Applications in Biotechnology (CAB8), pp. 19–24.
- Bernaerts, K., Servaes, R.D., Kooyman, S., Versyck, K.J., Van Impe, J.F., 2002. Optimal temperature input design for estimation of the Ratkowsky square root model parameters: parameter accuracy and model validity restrictions. International Journal of Food Microbiology 73 (2–3), 145–157.
- Bernaerts, K., Nhan Minh, T., Van Impe, J.F., 2003a. Critical evaluation of a nonlinear model from predictive microbiology using sensitivity analysis and optimal experimental design. In: Troch, I., Breitenecker, F. (Eds.), Proceedings 4th MathMod Vienna, pp. 1274–1280. ARGESIM Report No. 24, Technical University Vienna.
- Bernaerts, K., Gysemans, K., Nhan Minh, T., Van Impe, J.F., 2003b. Optimal experiment design for cardinal values estimation: instructions for data collection. In: Van Impe, J.F.M., Geeraerd, A.H., Leguérinel, I., Mafart, P. (Eds.), Predictive Modelling in Foods—Conference Proceedings. KULEuven/BioTeC, Belgium, ISBN: 90-5682-400-7, pp. 111–113.
- Bharathi, S., Ramesh, M.N., Varadaraj, M.C., 2001. Predicting the behavioural pattern of *Escherichia coli* in minimally processed vegetables. Food Control 12, 275–284.
- Box, M., Draper, N.R., 1971. Factorial designs, the $-X'X-$ criterion, and some related matters. Technometrics 13 (4), 731–742.
- Davies, K.W., 1993. Design of experiments for predictive microbial modeling. Journal of Industrial Microbiology 12, 296–300.
- Gysemans, K.P.M., Bernaerts, K., Van Impe, J.F., 2004. Constrained input optimization for optimal parameter estimation of a predictive biokinetic model. Proceedings of the 9th International Symposium on Computer Applications in Biotechnology (CAB9, Nancy (France), March 28–31, 2004), 6 p.
- Le Marc, Y., Huchet, V., Bourgeois, C.M., Guyonnet, J.P., Mafart, P., Thuault, D., 2002. Modelling the growth kinetics of *Listeria* as a function of temperature, pH and organic acid concentration. International Journal of Food Microbiology 73 (2–3), 219–237.
- Poschet, F., Bernaerts, K., Geeraerd, A.H., Scheerlinck, N., Nicolaï, B.M., Van Impe, J.F., 2004. Sensitivity analysis of microbial growth parameter distributions with respect to data quality and quantity by using Monte Carlo analysis. Mathematics and Computers in Simulation 65 (3), 231–243.
- Rosso, L., 1995. Modélisation et microbiologie prévisionnelle: Elaboration d'un nouvel outil pour l'agro-alimentaire. PhD thesis, Université Claude Bernard-Lyon 1. Villeurbanne Cedex (France).
- Rosso, L., Robinson, T.P., 2001. A cardinal model to describe the effect of water activity on the growth of moulds. International Journal of Food Microbiology 63, 265–273.
- Rosso, L., Lobry, J.R., Bajard, S., Flandrois, J.P., 1995. Convenient model to describe the combined effects of temperature and pH on microbial growth. Applied and Environmental Microbiology 61, 610–616.
- Sautour, M., Dantigny, P., Divies, C., Bensoussan, M., 2001. A temperature-type model for describing the relationship between fungal growth and water activity. International Journal of Food Microbiology 67, 63–69.
- Uljas, H.E., Schaffner, D.W., Duffy, S., Zhao, L., Ingham, S.C., 2001. Modeling of combined processing steps for reducing *Escherichia coli* O157:H7 populations in apple cider. Applied and Environmental Microbiology 67 (1), 133–141.
- Valdramidis, V.P., Geeraerd, A.H., Bernaerts, K., Devlieghere, F., Debevere, J., Van Impe, J.F., 2004. Accurate modelling of non-loglinear survivor 22 curves (in press).
- Van Impe, J.F., Bernaerts, K., Geeraerd, A.H., Poschet, F., Versyck, K.J., 2001. Modelling and prediction in an uncertain environment. In: Tijskens, L.M.M., Hertog, M.L.A.T.M., Nicolaï, B.M. (Eds.), Food Process Modelling, Chapter 8. Woodhead Publishing, Cambridge, UK, pp. 156–179. 496 pp.
- Versyck, K.J., Bernaerts, K., Geeraerd, A.H., Van Impe, J.F., 1999. Introducing optimal experimental design in predictive microbiology: a motivating example. International Journal of Food Microbiology 51 (1), 39–51.
- Walker, S.J., Jones, J.E., 1993. Protocols for data generation for predictive modelling. Journal of Industrial Microbiology 12, 273–276.
- Walter, E., Pronzato, L., 1997. Identification of Parametric Models from Experimental Data. Springer, Masson. 413 pp.
- Whiting, R.C., 1997. Microbial database building: what have we learned? Food Technology 51 (4), 82–86.
- Zwietering, M.H., Cuppers, H.G.A.M., De Wit, J.C., van't Riet, K., 1994. Evaluation of data transformations and validation of a model for the effect of temperature on bacterial growth. Applied and Environmental Microbiology 60 (1), 195–203.

Update

International Journal of Food Microbiology

Volume 110, Issue 1, 1 July 2006, Page 112–113

DOI: <https://doi.org/10.1016/j.ijfoodmicro.2006.02.001>



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

International Journal of Food Microbiology 110 (2006) 112–113

INTERNATIONAL JOURNAL
Food Microbiolog

www.elsevier.com/locate/ijfoodmi

Erratum

Erratum to “Optimal experiment design for cardinal values estimation: guidelines for data collection”

[International Journal of Food Microbiology 100 (2005) 153–165]

K. Bernaerts, K.P.M. Gysemans, T. Nhan Minh, J.F. Van Impe*

BioTeC-Bioprocess Technology and Control, Department of Chemical Engineering, Katholieke Universiteit Leuven, W. de Croylaan 46, B-3001 Leuven, Belgium

The publisher regrets that in the original printing of this article Fig. 3 is incorrect.
The figure is now reproduced correctly below.

DOI of original article: [10.1016/j.ijfoodmicro.2004.10.012](https://doi.org/10.1016/j.ijfoodmicro.2004.10.012).

* Corresponding author.

E-mail addresses: kristel.bernaerts@cit.kuleuven.be (K. Bernaerts), jan.vanimpe@cit.kuleuven.be (J.F. Van Impe).

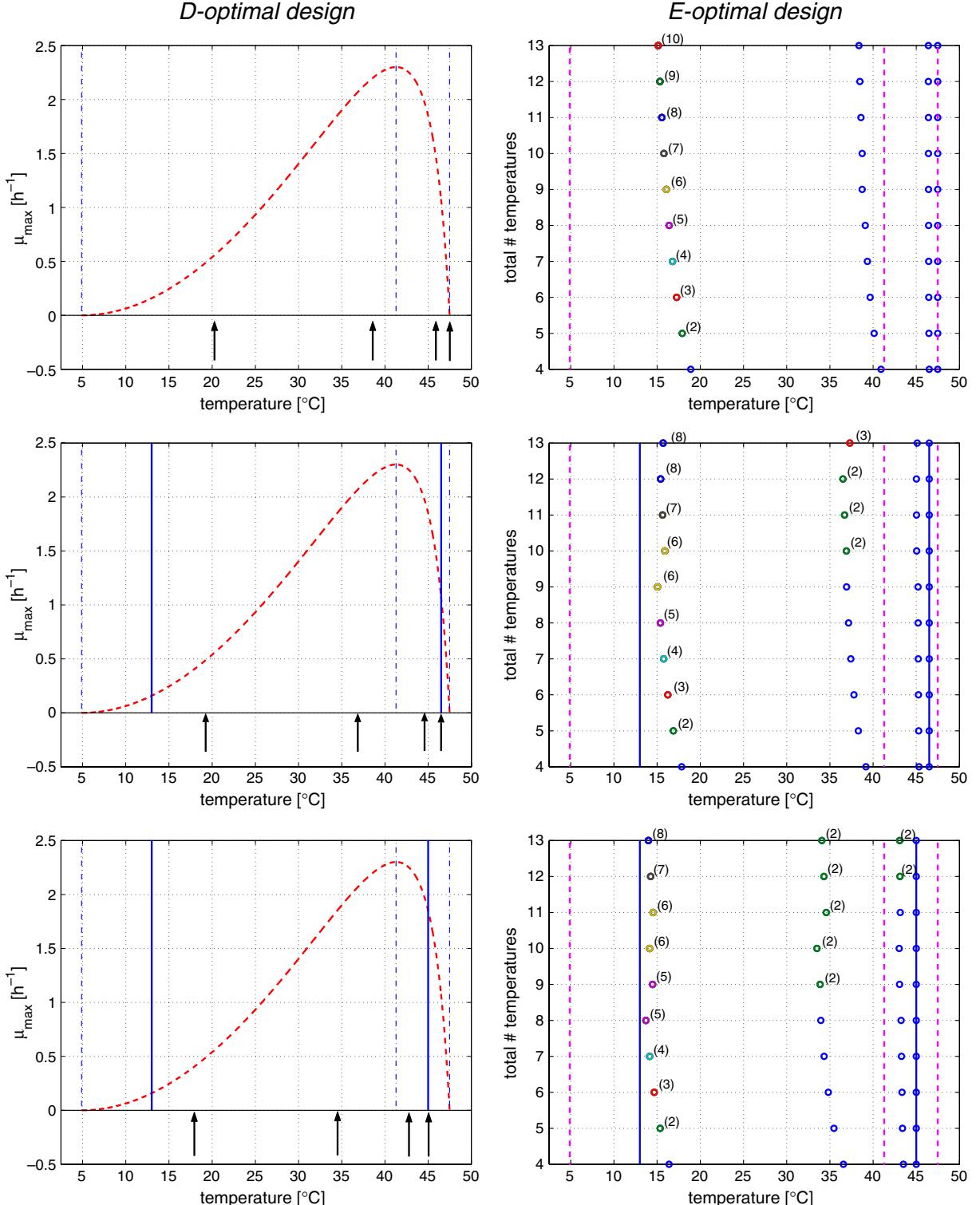


Fig. 3. *D*- and *E*-optimal design results for the CTMI with $T_i \in [T_{\min}^0; T_{\max}^0]$ (top plots), $T_i \in [13^\circ\text{C}, 46.5^\circ\text{C}]$ (middle plots, full lines), or $T_i \in [13^\circ\text{C}, 45^\circ\text{C}]$ (bottom plots, full lines). Left column: Informative temperature levels resulting from *D*-optimal design (indicated by arrows). In case $n_T > 4$, additional temperature treatments are evenly spread amongst these four temperature levels. Determinant values are not affected by the selected design. The dashed line depicts the nominal model of the system. Right column: Optimum placement of temperature levels according to an *E*-optimal design. Indices indicate the number of repeated experiments. Dash-dotted lines mark the nominal parameters. Optimum objective function values corresponding with *D*- and *E*-optimal designs with increasing number of temperature treatments are shown in Fig. 4.