A practical guide for optimal designs of experiments in the Monod model

Nikolay Strigul

Princeton University

Department of Ecology and

Evolutionary Biology

Princeton, NJ, USA

email: nstrigul@princeton.edu

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

email: holger.dette@ruhr-uni-bochum.de

FAX: +49 2 34 70 94 559

Viatcheslav B. Melas

St. Petersburg State University

Department of Mathematics

St. Petersburg

Russia

email: v.melas@pobox.spbu.ru

Keywords: Monod model, microbial growth, biodegradation kinetics, optimal experimental design, D-optimality

Abstract

The Monod model is a classical microbiological model much used in microbiology, for example to evaluate biodegradation processes. The model describes microbial growth kinetics in batch culture experiments using three parameters: the maximal specific growth rate, the saturation constant and the yield coefficient. However, identification of these parameter values from experimental data is a challenging problem. Recently, it was shown theoretically that the application of optimal design theory in this model is an efficient method for both parameter value identification and economic use of experimental resources (Dette et al., 2003). The purpose of this paper is to provide this method as a computational "tool" such

1

that it can be used by practitioners-without strong mathematical and statistical background-for the efficient design of experiments in the Monod model. The paper presents careful explanations of the principal theoretical concepts, and a computer program for practical optimal design calculations in Mathematica 5.0 software. In addition, analogous programs in Matlab software will be soon available at www.optimal-design.org.

1 Introduction

The Monod model was suggested by Nobel Laureate F. Monod in 1942 and for more than 60 years has been one of the most frequently used models in microbiology (Monod, 1949; Pirt, 1975; Koch, 1997; Kovarova-Kovar, Egli, 1998). Most models of chemostat growth are based on the Monod equations (Pirt, 1975; Waltman. Hsu, 1995), and numerous models of microbial ecology incorporate Monod growth kinetics (Koch, 1997; Strigul, Kravchenko, 2006). One of the very important practical applications of this model is the evaluation of the biodegradation kinetics of organic pollutants in environmental systems (Blok, 1994; Blok, Struys, 1996). The Monod model describes microbial growth with three parameters:

1) maximal specific growth rate;

2) a saturation constant;

3) a yield coefficient.

In the case of biodegradation kinetics, these parameters can be used as criteria for the biodegradability of organic pollutants (Blok, 1994).
One of the most important problems in practical applications of the Monod model is evaluating the parameters' values from experimental data. It was shown that in many cases, reasonable estimates of the parameters cannot be obtained by a simple application of non-linear least squares estimators (Holmberg, 1982; Saez, Rittmann, 1992). A possible way to circumvent this problem is the application of optimal experimental design techniques (Dette et al., 2005). Optimal design theory is a branch of statistics actively developed since 1960, when basic results by Kiefer and Wolfowitz were published (Kiefer, Wolfowitz, 1960; Kiefer, 1974). In general, the application of optimal experimental designs can be useful in

1) reducing the number of necessary experimental measurements;

2) improving the precision of model parameter value determinations

(Fedorov, 1972). Optimal design theory is well developed for linear regression models, and for several simple non-linear regression models (Dette et al., 2005). Recently, a complete theoretical examination of optimal designs for the Monod model was presented (Dette et al., 2003, 2005), where two different approaches to the problem of designing experiments for the Monod model were investigated , namely local optimal experimental designs (Dette et al., 2003); and maximin optimal experimental designs (Dette et al., 2005).

While optimal designs have numerous advantages, they are not widely used in practice. Indeed, applications of local optimal experimental designs for the different modifications of the Monod model have been recommended in the literature, and supported by demonstrative case studies (Munack, 1991; Vanrolleghem et al., 1995; Versyck et al., 1998; Berkholz et al. 2000; Smets et al., 2002). In particular, this technique was recommended for the design of biodegradation experiments (Merkel et al., 1996). However, there are only a few cases where this technique has been applied in real experimental research as a tool for parameter identification (Dette et al., 2005). We suspect that this gap between statistical theory and experimental practice is caused by the fact that practitioners often do not have the mathematical background or experience in numerical methods to comfortably use these methods. A theoretical method must not only be well described and mathematically investigated, but must also be delivered "as a tool" for practitioners, in order to be widely accepted in practice. It should include full explanations of the practical meaning of all explicit and implicit mathematical assumptions, and a readily available computer program for its applications in real experiments. While least squares estimators for models of non-linear differential equations are incorporated in numerous software products, for instance in the ModelMaker (SB Technology Ltd), sophisticated algorithms for optimal experimental design calculation are not included in widely distributed software products.

The present paper is an attempt to close this gap between applied mathematicians and practitioners, and is addressed to microbiologists, environmental scientists and other experimentalists. Our objective is to present optimal experimental designs for the Monod model "as a tool" which can be easily applied by practitioners. Firstly, we describe the Monod model with its theoretical background, and give general explanations of the application of optimal experimental designs in the Monod model. Secondly, we present a computer program for the numerical calculation of efficient experimental designs, which can be used directly in experimental studies. We try to avoid a technical description here: all mathematical details of local optimal design theory for the Monod model are presented in Dette et al. (2003); additionally, we direct practitioners to the recent review of optimal design applications in microbiology by Dette et al (2005).

2 The Monod model for a simple batch culture and its properties

In a simple homogeneous batch culture it is assumed that the growth conditions are similar for all cells (Pirt, 1975). Traditionally, a typical growth curve is divided into six phases: 1) lag, 2) accelerating, 3) exponential, 4) decelerating, 5) stationary, and 6) declining growth (Monod, 1949). In many cases cultures do not demonstrate this typical growth; also, the experimenter might be interested in only particular aspects of growth: for example, in predictive microbiology the duration of the lag phase is the main parameter characterizing growth inhibition. This has led to models simulating only a few focal phases of growth, for example in predictive microbiology (Dette et al., 2005), and to more sophisticated structured models able to describe typical growth curves in general (Koch, 1997). The role of the experimenter is to select the model that best reflects the characteristics of microbial growth that are his current interest, and to balance a model's complexity and its flexibility.

The Monod model uses a very convenient approximation of the batch growth process, especially when describing biodegradation kinetics (Blok, 1994; Dette et al., 2005); it recognizes only 3 growth phases, i.e. exponential and decelerating growth, and a stationary phase. It is assumed that lag and accelerating and declining growth phases do not exist. This assumption is satisfied in many practical cases. The cells used to inoculate batch cultures are often taken from another actively growing culture, and, therefore, growth starts immediately in exponential phase. However, one should not try to obtain Monod model parameters directly from experimental data if a lag phase is observed. The declining growth phase is usually not a consideration in biodegradation studies, but it can be analyzed separately, using-for example-a negative exponential model (Dette et al., 2005). Another important assumption of the Monod model is that microbial growth is limited solely by the substrate concentration and, therefore, it is very important that experimental conditions satisfy this requirement.

The Monod model suggests that the microbial growth rate $\mu(t)$ and the substrate concentration $s(t)$ at time t are related by the Michaelis-Menten function:

$$\mu(t) = \theta_1 \frac{s(t)}{s(t) + \theta_2}.$$

(1)

The microbial growth rate for the population is then given by the equation:

$$\eta'(t) = \mu(t)\eta(t)$$

(2)

Finally, the model assumes that some constant fraction of the consumed substrate is transformed into microbial biomass:

$$(3) \qquad\qquad s(t) - s_0 = \frac{\eta_0 - \eta(t)}{\theta_3}$$

where η_0 and s_0 are the initial values of microbial biomass and substrate concentration respectively, and θ_1, θ_2, and θ_3 denote the three model parameters, maximal specific growth rate (μ_m), saturation constant (K_s), and yield coefficient (Y), respectively.

The Monod differential equation (1) -(3) can be easily integrated by separation of variables and the solution gives a function $\eta(t)$ implicitly given by the equation:

$$(4) \qquad\qquad t = \frac{1}{\theta_1}\left[(1+b)\ln(\eta/\eta_0) - b\ln\frac{c-\eta}{c-\eta_0}\right],$$

where the constant b is defined by

$$(5) \qquad\qquad b = \frac{\theta_2\theta_3}{s_0\theta_3 + \eta_0}.$$

This function is depicted for a typical example in Figure 1 and has remarkable properties (Pirt, 1975):

1. It is a sigmoidal growth curve with one inflection point.

2. The microbial biomass tends to a horizontal asymptote determined by $\theta_3 s_0 - x_0$. This value is the microbial biomass at the stationary state. Only the yield coefficient determines the horizontal asymptote. The other two parameters do not affect it.

It is necessary to emphasize two important assumptions that should be taken into consideration by a practitioner attempting to identify Monod parameter values from experimental data. First, the sigmoidal function, which is a solution of the Monod model, starts in the exponential growth phase. Therefore, an experimenter must exclude measurements reflecting a lag phase. It is equivalent to shifting the experiment starting time to the right, to a time equal to the observed lag phase. Second, a solution of the Monod model demonstrates an infinite stationary phase, where microbial biomass tends to a stationary value. However, this is not observed in real life. Typically, a stationary phase is not observed for a long time, and, sometimes, it occurs only for a very short time, to be followed by a decline in microbial mass. Nevertheless, it is very important to identify the microbial biomass (or microbial concentration) at the stationary phase, since this value determines the yield coefficient. Therefore, when determining the Monod parameter values, a practitioner should determine when the stationary phase starts and the biomass at that moment, ignoring any decline in biomass that may follow.

3 Optimal design of experiments for the Monod model

Identification of the Monod values from experimental data is a very important practical problem. In many microbiological studies it is only possible to record biomass values at occasional time points. Then, the experimental data is a table of two columns containing the time of measurements and the corresponding measured mass values in one or several experimental replications. We will call the set of all times of measurements an experimental design. Experimental data for experiments in the Monod model usually conform to a non-linear regression model, which means that the measured biomass value is the sum of two components, a first deterministic component (i.e. exact value of the Monod model), and a second random error of measurements:

$$(6) \qquad y_{ji} = \eta(t_j, \theta) + \varepsilon_{ji}, \; i = 1, \ldots, r_j, \; j = 1, \ldots, n,$$

where y_{ji} are experimental observations taken under the experimental conditions t_j ($j = 1, \ldots, n$), $\eta(t_j, \theta)$ is the solution of the Monod equation at the given time t_j, ε_{ji} denotes independent random variables with expectation 0 and constant variance, and the vector $\theta = (\theta_1, \theta_2, \theta_3)$ denotes the parameters of the Monod model, that have to be estimated from available data $\{y_{ji}\}$. Then the typical practical problem is to evaluate the parameters of the non-linear regression model from the available data with some appropriate level of precision. It is important to notice several assumptions in this non-linear regression model. Usually it is accepted that the observations follow a normal distribution, where the mean equals the exact value of the deterministic component, with dispersion that can have a fixed value for all experimental conditions, or can depend on the value of the deterministic component. The latter assumption is usually referred to as homoscedasticity in the literature. However, this is a statistical assumption and the true error distribution can be different and depend on the experimental technique and perhaps remain unknown. Another assumption, and the most important, is that the deterministic component is determined by the Monod model.

The most popular technique to evaluate parameters of regression models is to apply the least squares estimator, where the best estimator of θ is a value that provides the minimal value for the nonlinear sum of squares:

$$(7) \qquad \sum_{j=1}^{n} \sum_{i=1}^{r_j} (y_{ij} - \eta(t_j, \theta))^2.$$

In general the least squares estimator is not uniquely determined, and there may exist several values for θ minimizing this sum. This property depends on the particular regression function

$\eta(t, \theta)$ under consideration. Also the least squares method, in general, does not guarantee that the obtained estimators are close to the true parameter values. The main and most serious problem with parameter identification for the Monod model is that the estimators of the maximal specific growth rate and the saturation constant are closely correlated (Holmberg, 1982). Thus, several solutions of the Monod equation with very different parameter values can have similar distances to the experimental data (see Dette at al., 2005, p. 157, Fig 6.7). It was shown theoretically (see Dette et al, 2003, section 2 for a detailed proof) that the least squares estimator for the Monod non-linear regression model is consistent and asymptotically normal. This theorem indicates that if one did a large number of independent experimental replications, the least squares estimator would be close to the "true" parameter values and that quantiles from the normal distribution could be used for the calculation of confidence intervals. However, when there are only a few experimental replications, there are no guarantees that the least squares estimates will be close to the true values. On the other hand, computer simulations demonstrate that for replications of several dozens, the parameter estimators are close to their "true" values, if the experimenter employs a uniform experimental design (e.g. do measurements in 20 equidistant time points; see Dette et al., 2005, p. 153, Table 6, for details). An interesting question in this context is if there exist alternative strategies to collect the data, or to a reduction of the total sample size, which yield more precise estimates of the parameters without loosing statistical significance.

Optimal experimental design techniques give a solution this problem. Informally, it directs experimental measurements to specific time points that carry maximal information concerning a given regression model. It was demonstrated by Dette et al. (2003, 2005) that the application of optimal experimental designs for statistical analysis in the Monod model allows one to resolve the problem of parameter correlation, and, at the same time, decrease the necessary number of measurements.

The basic concept of this method is easy to understand. Assume that there are N experimental observations obtained from r_i experiments under experimental conditions t_i $i = 1, \ldots, n$, so $N = \sum_{j=1}^{n} r_j$. The set $\xi = \{t_1, \ldots, t_n; \omega_1, \ldots, \omega_n\}$ is called a design and defines the relative proportion $\omega_j = r_j/N$ of total observations taken at each time point t_j, $j = 1, \ldots, n$. In practice the weights ω_j are not necessarily multiples of $1/N$ and a rounding procedure has be applied to obtain $r_j \approx N\omega_j$. For example, if $N = 100$ is the total sample size and $\xi = \{t_1, t_2, t_3; 1/3, 1/3, 1/3\}$, the experimenter could take 33, 33 and 34 observations at the points t_1, t_2 and t_3, respectively. Let θ^* be the vector of 'true' but unknown parameter values, and denote by $\hat{\theta}_N$ be the least square estimate obtained from the given N experimental observations. Then the precision of the estimates depends on the covariance matrix of the vector $(\hat{\theta}_N - \theta^*)$. Under some assumption of regularity (see Dette et al., 2005 p. 142-143 for

details) and for the sufficiently large sample size this matrix is given by

$$(8) \qquad \frac{\sigma^2}{N} M^{-1}(\xi, \theta^*),$$

where σ denotes the standard deviation of the errors in model (6) and the matrix $M(\xi, \theta)$ is defined by

$$(9) \qquad M(\xi, \theta) = \left(\sum_{k=1}^{n} \omega_k \frac{\partial \eta(t_k, \theta)}{\partial \theta_i} \frac{\partial \eta(t_k, \theta)}{\partial \theta_j} \right)_{i,j=0}^{m}$$

The matrix $M(\xi, \theta)$ is called the Fisher information matrix and consists of the multiplication of partial derivatives of the regression function by parameters. A "smaller" matrix $\frac{\sigma^2}{N} M^{-1}(\xi, \theta^*)$ or a "larger" matrix $\frac{N}{\sigma^2} M(\xi, \theta^*)$ means a more a precise estimate $\hat{\theta}_N$. Consequently the matrix $M(\xi, \theta)$ is the key object for the experimental design technique. Note that this matrix depends on both experimental design and parameter values and an optimal design maximizes a real valued function of the matrix $M(\xi, \theta)$, which is called the optimality criterion in the statistical literature. Some properties of this matrix have useful geometrical interpretations, which explain different optimality criteria (see Fedorov, 1972 and Hidalgo, Ayesa 2001, for the details of the geometrical interpretation). The eigenvectors of this matrix determines principal directions in the three dimensional parametric space, representing uncorrelated linear combinations of the original variables. For any given experimental design ξ, the square roots of the eigenvalues and determinant of this matrix define respectively the length of the axes of the information ellipsoid and its volume. Obviously, the more the information ellipsoid is similar to a sphere, the less dispersion of the parameter estimate and less correlation between its components is observed. Therefore the main purpose of the optimal design is to find such a design that minimizes the volume of the information ellipsoid. Such a design is called D-optimal design, and it is employed in this paper. In Figure 2 we display the projections of the ellipsoids corresponding to the local D- and a naively chosen uniform design on the two dimensional parameter space corresponding to μ_m, K_s. This figure gives a simple geometrical representation of the D-optimal design meaning: the elongated regression ellipsoid for the naive design corresponds to the case when parameter estimates are correlated, and in the optimal experimental design case, the regression ellipsoid looks like a ball, indicating that parameter estimates are not correlated. The mathematical definition of this criterion is (see also Dette et al., 2005) is as follows: a design is called local D-optimal design if it maximizes the quantity

$$(10) \qquad \det M(\xi, \theta^0),$$

where θ^0 is a given initial value for the true parameter vector. It is critical to notice that the local optimal design depends significantly on the initial guess for the true parameter - this is the most critical restriction for the local optimal designs.

It was shown by Dette at al. (2003, paragraph 4.2) that local optimal designs are robust with respect to a misspecification of the initial parameter values. Alternatively one could use a more sophisticated maximin design procedure to incorporate uncertainty with respect to the parameters in the construction of optimal designs (see Dette et al., 2005b).

We also note that one could use alternative properties of the information matrix to construct optimal designs for the Monod model, which would result in the E-, modified E-, and c-optimality criteria (see Dette et al., 2003; Dette et al., 2005). However, for the Monod model these criteria do not have real practical advantages compared to the local D-optimal design (see Dette et al., 2003, paragraph 5), and for this reason we restrict this paper to the consideration of local D-optimal designs.

In summary, the application of local D-optimal designs for the Monod model includes the following consecutive steps:

1. An experimenter should be convinced that the given process is determined by Monod kinetics. For example, it is important that only one consumed resource is a limiting growth factor, and that numerous environmental (or incubation) factors, such as temperature, water activity, salinity etc., do not vary significantly during the experiment.

2. The lag phase of microbial growth, which is not incorporated in the Monod model, should be excluded from the considered experimental data. The Monod model assumes infinite stationary phase with constant biomass values, and to meet this assumption experimental data should be modified in the following way: a) the starting point of the stationary phase and the biomass (concentration value) at that moment should be determined, and b) this value should be assumed to be constant for all incubation periods after the start of stationary phase.

3. One should have a respectably good initial "guess" about the possible parameters in the Monod model. Such parameter approximation can be obtained-for example-from preliminary experimental data or from results of similar studies. If such data are unavailable, then the use of maximin optimal designs (Dette et al., 2005b) should be considered. These designs require the specification of a range for the parameters in the Monod model.

4. One should establish initial microbial (η_0 or equally x_0) and substrate (s_0) concentrations and, having an initial guess of the parameters, should evaluate the value of T under the given incubation conditions.

5. Using the given computer program (described in the following section), calculate the local D-optimal experimental design for the experiment.

6. Taking into account the assumptions concerning lag phase and stationary phase, perform the experiment in at least three replicates (better 5 replicates).

7. Evaluate the Monod kinetics parameters, the maximal specific growth rate (θ_1 or μ_m), saturation constant (θ_2 or K_s), and yield coefficient (θ_3 or Y) by the least squares method.

We finally note that most of the presented results concern only the Monod model, and cannot automatically be generalized to other non-linear regression models (see also Dette et al., 2005).

4 Numerical calculation of local D-optimal design and the Kiefer criterion

Computer programs are presented in Mathematica 5.0. This software package has a detailed help system and we have omitted explanations of the programming language and most of functions used. These details can be easily obtained from the Mathematica Book incorporated into the software, or on-line at the website http://www.wolfram.com/.

There are several alternative algorithms that can be applied for calculation of the local D-optimal design. One algorithm, based on analytical results, was suggested in our previous paper (Dette at al., 2003, part 3). Following these results, a locally three-point D-optimal design can be determined as follows: 1) one of the optimal design points is always the right-side boundary point T of the experimental interval $[0, T]$. This result is easy to understand, since this point allows us to identify the yield coefficient, and consequently it should be the closest point of the experimental region $[0, T]$ to the asymptotic stationary value. Therefore, the problem of finding the three-point optimal design is reduced to the problem of finding a two-point optimal design (see also Dette et al., 2005a). 2) The other two points of the experimental design can be found with the function presented in theorem 4 (Dette at al., 2003, p. 731).

In this paper we present a Mathematica 5.0 program that employs a simpler algorithm, based on the numerical calculation of the Fisher information matrix. The advantage of this algorithm, compared to the analytical one, is that it can be easily adapted to different Monod model modifications, and, also, for other nonlinear regression models. In addition, we feel that consideration of this algorithm will be helpful to practitioners trying to understand the optimal design method.

The employed algorithm consists of four parts: 1) establish the initial parameter guess and experimental conditions, 2) arbitrarily select three design points in the interval $[0, T]$, 3) by random search find 3 time points at which the determinant of the Fisher information matrix takes its maximal value, 4) check if the obtained experimental design is the local D-optimal design by a calculation of the Kiefer criterion, which provides a simple condition to prove optimality of the optimal design. In the following we describe for each of these steps the corresponding Mathematica code.

The first computer program calculates the local optimal design for the Monod model. This program incorporates the following two steps:

Step 1:

A1: Specification of a guess for the initial parameter and initial biomass and resource concentrations. For illustration we choose the following values

$$\mu_m = 0.25$$
$$K = 0.5$$
$$Y = 0.25$$
$$s_0 = 1$$
$$x_0 = 0.03$$

At this stage it is already assumed that the first four requirements of the method mentioned at the end of Section 3 have been checked:

1) The investigated experimental setup corresponds to Monod kinetics.

2) Preliminary experimental data or literature sources have allowed an initial "guess" about likely parameter values to be made, and the assumptions concerning lag phase and stationary phase are taken into account. In this example, averaged realistic parameter values ($\theta_1 = \mu_m = 0.25\ h^{-1}$; $\theta_2 = K_s = 0.5\ mg/l$, $\theta_3 = Y = 0.25\ mg/mg$) were selected from John Pirt's classical book (1975) and it is assumed that there is no lag phase.

3) Initial microbial and substrate concentrations have been selected according to the experimental setup ($x_0 = 0.03\ mg/l$, $s_0 = 1\ mg/l$). Also, a possible experimental time was fixed to 400 hours, obviously an overestimation for the given parameter values, since Figure 1 clearly demonstrates that the maximal time T in this experiment could be limited to 24 hours.

11

B1: Determination of the partial derivatives of the solution of the Monod model, with respect to the Monod model parameters. These functions are components of the Fisher information matrix. All these functions have been obtained using a formula for the calculation of the derivative of an implicitly given function (presented in any advanced calculus text book). The partial derivatives with respect to the first (maximal specific growth coefficient), second (saturation constant), and the third (yield coefficient) parameters of the solution of the Monod differential equation with respect to the model parameters are respectively:

$$V[z\_, y\_] = \frac{y\, z\, (-y + Y\, s_0 + x_0)}{-y + K\, Y + Y\, s_0 + x_0}$$

$$W[y\_] = \frac{y\, Y\, \left(-\log(\frac{y}{x_0}) + \log(\frac{-y + Y\, s_0 + x_0}{Y\, s_0})\right)\, (-y + Y\, s_0 + x_0)}{(Y\, s_0 + x_0)\, (-y + K\, Y + Y\, s_0 + x_0)}$$

$$L[y\_] = \frac{K\, y\, \left(-1 - \log(\frac{y}{x_0}) + \log(\frac{-y + Y\, s_0 + x_0}{Y\, s_0})\right)\, x_0\, (-y + x_0)}{(Y\, s_0 + x_0)^2\, (-y + K\, Y + Y\, s_0 + x_0)} +$$

$$+ \frac{K\, y\, Y\, s_0\, \left(y + \left(-1 - \log(\frac{y}{x_0}) + \log(\frac{-y + Y\, s_0 + x_0}{Y\, s_0})\right)\, x_0\right)}{(Y\, s_0 + x_0)^2\, (-y + K\, Y + Y\, s_0 + x_0)}$$

Note that the first function $V[z\_, y\_]$ depends on both time and microbial biomass and the other two depend only on biomass.

C1: The Fisher information matrix is introduced by the formula:

$$FM[T\_, d\_] := Re[\{\{V(T, d).V(T, d), V(T, d).W(d), V(T, d).L(d)\},$$
$$\{W(d).V(T, d), W(d).W(d), W(d).L(d)\},$$
$$\{L(d).V(T, d), L(d).W(d), L(d).L(d)\}\}]$$

D1: An arbitrarily initial 3 point experimental design is specified

$$TV = \{1, 20, 30\};$$

$$y[t\_] = \text{NDSolve}[\{x'[t] == \frac{\mu_m \ (Y \ s_0 + x_0 - x(t)) \ x(t)}{K \ Y + Y \ s_0 + x_0 - x(t)}, x[0] == 0.03\},$$

$$x[t], \{t, 0, 40\}, MaxSteps \rightarrow 20000];$$

$$\text{DesignValues} = x[TV]/.y[TV];$$

E1: Three intermediate values, necessary for the computational cycle organization, are determined:

$$\text{CritDet} = Det[FM[TV, \text{DesignValues}[[1]]]];$$

$$\text{CritD11} = 1;$$

$$\text{Design11} = \{2, 2, 3\}$$

The first number is the determinant of the Fisher Information Matrix with respect to the arbitrary Time Vector (TV). The second number (CritD11) is the intermediate number, which will be continuously updated for comparison with the determinant of the Fisher Information Matrix (FM). And the third is the intermediate value of the experimental design, which will be updated during the cyclic calculations.

F1: The cyclic calculations determine the local D-optimal experimental design:

While[Design11 $\neq$ TV,

    For[des $= 1$, des < 4,

   Do[

      Design $=$ ReplacePart[TV, Random[Real, $\{0.001, 400\}, 5$], des];

      $y[t\_] =$ NDSolve[$\{x'[t] == \dfrac{\mu_m \ (Y\, s_0 + x_0 - x(t))\ x(t)}{K\, Y + Y\, s_0 + x_0 - x(t)}, x[0] == 0.03\}$,

        $x[t], \{t, 0, 400\}, MaxSteps \rightarrow 20000$];

      DesignValues $= x[\text{Design}]/.y[\text{Design}]$;

      NewDet $=$ Det[FM[Design, DesignValues[[1]]]];

      If[NewDet $>$ CritDet, $\{$Design11 $= TV$, CritDet $=$ NewDet, $TV =$ Design$\}$], $\{10000\}$];

      des $+ +$];

It is important to notice that the algorithm will stop when the design cannot be improved in one cycle, restricted by the number of the digits in the design points. For instance, in the program presented, all the design points are real numbers with 5 digits from the $\{0.001; 400\}$ interval, given by the function Random[Real, $(0.001, 400), 5$],des]. Typically for this precision, the algorithm converges in 4-5 cycles, and the condition Design1$\neq$TV can be replaced by 5 cyclic operations. It is also important to keep in mind that to observe and control any step in the calculations, one could insert the function print [parameter] at any point of this cyclic program.

Note that at this stage the general time of experiment T should be determined and included as the limit of integration in the function NDSolve, and all numbers included in the TV should be less than T, for instance, here T is equal to 400. Note that this is much longer than the actual time interval used (see Figure 1). In practical calculations with this model, random taken design points sometimes give very small values of the Fisher Information matrix components and an indeterminate message can occur: Det::mindet: Input matrix contains an indeterminate entry. This message should be ignored, because this indeterminate situation for one randomly taken entry does not affect the general results from numerous runs.

G1: After the cyclic calculations from step F1 are finished, the local D-optimal experimental design can be delivered by the following request: "Design". This concludes the local D-optimal design calculation.

The three-point local D-optimal experimental design calculated for this particular example is $\{9.5365, 16.741, 53.110\}$. However, note that the third design point is unrealistic, because it reflects an asymptotic stationary growth phase, which usually does not exist.

The important question is, does the described program really yield the local D-optimal experimental design. To answer this question it is necessary to show that the described algorithm is robust and does not depend on the initial design entry, that it converges and that the corresponding limit is the local D-optimal design. This question can be answered in general by employing random search methods of extreme values, but this goes beyond the scope of the present paper. As an alternative we propose to use the Kiefer criterion (Fedorov, 1972), which allows us to check if any arbitrary given experimental design is a local D-optimal design. Roughly speaking the Kiefer criterion yields for any design a curve on the experimental region (in our case $[0, T]$). If one of these curves stays below the line $y = 1$, the corresponding design is the local D-optimal design. The following program is for the Kiefer criterion calculation:

Step 2:

A2: The first two steps of this program are identical to the previous steps A1 and B1 of program. One should determine the initial parameter values and partial derivatives of the Monod model with respect to the parameters.

B2: Here the local optimal experimental design found in the previous step is introduced:

$T = \{\{9.5365, 16.741, 53.110\}\};$

$\mathrm{NDSolve}[\{x'[t] == \dfrac{\mu_m \ (Y \ s_0 + x_0 - x(t)) \ x(t)}{K \ Y + Y \ s_0 + x_0 - x(t)}, x[0] == 0.03\}, x[t], \{t, 0, 60\}];$

$d = \mathrm{Evaluate}[\{x[T[[1, 1]]], x[T[[1, 2]]], x[T[[1, 3]]]\} /. \%];$

The vectors T and d determine the optimal times of measurements and biomass values at these time points.

C2: The inverse matrix of the Fisher information matrix is determined:

$$IFM = \text{Inverse}[Re[\{\{V(T,d)[[1]].V(T,d)[[1]], V(T,d)[[1]].W(d)[[1]], V(T,d)[[1]].L(d)[[1]]\},$$
$$\{W(d)[[1]].V(T,d)[[1]], W(d)[[1]].W(d)[[1]], W(d)[[1]].L(d)[[1]]\},$$
$$\{L(d)[[1]].V(T,d)[[1]], L(d)[[1]].W(d)[[1]], L(d)[[1]].L(d)[[1]]\}\}]]$$

D2: Determination of the vector of partial derivatives of the Monod function by parameters:

$$\text{PartialDerivativesVector}[z\_, y\_] = (V[z,y] \; W[y] \; L[y])$$

E2: Introduction of the Kiefer Criterion:

$$\text{KieferCriterion}[z\_, y\_] =$$
$$\text{PartialDerivativesVector}[z,y].IFM.\text{Transpose}[\text{PartialDerivativesVector}[z,y]]$$

F2: Calculation of the Kiefer Criterion over the experimental region:

If the Kiefer criterion is less than or equal to 1 over the experimental region, it follows from the famous equivalence theorem of Kiefer and Wolfowitz (1960) that the chosen design is in fact local D-optimal. Moreover, at the experimental conditions specified by optimal design the curve must be in fact equal to 1 In order to obtain a graphical representation of the Kiefer criterion we first introduce $10,000$ randomly taken points in the considered time interval and calculate a biomass value at each point:

TimePoints $=$ Table[Random[Real, 60], $\{10000\}$];

$\text{NDSolve}[\{x'[t] == \dfrac{\mu_m \; (Y \; s_0 + x_0 - x(t)) \; x(t)}{K \; Y + Y \; s_0 + x_0 - x(t)}, x[0] == 0.03\}, x[t], \{t, 0, 60\}];$

BiomassValues $=$ Evaluate[Map[x, TimePoints]/.%];

Then the Kiefer criterion for each point is calculated:

CriterionValues=Re[KieferCriterion[TimePoints,BiomassValues[[1]]][[1]]];

Now one can plot the graph of the Kiefer criterion together with the list plot of the values of the Kiefer criterion at the optimal design points. For the considered example the situation is depicted in Figure 3.

pl1 = ListPlot[Transpose[{TimePoints,CriterionValues];

pl2=ListPlot[Transpose[$\{T[[1]], Ki\}$],PlotStyle $\rightarrow$ {PointSize[0.02], RGBColor[0,1,0]}];

Show[pl1,pl2];

The typical curve in Figure 3 demonstrates that the determined design is in fact local D-optimal.

5 Conclusions

The method of optimal experimental design for the Monod model can undoubtedly become a useful tool in everyday laboratory practice in environmental microbiology and biomedical research. The important advantages of optimal experimental designs (over naive designs for the Monod model) reduce both the cost and duration of experiments : 1) Application of experimental designs significantly reduces number of necessary experimental measurements, and the correlation between the parameter estimates is diminished. 2) Optimal experimental designs yield more precise estimates. However, to became an everyday tool this method should be derived as a tool. In this paper we have presented the theory of local D-optimal design for the Monod model as a tool ready for practical applications. This method has an important limitation, i.e. a good initial guess of parameter values should be available. If an experimenter has no such preliminary information, then the minimax optimal design should be employed. In addition to this presentation, we direct practitioners to the actively developing website where necessary software will be soon available at www.optimal-design.org.

References

[1] Berkholz, R., Rohlig, D., Guthke, R. (2000). Data and knowledge based experimental design for fermentation process optimization // *Enzyme Microb. Technol.*, 27, 10, 784–788

[2] Blok, J. (1994). Classification of biodegradability by growth kinetic parameters // *Ecotoxicol. Environ. Saf.* 27, 294–305

[3] Blok, J., Struys, J. (1996). Measurement and Validation of Kinetic Parameter Values for Prediction of Biodegradation Rates in Sewage Treatment // *Ecotoxicol. and Environ. Saf.*, 33, 3, 217–227

[4] Dette, H., Melas, V.B., Pepelyshev, A., Strigul, N. (2003). Efficient design of experiments in the Monod model // *J. R. Stat. Soc. B*, 65, Part 3, 725–742

[5] H. Dette, V.B. Melas, A. Pepelyshev and N.S. Strigul. (2005). Design of Experiments in the Monod model - Robust and Efficient Designs // *Journal of Theoretical Biology*, 234, 4, 537–550

[6] Dette,H., Melas,V.B., Strigul,N., (2005). Application of optimal experimental design in microbiology. In: Wong,W.,Berger,M. (Eds.),Optimal Designs: their Roles and Applications. Wiley and Sons Ltd, New York,pp. 137–180

[7] Fedorov, V.V. (1972). *Theory of optimal experiments*, Academic Press

[8] Hidalgo, M.E., Ayesa, E. (2001). A numerical identifiability test for state-space models-application to optimal experimental design // *Water Sci. Technol.*, 43, 7, 339–346

[9] Holmberg, A. (1982). On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities // *Math. Biosci.*, 62, 1, 23–43

[10] Kiefer, J., Wolfowitz, J. (1960). The equivalence of two extremum problems // *Canad. J. Math.*, 14, 363–366

[11] Kiefer, J.C., (1974). General equivalence theory for optimum designs (approximate theory) // *Ann. Statist.* 2, 849–879

[12] Koch, A.L.,(1997). The Monod model and its alternatives. In: Koch,A.L.,Robinson,J., Milliken, G.A. (Eds.), Mathematical Modeling in Microbial Ecology. Chapman and Hall,New York, 62–93

[13] Kovarova-Kovar K., Egli T., (1998). Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics // *Microbiology and molecular biology reviews*, 62, 3, 646–666

[14] Merkel, W., Schwarz, A., Fritz, S., Reuss, M., Krauth, K. (1996). New strategies for estimating kinetic parameters in anaerobic wastewater treatment plants // *Water Sci. Technol.*, 34, 5–6, 393–401

[15] Monod, J. (1949). The growth of bacterial cultures // *Ann. Rev. Microbiol.*, 3, 371–393

[16] Munack, A. (1991). Optimization of sampling // In: *Biotechnology*, 4, 252–264, VCH Weinheim

[17] Pirt, S.J. (1975). *Principles of microbe and cell cultivation*, Wiley, New York

[18] Saez, P.B., Rittmann, B.E. (1992). Model-parameter estimation using least squares // *Water Res.*, 26, 6, 789–796

[19] Smets, I.Y.M., Versyck, K.J.E., Van Impe, J.F.M. (2002). Optimal control theory: A generic tool for identification and control of (bio-)chemical reactors // *Ann. Rev. Control*, 26, 1, 57–73

[20] Strigul, N.S., Kravchenko, L.V. (2006). Mathematical Modeling of PGPR Inoculation into the Rhizosphere // *Eviron. Model Softw.*, 21, 8, 1158–1171

[21] Vanrolleghem, P.A., Van Daele, M., Dochain, D. (1995). Practical identifiability of a biokinetic model of activated sludge respiration // *Water Res.*, 29, 11, 2561–2570

[22] Versyck, K.J., Bernaerts, K., Geeraerd, A.H., Van Impe, J.F. (1999). Introducing optimal experimental design in predictive modeling: A motivating example // *Int. J. Food Microbiol.*, 51, 1, 39–51
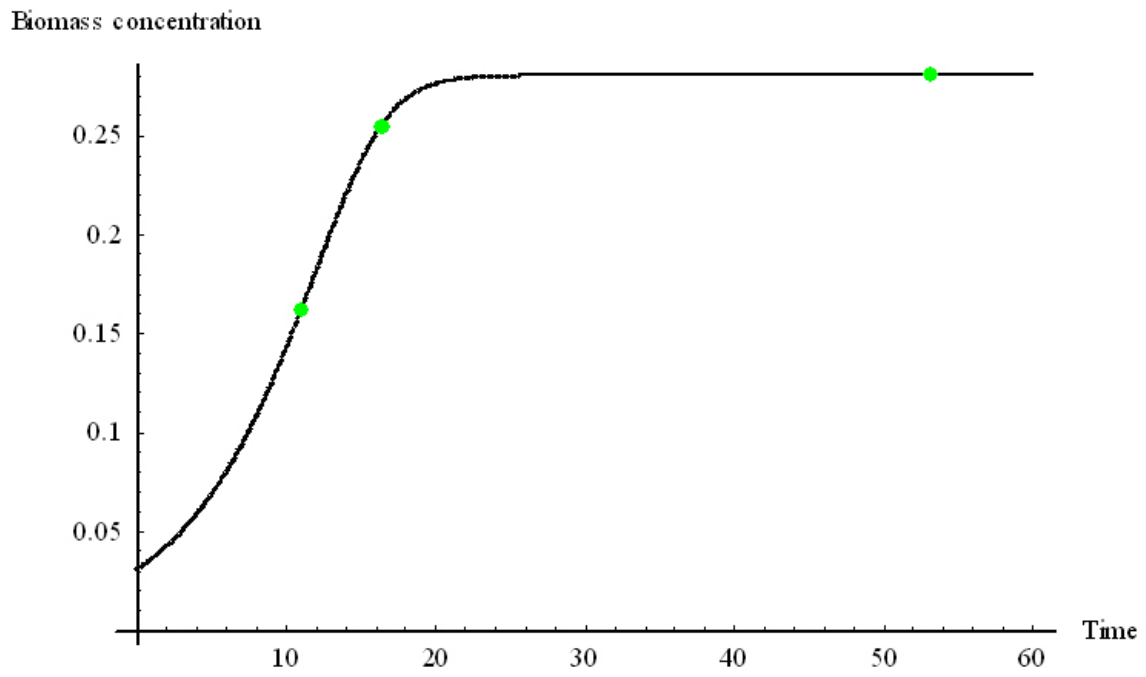
Figure 1: Solution of the Monod differential equation. Large points indicate biomass values at local D-optimal design points.
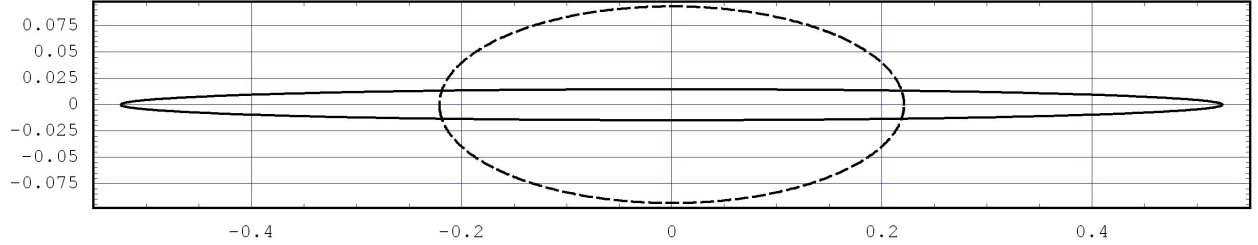
Figure 2: Comparison of information ellipses for the D-local optimal design (dashed line) and naive 10-point design (solid line) for the two parametric regression (μ_m, K_s). Areas of information ellipses for the local optimal design and the naive design are 0.000118 and 0.000042, respectively. For demonstration purposes the ellipses were centered, and the length of the major axis of the naive design ellipse was multiplied by 10.
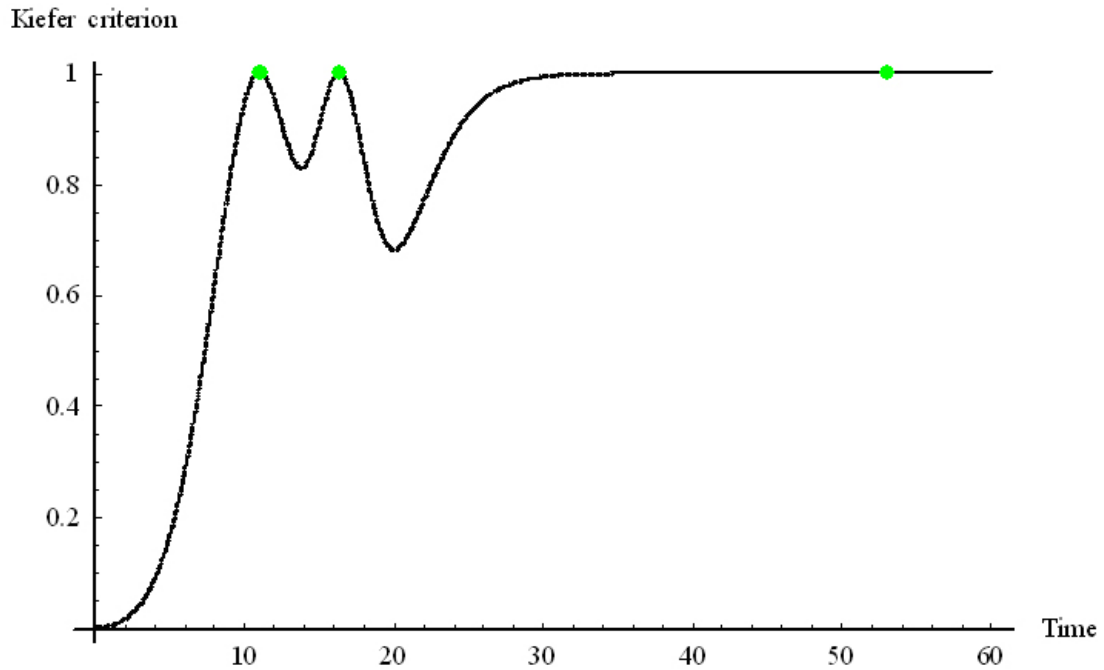

Figure 3: Kiefer criterion at different time points. Large points indicate Kiefer criterion values at local D-optimal design points.