

Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models

Rafael Muñoz-Tamayo, Pierre Martinon, Gaël Bougaran, Francis Mairet, Olivier Bernard

► **To cite this version:**

Rafael Muñoz-Tamayo, Pierre Martinon, Gaël Bougaran, Francis Mairet, Olivier Bernard. Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models. *Journal of Process Control*, Elsevier, 2014, 24 (6), pp.991-1001. 10.1016/j.jprocont.2014.04.021 . hal-00998525

HAL Id: hal-00998525

<https://hal.inria.fr/hal-00998525>

Submitted on 4 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models

Rafael Muñoz-Tamayo^a, Pierre Martinon^b, Gaël Bougaran^c, Francis Mairet^a, Olivier Bernard^{a,d}

^a*BIOCORE-INRIA, BP93, 06902 Sophia-Antipolis Cedex, France.*

^b*Commandes INRIA Saclay, Ecole Polytechnique, CMAP 91128 Palaiseau, France.*

^c*Ifremer, Laboratoire Physiologie et Biotechnologie des Algues, rue de l'île d'Yeu BP 21105 44311 Nantes cedex 3, France*

^d*LOV-UPMC-CNRS, UMR 7093, Station Zoologique, B.P. 28 06234, Villefranche-sur-mer, France*

Abstract

Mathematical models are expected to play a pivotal role for driving microalgal production towards a profitable process of renewable energy generation. To render models of microalgae growth useful tools for prediction and process optimization, reliable parameters need to be provided. This reliability implies a careful design of experiments that can be exploited for parameter estimation. In this paper, we provide guidelines for the design of experiments with high informative content based on optimal experiment techniques to attain an accurate parameter estimation. We study a real experimental device devoted to evaluate the effect of temperature and light on microalgae growth. On the basis of a mathematical model of the experimental system, the optimal experiment design problem was formulated and solved with both static (constant light and temperature) and dynamic (time varying light and temperature) approaches. Simulation results indicated that the optimal experiment design allows for a more accurate parameter estimation than that provided by the existing experimental protocol. For its efficacy in terms of the maximum likelihood properties and its practical aspects of implementation, the dynamic approach is recommended over the static approach.

Keywords: biofuel; biological processes; modelling; parameter identification; optimal experiment design

Email addresses: rafaun@yahoo.com (Rafael Muñoz-Tamayo), pierre.martinon@inria.fr (Pierre Martinon), gael.bougaran@ifremer.fr (Gaël Bougaran), francis.mairet@inria.fr (Francis Mairet), olivier.bernard@inria.fr (Olivier Bernard)

1. Introduction

Microalgae have received a specific attention in the framework of renewable energy generation [1]. However, optimizing productivity in large scale systems is a difficult task since microalgae growth is driven by multiple factors including light intensity, temperature and pH [2]. Mathematical modelling is thus required for quantifying the effect of environmental factors on microalgae dynamics.

In order to obtain reliable models that can be used in prediction and optimization of large scale systems, the model calibration stage requires carefully designed experiments with high informative content. Providing accurate parameters is indeed crucial since model-based optimality might be sensitive to parameters values as shown in [3]. Moreover, assessing the effect of operational factors via sensitivity analysis can provide useful information for improving configuration design of photobioreactors [4].

Looking for high informative experiments is the objective of optimal experiment design (OED) for parameter estimation. Extensive work has been done for tackling the OED problem for dynamical systems (see, *e.g.*, [5, 6, 7, 8]). **The OED problem can be formulated as an optimal control problem. For low dimension models, analytical solutions may be obtained by the application of Pontryagin's Maximum Principle, which provides necessary conditions to be satisfied by the optimal inputs (see, *e.g.* [9]). When model complexity increases, analytical solutions are arduous to obtain and thus the solution of the OED problem relies on numerical optimization techniques (see, *e.g.* [10, 11]).**

When dealing with biological systems, OED approaches are based either on static or dynamic experiments (see, *e.g.*, [12, 13]). In this work, we analyze these two strategies and capitalize the available tools for OED to provide guidelines for the design of optimal experiments that allow an efficient assessment of the effect of temperature and light on microalgae growth. The model under investigation represents a real experimental device used to assess optimal growth conditions under batch mode. This device is operated at IFREMER Nantes, France.

The paper is organized as follows. Section 2 presents the system under study and its

29 mathematical description, which corresponds to a simplified model of microalgae growth.
30 The OED framework based on this model is detailed in Section 3. In Section 4 we show
31 the results of solving the OED problem. **Two strategies are analyzed, namely static and**
32 **dynamic approaches.** Furthermore, we discuss about the relevance of OED for model-
33 driven decisions on raceway performance. For that, we make use of a local sensitivity
34 analysis of a more complex model describing microalgae growth in an outdoor pond. In
35 Appendix A, we discuss about the structural and practical identifiability of the model.
36 The main conclusions of the study are summarized in Section 5.

37 2. Modelling

38 We focus our study on the effect of temperature and light on the growth of microalgae.
39 More precisely, we aim at designing efficient experimental protocols for a real experimental
40 system that allow an accurate estimation of the model parameters. The experimental
41 apparatus, named the TIP (Fig. 1), consists of 18 batch photobioreactors located inside
42 an incubator (see [14] for more details). In each photobioreactor, it is possible to regulate
43 the temperature, pH and light intensity.

44 Following the models developed for microalgae growth [15, 16], we study here a sim-
45 plified model of microalgae growth under the hypotheses that the experiment is carried
46 out at low cellular concentrations and under conditions of non-limiting nutrients. The
47 first hypothesis implies that light is homogeneous along the depth of the photobioreactor.
48 The second hypothesis implies that the cells grow in exponential phase. The resulting
49 mass balance equation on the TIP system reads

$$\dot{x} = f(x, \boldsymbol{\theta}, I, T, t) = \bar{\mu}(\boldsymbol{\theta}, I, T)x, \quad x(0) = x_0, \quad (1)$$

50 with x the biomass concentration, I the light intensity and T the temperature in the
51 reactor, $\boldsymbol{\theta}$ the parameter vector and $\bar{\mu}(\cdot)$ the specific growth rate $\bar{\mu}(\cdot)$ defined by

$$\bar{\mu}(\boldsymbol{\theta}, I, T) = \mu_{\max} \phi_I \phi_T. \quad (2)$$

52 with μ_{\max} the maximal specific growth rate. The factors ϕ_I, ϕ_T , detailed below, represent
53 the effects of light and temperature on microalgae growth.

54 Temperature has a homogeneous effect on uptake and growth rates [17]. The effect of
 55 temperature is described by the cardinal model developed for bacteria [18] and validated
 56 for microalgae [16].

$$\phi_T = \begin{cases} 0, & T < T_{\min} \\ \frac{(T-T_{\max})(T-T_{\min})^2}{(T_{\text{opt}}-T_{\min})[(T_{\text{opt}}-T_{\min})(T-T_{\text{opt}})-(T_{\text{opt}}-T_{\max})(T_{\text{opt}}+T_{\min}-2T)]}, & T \in [T_{\min}, T_{\max}] \\ 0, & T > T_{\max}. \end{cases} \quad (3)$$

57 The effect of light (ϕ_I) on microalgae growth is often represented by a Haldane type
 58 kinetics that accounts for photoinhibition [19]. The following parameterization of the
 59 standard Haldane equation is used [16]

$$\phi_I = \frac{I}{I + \frac{\mu_{\max}}{\alpha} \left(\frac{I}{I_{\text{opt}}} - 1 \right)^2}, \quad (4)$$

60 where α is the initial slope of the growth response curve w.r.t. light.

61 In terms of practical identifiability properties, Eq. (4) excels the standard Haldane
 62 kinetics. For a brief discussion, the reader is referred to Appendix A.

63 The above equations implies that microalgae exhibit a maximal growth rate at optimal
 64 conditions of light (I_{opt}) and temperature (T_{opt}).

The model is then determined by the parameter vector θ

$$\theta = [\mu_{\max}, \alpha, I_{\text{opt}}, T_{\min}, T_{\max}, T_{\text{opt}}].$$

65 In the next Section, we tackle the OED problem locally, that is the design of op-
 66 timal experiments is carried out on the basis of nominal values $\hat{\theta}$. Table 1 shows the
 67 nominal values of the model parameters used in this study. They correspond to the mi-
 68 croalgae *Isochrysis* aff. *galbana*, currently named as *Tisochrysis lutea* [20]. Parameter
 69 values were mainly obtained from [3] and [21]. The temperature parameters are those
 70 of *Nannochloropsis oceanica* [16] whose maximal and optimal temperatures are close to
 71 those of *Tisochrysis lutea* [22].

3. OED problem for parameter estimation

The problem of OED for parameter estimation consists in designing an experimental protocol that provides data with high informative content to allow an accurate identification of the model parameters, that is to provide estimates with small confidence intervals. Classical approaches of OED for parameter estimation rely on the optimization of a scalar function of the Fisher information matrix (FIM), since this matrix is the core for the calculation of the confidence intervals of the parameter estimates (see, *e.g.*, [6], [8]). Recent approaches such as the Sigma Point Method have been proposed to estimate parameter uncertainty without the explicit calculation of the FIM [23]. Here, we will focus on the classical approach.

Let us recall some basic principles of parameter identification. We consider here a local design approach. Our aim is to design optimal experiments on the basis of the nominal parameter vector $\hat{\boldsymbol{\theta}}$. We first assume that the i th measurement (observation) y_i of our experiment is modelled as:

$$y_i = y_{m_i}(\boldsymbol{\theta}^*) + \varepsilon, \quad (5)$$

where $y_{m_i}(\boldsymbol{\theta}^*)$ is the deterministic output of the model and $\boldsymbol{\theta}^*$ the true value of the parameter vector. The measurement error ε is here assumed to follow a normal distribution $\varepsilon \sim \mathbf{N}(0, \sigma^2)$. Note that (5) implies that a deterministic model is available and represents adequately the system. Moreover, the model structure must be structurally identifiable. In Appendix A, structural identifiability of the model is checked.

The maximum likelihood (ML) estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ minimizes the cost function

$$J(\boldsymbol{\theta}) = \frac{1}{\sigma_s^2} \sum_{i=1}^n (y_i - y_{m_i}(\boldsymbol{\theta}))^2, \quad (6)$$

with n the number of data points.

The covariance matrix $\hat{\mathbf{P}}$ of $\hat{\boldsymbol{\theta}}$ can be approximated to

$$\hat{\mathbf{P}} = \mathbf{F}^{-1}(\hat{\boldsymbol{\theta}}), \quad (7)$$

where \mathbf{F} is the Fisher information matrix. An estimate of the standard deviation of $\hat{\theta}_j$ is given by

$$\eta_j = \sqrt{\hat{\mathbf{P}}_{jj}}. \quad (8)$$

96 We will be then interested in designing an experiment that render η_j small.

97 In our case study, we aim at determining optimal profiles (or levels) of temperature
 98 (T) and light intensity (I) for attaining an accurate estimation of parameters. Optimal
 99 experiments are built w.r.t. the D-optimality criterion, which maximizes the determinant
 100 of the FIM. Maximizing the determinant implies minimizing the volume of the confidence
 101 ellipsoids for the parameters [6].

102 By means of simulations, we tested also other optimality criteria, namely E-optimality
 103 (maximization of the smallest eigenvalue of the FIM) and modified E-optimality ((min-
 104 imization of the condition number of the FIM)). D-optimality provided the best results
 105 in terms of the volume of the confidence ellipsoids. Therefore, we chose D-optimality as
 106 criterion of optimal design. Interestingly, the modified E-optimality criterion resulted in
 107 large confidence intervals. Indeed, it has been noted that since the modified E-optimality
 108 criterion is a criterion of shape of the ellipsoids, it is possible to obtain circular confidence
 109 regions with large volumes [24].

110 It should be noted that the performance of the obtained optimal experiment strongly
 111 depend on the nominal values of the estimates of the parameter vector. Ideally, $\hat{\boldsymbol{\theta}}$ should
 112 be as close as possible to $\boldsymbol{\theta}^*$. In our case study, the nominal values of the parameters
 113 used are expected to be close to the true values, since the selection of *priors* was based
 114 on published experimental studies.

115 The OED problem is tackled by means of two strategies, namely dynamic and static
 116 approaches, which are detailed in the following.

117 3.1. Dynamic approach

118 The OED by the dynamic approach is directly applied on the dynamic (primary)
 119 model (1). Here, the temperature and light intensity can be set to vary in time.

120 For the dynamic approach, the FIM reads as follows

$$121 \mathbf{F}_d(\hat{\boldsymbol{\theta}}) = \frac{2}{\sigma_d^2} \sum_{k=1}^{n_e} \sum_{i=1}^{n_t} \left[\frac{\partial y_{m_{k,i}}}{\partial \boldsymbol{\theta}} \right]_{\hat{\boldsymbol{\theta}}}^T \left[\frac{\partial y_{m_{k,i}}}{\partial \boldsymbol{\theta}} \right]_{\hat{\boldsymbol{\theta}}} = \frac{2}{\sigma_d^2} \widehat{\mathbf{M}}_d, \quad (9)$$

122 with n_t the number of sampling times. Here, $y_{m_{k,i}}$ is the biomass concentration predicted
 by the model (1) for the k th experiment at the i th time and σ_d^2 is the noise variance of the

123 measurement of biomass concentration. $\widehat{\mathbf{M}}_d$ is the matrix resulting from the summation
 124 term. This formulation is made for facilitating further discussion. The terms in brackets
 125 in (9) contains the local sensitivities of the model output w.r.t. the parameter vector $\boldsymbol{\theta}$.
 126 The sensitivity functions were computed automatically with the Matlab Toolbox IDEAS
 127 [25]. The toolbox is devoted to estimate parameters of ODE models. It uses symbolic
 128 differentiation to calculate the sensitivity functions for the evaluation of the FIM.

129 An approximate noise variance $\sigma_d^2 = 9.31$ was calculated from the data reported in
 130 [26] and the mathematical model developed in [27].

131 The OED problem is defined as

$$\min_{\boldsymbol{\varphi}_d} \text{Det}(\mathbf{F}), \quad (10)$$

132 with $\boldsymbol{\varphi}_d$ the design vector

$$\boldsymbol{\varphi}_d = [T_1(t), I_1(t), \dots, T(t)_{n_e}, I(t)_{n_e}],$$

such that

$$T_L = 12 \leq T_k(t) \leq T_U = 33.2^\circ\text{C} \quad (11)$$

$$I_L = 20 \leq I_k(t) \leq I_U = 1200 \mu\text{E m}^{-2}\text{s}^{-1}$$

$$\dot{T}_L = -5 \leq \dot{T}_k(t) \leq \dot{T}_U = 15^\circ\text{C},$$

133 with n_e the number of distinct experiments. We set $n_e = 9$ with duplicate experiments.
 134 The boundaries in (11) correspond to the physical boundaries of the TIP system. Note
 135 that the rate of temperature change (\dot{T}) is imposed. This constraint is bounded by the
 136 thermal dynamics of the equipment but also it must account for the potential thermal
 137 stress induced to the microalgal cells.

138 No boundaries were imposed to the rate of change of light, since it can be changed
 139 instantaneously. We assumed that microalgae respond instantaneously to light changes.
 140 However, it is known that microalgae can adapt its photosynthetic system to changes of
 141 light [17]. Here, we consider time scales larger than the photosynthesis response time
 142 (in the range of minutes for photoinhibition). In this case study, we neglect however
 143 photoacclimation (adaptation of the pigment content to light intensity, at the scale of

144 weeks). Further experiments will be needed to assess the dynamics of such an adaptation.

145 Note that φ_d is of infinite dimension. However, φ_d will be further transformed into a
146 finite dimension vector to solve the optimization problem numerically.

147 Before attempting to solve the full OED problem, we first partitioned the original OED
148 problem into simpler subproblems in which we study the effect of either temperature or
149 light. This strategy was for instance used in [28] to estimate the cardinal temperatures
150 for *E. coli*.

151 Each subproblem is dedicated to improve the accuracy of the estimation of a couple
152 of parameters, while the other parameters are assumed to be known. In this case, \mathbf{F}_d is
153 a square matrix of dimension 2×2 for each subproblem (for the full OED problem, the
154 FIM is of dimension 6×6). The initial concentration of biomass was set to $x_0 = 10$ mg/L
155 and. The duration of the experiment to $t_f = 4$ d with ten equidistant sampling times.

156 When studying the temperature parameters, the light was set to $I = 547 \mu\text{E m}^{-2}\text{s}^{-1}$,
157 and when studying the light, the temperature was set to $T = 26.7$ °C. These constant
158 values correspond to the optimal values for growth obtained from the model parameters
159 (Table 1). This choice is supported by the fact that the FIM of each subproblem only
160 involves either parameters related to the effect of light or to the effect of temperature,
161 therefore the other experimental input only affects relatively the calculation of the sen-
162 sitivity functions. By setting the experimental inputs to their optimal values, we favor
163 growth.

164 A total number of nine subproblems was obtained. Table 2 shows the combination of
165 parameters and the experiment input (T or I) for each subproblem. In practice, the nine
166 solutions will be implemented in duplicates in the TIP.

167 The resulting subproblems were solved numerically via two discretization methods,
168 namely sequential and simultaneous. The discretization allows to convert the original in-
169 finite dimensional optimization problem into a finite dimension problem. In the sequential
170 approach (also called control vector parametrization (CVP)), the control variables are ap-
171 proximated by a set of basis functions that depend on a finite number of real parameters.
172 In the simultaneous approach, all state and control variables are discretized w.r.t. time.
173 Hence, this method is also known as total discretization. In this case, the dimension of

174 the optimization problem depends on the number of discretization steps [29, 30].

175 The simultaneous method was implemented with the open source toolbox BOCOP
 176 [31](<http://bocop.org>), based on the IPOPT solver [32]. The simultaneous method used
 177 a Midpoint discretization with 1000 steps, with a 10^{-14} tolerance for solving the discretized
 178 problem. All state and control variables were initialized with constant values.

179 Numerical solutions of the CVP approach were obtained with the SSMGO tool-
 180 box (<http://www.iim.csic.es/gingproc/ssmGO.html>), with the parameterization de-
 181 picted in Fig. 2. SSmGo performs global optimization by using a scatter search method
 182 [33, 34].

183 The experiment inputs are thus defined by four parameters, namely u_1, u_2, t_1, t_2 . The
 184 dimension of the optimization problem is therefore 9×4 with the decision vector

$$\varphi_d = [u_1(1), u_2(1), t_1(1), t_2(1) \dots, u_1(n_e), u_2(n_e), t_1(n_e), t_2(n_e)]. \quad (12)$$

185 3.2. Static approach

186 The OED by the static approach is based on the secondary model of growth (here
 187 represented in (2)). In this approach, one experiment is characterized by a constant
 188 environment (T, I in our case). The dynamic data of the biomass evolution for a given
 189 experiment is first used to calculate the maximal growth. Once different growth rates
 190 are calculated at different conditions of temperature and light intensity, the parameter
 191 estimation procedure is applied on the growth model (2).

192 Since the TIP system allows to run 18 experiments simultaneously, a parallel design
 193 procedure is here used. Hence, the following OED strategies aim at finding the nine best
 194 experiment conditions to account for duplicate experiments.

195 For the static approach, the FIM is computed as

$$\mathbf{F}_s(\hat{\boldsymbol{\theta}}) = \frac{2}{\sigma_s^2} \sum_{k=1}^{n_e} \left[\frac{\partial y_{m_k}}{\partial \boldsymbol{\theta}} \right]_{\hat{\boldsymbol{\theta}}}^T \left[\frac{\partial y_{m_k}}{\partial \boldsymbol{\theta}} \right]_{\hat{\boldsymbol{\theta}}} = \frac{2}{\sigma_s^2} \widehat{\mathbf{M}}_s, \quad (13)$$

196 where y_{m_k} is the maximal growth predicted by the model (2) for the k th experiment, n_e
 197 is the number of distinct experiments ($n_e = 9$) and σ_s^2 is the noise variance associated
 198 to measurement of the maximal growth. To provide an approximate value of the noise

199 variance, the dynamic model was simulated for nine experiments. Each of them charac-
 200 terized by a level of temperature and light intensity. Normal distributed data of biomass
 201 concentration was further generated by taking the value of noise variance of biomass
 202 σ_d^2 . The generated noisy data was used to calculate the variance of specific growth. An
 203 approximate value of $\sigma_s^2 = 3.8 \cdot 10^{-3}$ was obtained.

204 The OED problem is then defined as

$$\min_{\varphi_s} \text{Det}(\mathbf{F}), \quad (14)$$

205 with φ_s the design vector

$$\varphi_s = [T_1, I_1, \dots, T_{n_e}, I_{n_e}],$$

such that

$$\begin{aligned} T_L = 12 \leq T_k \leq T_U = 33.2^\circ\text{C} \\ I_L = 20 \leq I_k \leq I_U = 1200 \mu\text{E m}^{-2}\text{s}^{-1}. \end{aligned} \quad (15)$$

206 The design vector $\varphi_s \in \mathbb{R}^{n_e}$.

207 To solve the OED problem of the static approach, the Matlab optimization toolbox
 208 SSmGo was used.

209 4. Results and Discussion

210 Before presenting the resulting optimal experiments for both static and dynamic ap-
 211 proaches, we should keep in mind that in our case study the D-optimal experiments do
 212 not depend on the value of the noise variance σ^2 , given that we assumed that the measure-
 213 ment errors are homoscedastic. Indeed, the optimal experiment inputs depend only on
 214 the matrix $\widehat{\mathbf{M}}$, defined previously in (9,13). On the other hand, the confidence intervals
 215 of the estimates do depend on the actual value of σ since the estimate of the standard
 216 deviation of the parameter θ_j is given by

$$\eta_j = \frac{\sigma}{\sqrt{2}} \sqrt{(\widehat{\mathbf{M}}_{jj})^{-1}}. \quad (16)$$

217 4.1. OED by the static approach

218 The nine D-optimal experiments are given in Table 3. These experiments are defined
219 by six levels of light intensity and five levels of temperature (if the decimal digits are omit-
220 ted). Note that the nine experiments include the repetition of two experimental conditions
221 (experiments 1,2 and experiments 5,6), which results in seven distinct experimental con-
222 ditions. This result is not surprising since D-optimal often calls for the repetition of a
223 small number of experimental conditions [6]. Simulated data resulted from the D-optimal
224 experiments are illustrated in Fig 5A.

225 The performance of the D-optimal experiments was compared by means of simulation
226 with a equidistant full 3^2 factorial design including duplicates and with the central com-
227 posite design currently used in the TIP device [14]. This composite design involved 17
228 experiments with five levels for the environmental variables temperature, pH and light
229 intensity. Since in our study the effect of the pH is not considered in the OED problem,
230 we only took into account the levels for temperature and light of the 17 experiments. The
231 maximal level of temperature used in [14] was 33.7°C . We set the maximal temperature
232 of culture to 33.2°C , which is lower than the nominal value of the upper temperature for
233 algae growth (T_{\max}).

234 Table 4 illustrates the advantage of the D-optimal experiments over the factorial de-
235 signs. Firstly, we notice that with the equidistant full factorial design the determinant
236 of the FIM is zero, implying that the FIM is singular. Indeed, the inverse of the condi-
237 tion number of the FIM (defined as ratio of the largest eigenvalue to the smallest one) is
238 smaller than the precision of floating point format ($2 \cdot 10^{-16}$). In this case, confidence in-
239 tervals for the parameter estimates can not be computed on the basis of the density of the
240 estimator. To identify alternatives for guaranteeing a non-singular FIM for a full factorial
241 design, a series of computations was performed. From the computations, it is concluded
242 that a minimum number of four levels need to be considered in a full factorial design to
243 provide a well-conditioned FIM. Other option to avoid an ill-conditioned FIM is to reduce
244 the dimension of the matrix by splitting the full problem into subproblems (as we did for
245 the dynamic OED). Our computations indicated that for combinations of five parameters
246 $\binom{6}{5}$, five out of six possible combinations of parameters led to a well-conditioned FIM.

247 The combination that resulted in a singular FIM was $[\mu_{\max}, \alpha, T_{\min}, T_{\max}, T_{\text{opt}}]$. For
248 combinations of four parameters (FIM has dimension 4×4), the FIM was well-conditioned
249 for all the fifteen combinations.

250 It should be noted that in a simulation study performed in [35], full factorial design was
251 applied for a cardinal model describing the effects of temperature, pH and water activity
252 on the microbial growth rate, and the estimated values were close to the nominal values
253 used in the simulation. However, we should not be content only with this result, since the
254 actual values need to be supported by their corresponding confidence intervals in order
255 to identify practical identifiability problems and to provide a quantitative measurement
256 of the accuracy of the estimation.

257 We note that the composite factorial design does provide a well-conditioned FIM.
258 However the determinant of the FIM for this design is much lower than that obtained
259 with the D-optimal design, and this is actually reflected on the accuracy of the estimates.
260 The second row of Table 4 shows the ratio of the standard deviations of the parameters
261 obtained with the D-optimal design to the standard deviations obtained with the compos-
262 ite design. D-optimal design provides lower standard deviations, 36% better in average.
263 This result establishes the benefit of designing optimal experiments with OED techniques
264 for obtaining accurate parameter estimates.

265 *4.2. OED by the dynamic approach*

266 As it was mentioned in Section 3, the dynamic OED problem (10-11) was solved via
267 the simultaneous and CVP approaches. While the CVP method reduces substantially
268 the dimension of the original optimization problem, the simultaneous approach allows to
269 find solutions without restricting the shape of the controls. These solutions potentially
270 give better objective values, but may not be fit for practical use, if the controls have
271 a complicated shape. Comparing the two methods also give a hint at what we lose by
272 restricting the control shape to simple functions.

273 The CVP and the simultaneous approaches were compared for the case when the full
274 problem was partitioned into nine subproblems devoted to improve the accuracy of the
275 estimation of a couple of parameters. Overall, the CVP and the simultaneous methods

276 find very similar solutions with the exception of the experiment 7 (see Fig. 3). The
277 controls found with the simultaneous approach are often quite close in shape to piecewise
278 linear functions. This confirms that our choice of shape for the controls in the CVP was
279 a sensible one.

280 Table 5 compares the optimality cost functions provided by the simultaneous (J_{Sim})
281 and the CVP (J_{CVP}) methods. For all the nine experiments, the simultaneous approach
282 converges to better solutions than the CVP ones. However, the CVP approach provides
283 optimality cost functions very close to those obtained with the simultaneous approach.
284 In average, the cost functions obtained by the CVP approach are 95% of those obtained
285 with the simultaneous approach.

286 From the study of the subproblems, we can conclude that the CVP approach with a
287 simple piecewise parametrization seems well suited to design highly informative exper-
288 iments. We now apply the CVP approach to the full OED problem, with the FIM of
289 dimension 6×6 . The optimal experiment inputs obtained are displayed in Fig. 4. The
290 simulation of the nine D-optimal experiments is displayed in Fig 5B. Note that in the
291 experiment 9, the biomass concentration exhibits, for a certain time interval, a behavior
292 close to the steady state. This is due to the fact that the temperature reaches a very
293 close value to T_{max} and thus the growth rate becomes close zero. When performing the
294 experiments, caution should be made for the selection of the maximal operational temper-
295 ature. Indeed, an erroneous *a priori* on T_{max} with a higher value than the real maximal
296 temperature would lead to cell inactivation [13]. For microalgae cultures, T_{max} must be
297 well characterized to avoid operations that can be detrimental for attaining maximal pro-
298 ductivities [38]. In our case study, we were conservative in the selection of the *prior* of
299 T_{max} . By setting the *prior* lower than the maximal value reported in [14], we assured that
300 the temperature will allow growth in all the experiments.

301 Additionally, we wanted to assess the accuracy of the estimates when applying the
302 optimal solutions obtained from the nine small subproblems to the full OED. Table 6 shows
303 the ratio of the standard deviations of the estimates obtained from the full OED solutions
304 to those obtained from the solutions of the OED subproblems. We observe that the
305 standard deviations obtained when solving the full OED problem are usually smaller than

306 those for the partitioned subproblems, 70% in average. For T_{\max} the estimated standard
 307 deviations of the two approaches are very close. For the light associated parameters, in
 308 particular, the accuracy of the estimation provided by the solution of the full OED problem
 309 is substantially better. A higher determinant of the FIM (two orders of magnitude) is
 310 obtained with the full OED solution while the condition numbers for the partitioned and
 311 the full OED are of the same magnitude. On the other hand, we also see that partitioning
 312 the full OED problem into small subproblems gives satisfactory results. This strategy of
 313 simplification could be easier to implement when dealing with the full model, even if a
 314 better accuracy is achieved by considering the full FIM.

315 4.3. Static vs Dynamic OED

316 A complete comparison between the static and the dynamic approaches for OED
 317 would require the knowledge on the noise variance for the measurements of the maximal
 318 growth rate (σ_d^2) and the biomass concentration (σ_s^2). However, even if this information
 319 is unknown *a priori* we can still draw a comparative analysis of the performance of these
 320 methods, assuming that the data is generated by (5).

321 The unbiased estimator of the noise variance reads as

$$\sigma^2 = \frac{1}{n - n_p} \sum_{i=1}^n [y_i - y_{m_i}(\boldsymbol{\theta}^*)]^2, \quad (17)$$

322 with n the total number of data measurements and n_p the number of parameters . Since
 323 the ML estimator is efficient asymptotically (as $n \rightarrow \infty$), we can infer that the dynamic
 324 approach provides a more efficient estimator than the static approach. Indeed, for our case
 325 study, while the number of data points in the static approach is only 3 times the number
 326 of parameters, when applying the dynamic approach we get a number of experimental
 327 points that is 30 times the number of parameters.

328 To allow for a quantitative comparison, we used the approximated noise variances
 329 previously estimated $\sigma_d^2 = 9.31$ and $\sigma_s^2 = 3.8 \cdot 10^{-3}$ to generate random simulated data for
 330 tackling the parameter estimation problem for both methods. Table 7 shows the estimated
 331 values and their confidence intervals for both approaches. The standard deviations of the
 332 parameters obtained with the dynamic approach are in average 42% lower than those
 333 given by the static approach. For the parameters μ_{\max} , I_{opt} the dynamic approach excels

334 substantially the static approach by providing standard deviations 13% lower. Finally, it
 335 is worth noting that for the static approach to equal in average the dynamic approach, it
 336 is required to reduce significantly the value of σ_s , which is only possible for $n \gg n_p$.

337 The correlation matrix of the estimates for the dynamic approach was

$$\begin{array}{rcccccc}
 \mu_{\max} & 1.0 & & & & & \\
 \alpha & -0.47 & 1.0 & & & & \\
 I_{\text{opt}} & -0.08 & -0.36 & 1.0 & & & \\
 T_{\text{min}} & 0.34 & -0.04 & -0.07 & 1.0 & & \\
 T_{\text{max}} & -0.19 & 0.10 & -0.04 & 0.33 & 1.0 & \\
 T_{\text{opt}} & -0.14 & 0.01 & 0.05 & -0.35 & -0.34 & 1.0
 \end{array}$$

338 For the static approach, The correlation matrix of the estimates was

$$\begin{array}{rcccccc}
 \mu_{\max} & 1.0 & & & & & \\
 \alpha & -0.47 & 1.0 & & & & \\
 I_{\text{opt}} & -0.32 & 0.32 & 1.0 & & & \\
 T_{\text{min}} & 0.49 & -0.19 & -0.15 & 1.0 & & \\
 T_{\text{max}} & -0.07 & 0.03 & 0.03 & 0.18 & 1.0 & \\
 T_{\text{opt}} & -0.21 & 0.04 & -0.02 & -0.39 & -0.33 & 1.0
 \end{array}$$

339 The condition numbers of both approaches are comparable (see Table 4 and Table 6).
 340 The correlation matrices for both approaches indicated a low correlation between the
 341 parameters despite the high condition numbers . This is indeed thanks to the practical
 342 identifiability properties of the cardinal model as discussed in Appendix A.

343 For the dynamic approach the mean squared error (MSE) of the estimated parameters
 344 w.r.t the nominal parameters was 1.42 while for the static approach $\text{MSE} = 2.30 \cdot 10^3$,
 345 indicating the dynamic approach provides closest estimates to the nominal values in com-
 346 parison to the static approach. Only for T_{max} , both approaches perform equally.

347 Practical aspects as the labor of performing a two-step identification [12] place the
 348 static approach in disadvantage compared to the dynamic approach. These reasons lead
 349 us to favor the dynamic approach. Another benefit is that the sampling times could be
 350 further optimized within the experimental protocol, giving additional degrees of freedom.

351 From the mechanistic point of view, by stimulating the system with time-varying in-
352 puts, the dynamic approach allows a better characterization of the system behavior. On
353 the opposite, the static approach can hide the relevance of certain important phenonema.
354 This factor is critical to our case study where microalgae are meant to grow in a dynamic
355 environment that is periodically forced by daily variations of light and temperature. How-
356 ever, we should keep in mind that to take advantage fully of the dynamic approach, a
357 step forward in the mathematical description of the process needs to be done for account-
358 ing important phenomena such as acclimation to light and temperature [17, 36] and cell
359 inactivation due to high temperatures. For the sake of generality, a further study should
360 be done to include the impact of the initial conditions and the physiological state of the
361 cells on the determination of optimal experiment inputs. We also recommend to perform
362 a preliminary experiment for which the cells get acclimated to their light and temper-
363 ature growth conditions. This experiment will allow a dynamic characterization of the
364 adaptation phenomena.

365 4.4. Relevance of accurate estimation on model-driven optimization

One of the ultimate goals of developing microalgae growth models is to provide a platform for model predictions and for the design of optimal control strategies for systems operated at large scale. Following this aim, we wanted to assess the relevance of providing accurate parameter estimates on the quality of the predictions for a more complex model representing the continuous cultivation of microalgae on an outdoor pond. For that, we used the raceway model described in [3]. The model takes the configuration of a pilot-scale raceway (Algotron) located at INRA LBE, Narbonne (France). The model is described by

$$\dot{s} = D(s_{in} - s) - \rho(\cdot)x, \quad (18)$$

$$\dot{q}_n = \rho(\cdot) - (\mu(\cdot) - R(\cdot))q_n, \quad (19)$$

$$\dot{x} = (\mu(\cdot) - D - R(\cdot))x, \quad (20)$$

366 where s (mg N/L) is the extracellular nitrogen concentration and x (mg C/L) is the
367 concentration of carbon biomass. The term q_n (g N/g C) denotes the intracellular nitrogen

368 quota, that is the concentration of nitrogen per biomass unit. D is the dilution rate, $\mu(\cdot)$
 369 is the specific growth rate, ρ is the nitrogen uptake rate and $R(\cdot)$ the respiration rate. For
 370 more details, the reader is referred to [3].

371 Firstly, we evaluated the sensitivity of the biomass concentration with respect to the
 372 model parameters along a year of operation. Meteorological data was used for the location
 373 of Narbonne to calculate the temperature and light intensities for each month. The
 374 normalized sensitivity matrix \mathbf{S}_y was computed for each month. The element (k, j) of \mathbf{S}_y
 375 is calculated as [37]

$$\mathbf{S}_y(k, j) = \sum_{i=1}^{n_t} \left| \bar{s}_j^k(t_i, \hat{\boldsymbol{\theta}}) \right|, \quad (21)$$

376 where \bar{s}_j^k is the normalized sensitivity of the model output y_{m_k} w.r.t. θ_j ,

$$\bar{s}_j^k(t_i, \hat{\boldsymbol{\theta}}) = \frac{\hat{\theta}_j}{y_{m_k}(t_i, \hat{\boldsymbol{\theta}})} \left[\frac{\partial y_{m_k}}{\partial \theta_j} \right]_{(t_i, \hat{\boldsymbol{\theta}})}. \quad (22)$$

377 Figure 6 shows a graphical representation of the sensitivity matrices for four months.
 378 January is the coldest month in Narbonne, while August is the warmest. October is
 379 an intermediate month. The sensitivities of June are also presented for illustration. It
 380 is interesting to observe that the influence of the parameters on the model response is
 381 modulated by the environmental conditions. Indeed, we can see a specific pattern of
 382 parameter influence for each month. In terms of the tuning importance, that is the
 383 importance of parameter changes around their nominal value for the model output [37],
 384 we observe that, overall, the most dominant parameter is T_{opt} . In August the most
 385 dominant parameter is μ_{max} . This month exhibits the most homogeneous distribution of
 386 the influence of parameters in comparison with months like January where the influence of
 387 two parameters ($T_{\text{max}}, T_{\text{opt}}$) exceeds substantially the influence of the rest of parameters.
 388 In cold months (*e.g.* January-March), the influence of T_{max} is higher than the influence
 389 of μ_{max} . This pattern is switched in warm months (*e.g.* June-August). In the figure, T_{min}
 390 appears as the less influential parameter. This effect may be inverted in cold regions.
 391 Indeed, the minimum average temperature in Narbonne used in our simulation is 4.76 °C,
 392 which is very high compared to the nominal value of T_{min} .

393 Figure 7 shows the dramatic effect of an uncertainty of 5% on the nominal value of
394 T_{opt} on the quality of model predictions for the month of January. Both overestimation
395 ($1.05T_{\text{opt}}$) and underestimation ($0.95T_{\text{opt}}$) of the optimal temperature results in important
396 discrepancies between the response of the model with the nominal value of T_{opt} and those
397 obtained with a small perturbation of 5% on the nominal value. Hence the importance of
398 providing accurate estimates since small changes on the parameter values can induce large
399 changes on the biomass dynamics. Model-driven decisions are thus strongly dependent
400 on the accuracy of the parameter estimates.

401 The previous result strengthens the relevance of the temperature effect for outdoor
402 cultivation as discussed in [38]. It should be noted that with the meteorological data used
403 here, the temperature of the culture (T) never exceeded T_{max} , so the effect of temperature
404 ϕ_T was always higher than zero. We recalled that an overestimation on T_{max} will have
405 a strong impact on model predictions and system operation. In particular, when the
406 temperature exceeds T_{max} phenomena as cell inactivation and mortality take place. These
407 phenomena, detrimental for attaining maximal productivities, need to be characterized
408 by an approach combining both experiments and modelling in order to provide guidelines
409 to mitigate negative effects.

410 5. Conclusions

411 We solved the OED problem for a simplified model of microalgae growth. We have
412 determined optimal experiment conditions to provide an accurate estimation of the pa-
413 rameters that drive microalgae growth by modulating the effects of light and temperature.
414 Both static and dynamic approaches were evaluated to find D-optimal experiments. From
415 our results, we recommend the use of the dynamic approach in virtue of the efficacy in
416 terms of the maximum likelihood properties of the estimator. The protocol of experiment
417 inputs determined in this study will be further implemented in the TIP system used at
418 Ifremer Institute.

419 For the dynamic case, we showed that a parameterization of the control input by
420 piecewise linear functions (CVP approach) provides efficient results as compared as the
421 simultaneous approach. Moreover, the strategy of partitioning the full OED problem

422 into subproblems dedicated to improve the accuracy of the estimation of a couple of
423 parameters was shown to be satisfactory. The CVP method and the partitioning of
424 the full OED into subproblems are suitable approaches for solving the OED problem in
425 microalgae growth models by reducing, additionally, the problem complexity. Finally,
426 with the use of sensitivity analysis of a more complex model describing the cultivation of
427 microalgae in a raceway, we showed the relevance of providing accurate parameters for
428 enabling reliable model-driven decisions.

429 **6. Acknowledgment**

430 This work benefited from the support of the Facteur 4 research project founded by
431 the French National Research Agency (ANR).

432 **References**

- 433 [1] P. J. B. Williams, L. M. L. Laurens, Microalgae as biodiesel & biomass feedstocks:
434 Review & analysis of the biochemistry, energetics & economics, *Energy and Environ.*
435 *Sci.* 3 (2010) 554–590.
- 436 [2] C. Posten, Design principles of photo-bioreactors for cultivation of microalgae, *Eng.*
437 *Life Sci.* 9 (2009) 165–177.
- 438 [3] R. Muñoz-Tamayo, F. Mairet, O. Bernard, Optimizing microalgal production in race-
439 way systems., *Biotechnol Prog* 29 (2013) 543–552.
- 440 [4] D. A. Pereira, V. O. Rodrigues, S. V. Gómez, E. A. Sales, O. Jorquera, Parametric
441 sensitivity analysis for temperature control in outdoor photobioreactors, *Bioresource*
442 *technology* 144 (2013) 548–553.
- 443 [5] G. Goodwin, R. Payne, *Dynamic System Identification: Experiment Design and Data*
444 *Analysis*, Academic Press, New York, 1977.
- 445 [6] E. Walter, L. Pronzato, *Identification of Parametric Models from Experimental Data*,
446 Springer, London, 1997.

- 447 [7] K. Keesman, System Identification: an Introduction, Springer-Verlag, London,, 2011.
- 448 [8] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter
449 precision: state of the art, Chemical Engineering Science 63 (19) (2008) 4846–4872.
- 450 [9] J. Stigter, K. Keesman, Optimal parametric sensitivity control of a fed-batch reactor,
451 Automatica 40 (2004) 1459 – 1464.
- 452 [10] I. Bauer, H. G. Bock, S. Körkel, J. P. Schlöder, Numerical methods for optimum
453 experimental design in DAE systems, Journal of Computational and Applied Math-
454 ematics 120 (2000) 1–25.
- 455 [11] E. Balsa-Canto, A. A. Alonso, J. Banga, Computational procedures for optimal ex-
456 perimental design in biological systems, IET Syst. Biol. 2 (2008) 163–172.
- 457 [12] K. Bernaerts, K. J. Versyck, J. F. Van Impe, On the design of optimal dynamic exper-
458 iments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal
459 temperatures, International journal of food microbiology 54 (1) (2000) 27–38.
- 460 [13] K. Bernaerts, K. P. M. Gysemans, T. Nhan Minh, J. F. Van Impe, Optimal experi-
461 ment design for cardinal values estimation: guidelines for data collection., Int J Food
462 Microbiol 100 (2005) 153–165.
- 463 [14] J. Marchetti, G. Bougaran, L. L. Dean, C. Mégrier, E. Lukomska, R. Kaas, E. Olivo,
464 R. Baron, R. Robert, J. P. Cadoret, Optimizing conditions for the continuous culture
465 of *Isochrysis affinis galbana* relevant to commercial hatcheries, Aquaculture 326-329
466 (2012) 106–115.
- 467 [15] O. Bernard, Hurdles and challenges for modelling and control of microalgae for CO₂
468 mitigation and biofuel production, Journal of Process Control 21 (2011) 1378–1389.
- 469 [16] O. Bernard, B. Remond, Validation of a simple model accounting for light and tem-
470 perature effect on microalgal growth, Bioresour Technol 123 (2012) 520–527.

- 471 [17] R. J. Geider, Light and temperature dependence of the carbon to chlorophyll a
472 ratio in microalgae and cyanobacteria: implications for physiology and growth of
473 phytoplankton, *New Phytologist* 106 (1987) 1–34.
- 474 [18] L. Rosso, J. R. Lobry, J. P. Flandrois, An unexpected correlation between cardinal
475 temperatures of microbial growth highlighted by a new model., *J Theor Biol* 162 (4)
476 (1993) 447–463.
- 477 [19] J. C. H. Peeters, P. Eilers, The relationship between light intensity and photosyn-
478 thesis—A simple mathematical model, *Hydrobiological Bulletin* 12 (1978) 134–136.
- 479 [20] E. M. Bendif, I. Probert, D. C. Schroeder, C. de Vargas, On the description of
480 *Tisochrysis lutea* gen. nov. sp. nov. and *Isochrysis nuda* sp. nov. in the Isochrysidales,
481 and the transfer of *Dicrateria* to the Prymnesiales (Haptophyta), *Journal of Applied*
482 *Phycology* .
- 483 [21] F. Mairet, O. Bernard, P. Masci, T. Lacour, A. Sciandra, Modelling neutral lipid pro-
484 duction by the microalga *Isochrysis aff. galbana* under nitrogen limitation., *Bioresour*
485 *Technol* 102 (2011) 142–149.
- 486 [22] S. M. Renaud, L.-V. Thinh, G. Lambrinidis, D. L. Parry, Effect of temperature
487 on growth, chemical composition and fatty acid composition of tropical Australian
488 microalgae grown in batch cultures, *Aquaculture* 211 (1) (2002) 195–214.
- 489 [23] R. Schenkendorf, A. Kremling, M. Mangold, Optimal experimental design with the
490 sigma point method, *IET Syst. Biol* 3 (2009) 10–23.
- 491 [24] D. De Pauw, Optimal experimental design for calibration of bioprocess models: a
492 validated software toolbox, Ph.D. thesis, University of Ghent, 2005.
- 493 [25] R. Muñoz-Tamayo, B. Laroche, M. Leclerc, E. Walter, IDEAS: a Pa-
494 rameter Identification Toolbox with Symbolic Analysis of Uncertainty and
495 its Application to Biological Modelling, in: *Preprints of the 15th IFAC*
496 *Symposium on System Identification*, Saint-Malo, France, 1271–1276, URL
497 <http://www.inra.fr/miaj/public/logiciels/ideas/index.html>, 2009.

- 498 [26] T. Lacour, A. Sciandra, A. Talec, P. Mayzaud, O. Bernard, Diel variations of car-
499 bohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat
500 cultures of *Isochrysis* sp., *Journal of Phycology* 48 (2012) 966–975.
- 501 [27] F. Mairet, O. Bernard, T. Lacour, A. Sciandra, Modelling microalgae growth in
502 nitrogen limited photobioreactor for estimating biomass, carbohydrate and neutral
503 lipid productivities, in: *Proc. 18th World Congress The International Federation of*
504 *Automatic Control*, Milano, Italy, 10591–10596, 2011.
- 505 [28] E. Van Derlinden, K. Bernaerts, J. F. Van Impe, Accurate estimation of cardinal
506 growth temperatures of *Escherichia coli* from optimal dynamic experiments., *Int J*
507 *Food Microbiol* 128 (2008) 89–100.
- 508 [29] B. Chachuat, A. B. Singer, P. I. Barton, *Global Methods for Dynamic Optimization*
509 *and Mixed-Integer Dynamic Optimization*, *Ind. Eng. Chem.* 45 (2006) 8373–8392.
- 510 [30] L. T. Biegler, *Nonlinear Programming: Concepts, Algorithms, and Applications to*
511 *Chemical Processes*, Society for Industrial and Applied Mathematics and the Math-
512 *ematical Optimization Society*, Philadelphia,, 2010.
- 513 [31] J. Bonnans, Frédéric, P. Martinon, V. Grélard, Bocop - A collection of examples,
514 *Tech. Rep.*, INRIA, URL <http://hal.inria.fr/hal-00726992>, rR-8053, 2012.
- 515 [32] A. Wächter, T. Biegler, On the implementation of an interior-point filter line-search
516 algorithm for large-scale nonlinear programming, *Mathematical Programming* 106
517 (2006) 25–57.
- 518 [33] M. Rodríguez-Fernández, J. A. Egea, J. R. Banga, Novel metaheuristic for parameter
519 estimation in nonlinear dynamic biological systems., *BMC Bioinformatics* 7 (2006)
520 483.
- 521 [34] J. A. Egea, M. Rodríguez-Fernández, J. R. Banga, R., Martí, Scatter search for
522 chemical and bio-process optimization, *Journal of Global Optimization* 37 (2007)
523 481–503.

- 524 [35] E. Van Derlinden, L. Mertens, J. F. Van Impe, The impact of experiment design on
525 the parameter estimation of cardinal parameter models in predictive microbiology,
526 Food Control 29 (2013) 300–308.
- 527 [36] D. Zou, K. Gao, Thermal acclimation of respiration and photosynthesis in the marine
528 macroalga *Gracilaria lemaneiformis* (Gracilariales, Rhodophyta), Journal of Phycol-
529 ogy .
- 530 [37] T. Turányi, Sensitivity analysis of complex kinetic systems. Tools and applications,
531 Journal of Mathematical Chemistry 5 (3) (1990) 203–248.
- 532 [38] M. Ras, J.-P. Steyer, O. Bernard, Temperature effect on microalgae: a crucial factor
533 for outdoor production, Reviews in Environmental Science and Biotechnology 12
534 (2013) 153–164.
- 535 [39] H. Pohjanpalo, System Identifiability Based on the Power Series Expansion of the
536 Solution, Math Biosci 41 (1978) 21–33.



Figure 1: The TIP system. The device has 18 batch photobioreactors for microalgae cultivation.

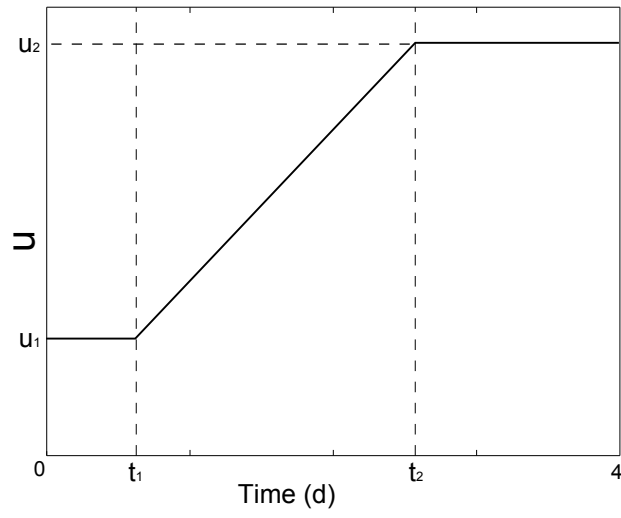


Figure 2: Parameterization of the experiment inputs $u(T, I)$ for the CVP approach.

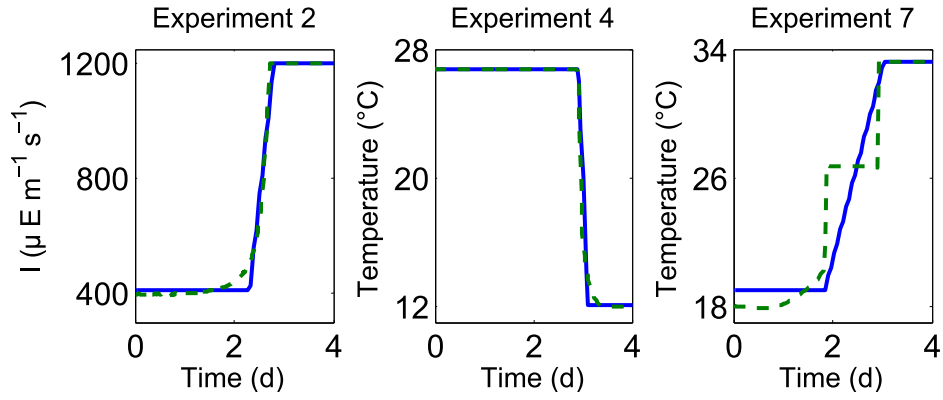


Figure 3: Optimal experiment inputs given by the CVP approach (solid lines) and the simultaneous approach (dashed lines) for the partitioned OED problem.

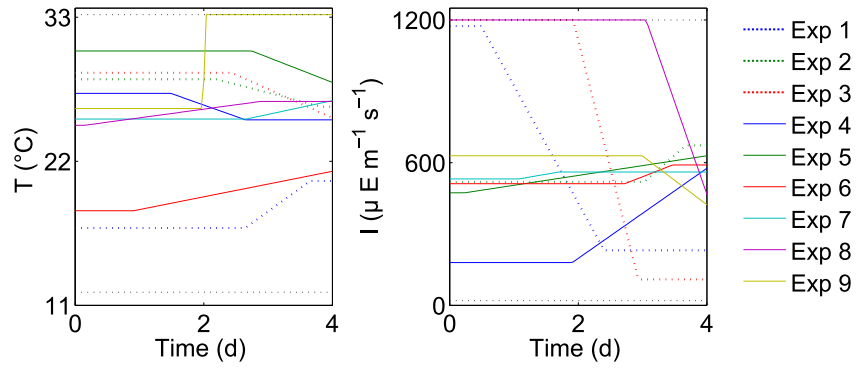


Figure 4: Optimal experiment inputs obtained for the full OED problem.

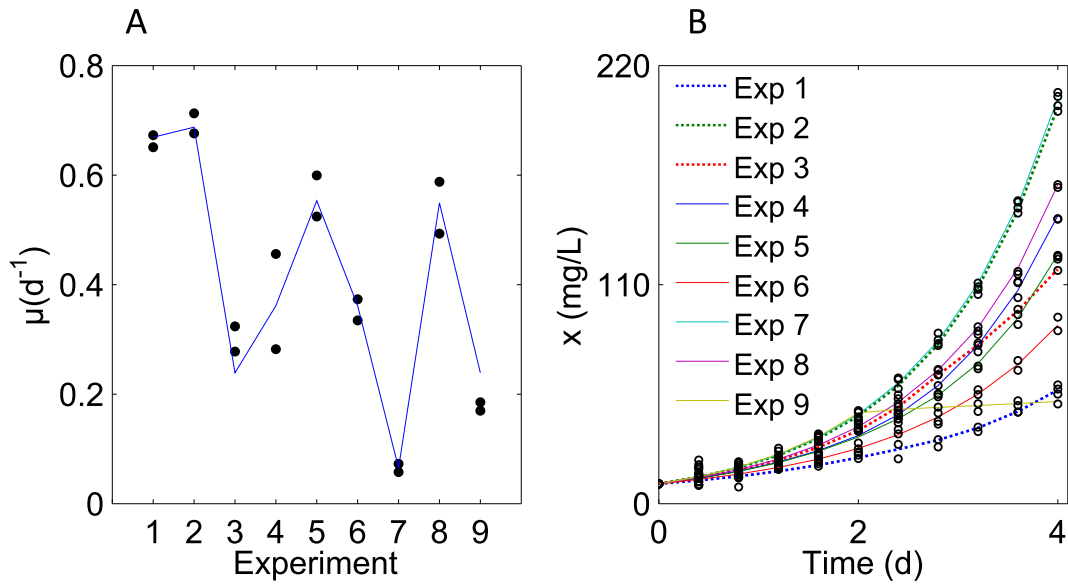


Figure 5: Simulated data resulted from D-optimal experiments including duplicates and responses of the identified models. A. Measurements of specific growth (circles) for the static approach. B. Measurements of biomass concentrations (circles) for the dynamic approach obtained for the full OED problem. The responses of the identified models (solid and dotted lines) for both static and dynamic approaches described satisfactorily the simulated data.

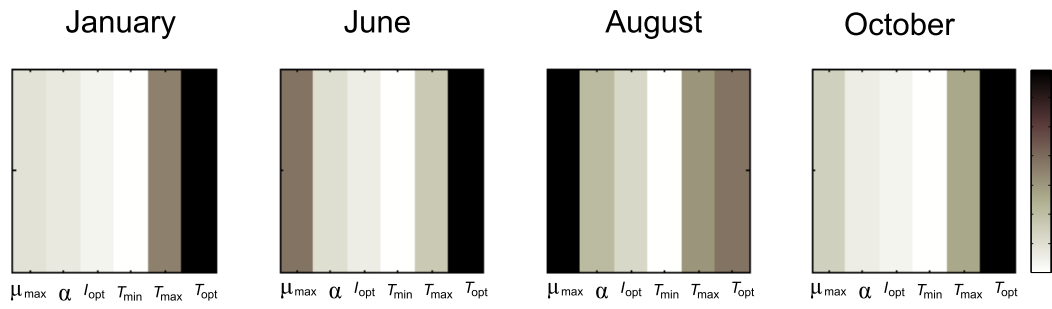


Figure 6: Overall sensitivity of the biomass concentration to the parameters in the complete raceway model developed in [3]. Results are shown for four months illustrating how the influence of the parameters on the model output is modulated by the environmental conditions.

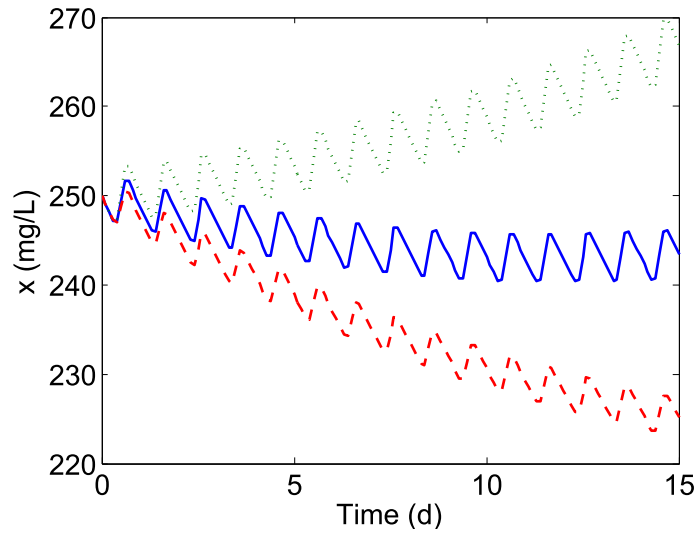


Figure 7: A small uncertainty of 5% on the value of T_{opt} leads to important mismatches on model predictions. The dynamic of the biomass concentration of the month of January with the nominal value of T_{opt} (solid blue line) is compared to the response of the model with $0.95T_{opt}$ (dotted green line) and $1.05T_{opt}$ (dashed red line) .

Table 1: Nominal values of the model parameters.

Parameter	Definition	Value
α	Initial slope of the growth response curve w.r.t. light	$0.008 (\mu\text{E m}^{-2}\text{s}^{-1} \text{d})^{-1}$
μ_{\max}	Maximal specific growth rate	0.76 d^{-1}
I_{opt}	Optimal light intensity	$548 \mu\text{E m}^{-2}\text{s}^{-1}$
T_{\min}	Lower temperature for microalgae growth	$-0.20 \text{ }^\circ\text{C}$
T_{\max}	Upper temperature for microalgae growth	$33.30 \text{ }^\circ\text{C}$
T_{opt}	Temperature at which growth rate is maximal	$26.70 \text{ }^\circ\text{C}$

Table 2: Subproblems of the dynamic OED strategy.

Experiment	Couple of parameters	Experiment input
1	(μ_{\max}, α)	I
2	$(\mu_{\max}, I_{\text{opt}})$	I
3	(α, I_{opt})	I
4	(μ_{\max}, T_{\min})	T
5	(μ_{\max}, T_{\max})	T
6	$(\mu_{\max}, T_{\text{opt}})$	T
7	(T_{\min}, T_{\max})	T
8	$(T_{\min}, T_{\text{opt}})$	T
9	$(T_{\max}, T_{\text{opt}})$	T

Table 3: Experimental conditions for the static approach.

Full factorial design									
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	12.0	22.60	33.20	12.0	22.60	33.20	12.0	22.60	33.20
Light intensity ($\mu\text{E m}^{-2}\text{s}^{-1}$)	20	20	20	610	610	610	1200	1200	1200
Composite factorial design									
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	15.30	19.0	19.0	19.0	19.0	24.5	24.5	24.5	24.5
Light intensity ($\mu\text{E m}^{-2}\text{s}^{-1}$)	560	863	257	257	863	560	560	560	560
	10	11	12	13	14	15	16	17	
Temperature (°C)	24.50	24.50	24.50	30.0	30.0	30.0	30.0	33.20	
Light intensity ($\mu\text{E m}^{-2}\text{s}^{-1}$)	560	50	1070	863	257	257	863	560	
D-optimal design									
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	12.10	12.10	24.30	24.60	26.70	26.70	30.60	30.70	33.20
Light intensity ($\mu\text{E m}^{-2}\text{s}^{-1}$)	536	536	1200	409	74	74	1200	395	547

Table 4: Comparison of D-optimal design with factorial design for the static approach.

	μ_{\max}	α	I_{opt}	T_{\min}	T_{\max}	T_{opt}
$\frac{\eta_j \text{D-optimal}}{\eta_j \text{Composite factorial design}}$	0.50	0.48	0.76	0.16	0.02	0.25
	$\text{Det}(\mathbf{F}_s)$		$\lambda_{\max}/\lambda_{\min}$			
Full factorial design	0		$5.30 \cdot 10^{20}$			
Composite factorial design	381.74		$1.89 \cdot 10^9$			
D-optimal	$7.90 \cdot 10^6$		$3.85 \cdot 10^9$			

Table 5: Comparison of the CVP and sequential strategies for the partitioned OED problem in the dynamic approach.

Experiment	$-\log \text{Det}(\mathbf{F}_d)$		
	Simultaneous (J_{Sim})	CVP (J_{cvp})	$J_{\text{cvp}}/J_{\text{Sim}}$
1	-26.1949	-25.8114	0.9854
2	-5.37552	-5.0026	0.9306
3	-10.2537	-9.8398	0.9596
4	-11.3121	-10.8970	0.9633
5	-16.9941	-16.5543	0.9741
6	-14.856	-14.6943	0.9891
7	-5.0363	-4.4793	0.8894
8	-4.5046	-4.0965	0.9094
9	-9.9769	-9.4997	0.9522

Table 6: Comparison of the accuracy of the estimation obtained with the solutions of the full and partitioned OED problems in the dynamic approach.

	μ_{\max}	α	I_{opt}	T_{\min}	T_{\max}	T_{opt}
$\frac{\eta_j^{\text{Full}}}{\eta_j^{\text{Partitioned}}}$	0.85	0.57	0.41	0.76	1.05	0.59
	Det(\mathbf{F}_d)			$\lambda_{\max}/\lambda_{\min}$		
Full OED	$1.76 \cdot 10^{12}$			$1.10 \cdot 10^9$		
Partitioned OED	$4.20 \cdot 10^{10}$			$4.44 \cdot 10^9$		

Table 7: Estimated parameters with their approximate confidence intervals for the static and dynamic OED approaches. The parameter estimation was performed with simulated noisy data.

	$\hat{\theta} \pm 2\eta_j$					
	μ_{\max}	α	I_{opt}	T_{min}	T_{max}	T_{opt}
Static	0.74±0.070	0.0075±0.0029	665.03±274.52	-0.86±4.84	33.34±0.30	26.80±0.97
Dynamic	0.76±0.0091	0.008±0.0014	550.64±35.44	-0.26±2.80	33.34±0.28	26.66±0.25

537 **Appendix A. Comments on structural and practical identifiability of the model**

538 *Appendix A.1. Structural identifiability*

539 To check the structural identifiability of the model, we used the time power series
540 method [39], briefly described below.

Consider the following model

$$\dot{x}(t) = f(x(t), u(t), \boldsymbol{\theta}, t), \quad (\text{A.1})$$

$$y_m(t) = h(x(t), \boldsymbol{\theta}), \quad (\text{A.2})$$

541 with x the state vector, y_m the vector of model outputs(observations) and $u(t)$ the control
542 vector. The function vector $f(\cdot)$ is assumed to have infinitely many derivatives with
543 respect to time and the input and state vector components. In the same manner, $h(\cdot)$
544 is infinitely differentiable w.r.t. the state vector and $u(t)$ is infinitely differentiable w.r.t.
545 time. Both the state and the outputs are infinitely differentiable w.r.t. time. The outputs
546 can therefore be represented by the corresponding Taylor series expansion around $t = 0$.

547 Consider the k th time derivative (a_k) of the output.

$$a_k(\boldsymbol{\theta}) = \lim_{t \rightarrow 0^+} \frac{d^k}{dt^k} y_m(t). \quad (\text{A.3})$$

548 Since all the time derivatives of the outputs are unique, it follows that a sufficient
549 condition for the identifiability of the model is that set of equations

$$h^{(k)}(x(0), \boldsymbol{\theta}) = a_k(0) \quad (\text{A.4})$$

550 have a unique solution for $\boldsymbol{\theta}$ [39].

For our case study, let us consider first the identifiability of the temperature parameters
of the cardinal model $T_{\min}, T_{\max}, T_{\text{opt}}$. At constant light, the model is given by

$$\dot{x}(t) = \mu_I \phi_T(t) x(t), \quad x(0) = x_0, \quad (\text{A.5})$$

$$y_m(t) = x(t), \quad (\text{A.6})$$

551 with $\mu_I = \mu_{\max} \phi_I$ and x_0 a known initial concentration of biomass. The effects of light
552 ϕ_I and temperature ϕ_T on microalgae growth are here recalled:

$$\phi_I = \frac{I}{I + \frac{\mu_{\max}}{\alpha} \left(\frac{I}{I_{\text{opt}}} - 1 \right)^2}, \quad (\text{A.7})$$

553

$$\phi_T = \begin{cases} 0, & T < T_{\min} \\ \frac{(T-T_{\max})(T-T_{\min})^2}{(T_{\text{opt}}-T_{\min})[(T_{\text{opt}}-T_{\min})(T-T_{\text{opt}})-(T_{\text{opt}}-T_{\max})(T_{\text{opt}}+T_{\min}-2T)]}, & T \in [T_{\min}, T_{\max}] \\ 0, & T > T_{\max}. \end{cases} \quad (\text{A.8})$$

554

555

556

557

By simple inspection of (A.8) and given the biological meaning of the parameters of the cardinal model, we can infer that a series of adequate experiments running at different temperature conditions in the interval $[T_{\min}, T_{\max}]$ will allow to identify uniquely the temperature parameters.

558

559

560

Coming back to the time power series method, let us consider the case of a specific input $T(t)$ that is infinitely differentiable w.r.t. to time. For simplicity, we chose $T(t) = c_1 t + c_2$ with $c_1 > 0, c_2 > 0$ and $T(t) \in [T_{\min}, T_{\max}]$ in the experimentation time.

The Taylor series coefficients are thus

$$a_0 = x_0, \quad (\text{A.9})$$

$$a_1 = \mu_I \phi_T(0) x_0, \quad (\text{A.10})$$

$$a_2 = \left((\mu_I \phi_T)^2 + \mu_I \left[\frac{\partial \phi_T}{\partial T} \right]_{T=T(0)} \dot{T}(0) \right) x_0. \quad (\text{A.11})$$

Given the shape of ϕ_T and applying the first-optimality condition, the following cases provide the parameters to be uniquely identifiable:

$$a_1 = 0, \text{ and, } a_2 \geq 0, \Rightarrow T_{\min} = T(0), \quad (\text{A.12})$$

$$a_1 = 0, \text{ and, } a_2 < 0, \Rightarrow T_{\max} = T(0), \quad (\text{A.13})$$

$$a_1 > 0, \text{ and, } a_2 = a_1^2/x_0, \Rightarrow T_{\text{opt}} = T(0). \quad (\text{A.14})$$

561

The previous conditions can be reached by making $T(t)$ vary along the interval $[T_{\min}, T_{\max}]$.

Following the same procedure, we can now check the identifiability of the light parameters I_{opt}, α and the maximal specific growth rate μ_{\max} . Since T_{opt} is structurally identifiable, let us consider a constant temperature $T = T_{\text{opt}}$. The model is thus

$$\dot{x}(t) = \mu_{\max} \phi_I(t) x(t), \quad x(0) = x_0, \quad (\text{A.15})$$

$$y_m(t) = x(t). \quad (\text{A.16})$$

562 By considering the case of a specific input $I(t) = c_1 t + c_2$, the series expansion provides
 563 the following coefficients

$$a_0 = x_0, \quad (\text{A.17})$$

$$a_1 = \mu_{\max} \phi_I(0) x_0, \quad (\text{A.18})$$

$$a_2 = \left((\mu_{\max} \phi_I(0))^2 + \mu_{\max} \left[\frac{\partial \phi_I}{\partial I} \right]_{I=I(0)} \dot{I}(0) \right) x_0. \quad (\text{A.19})$$

564 By applying the first-order optimality condition on ϕ_I , we get

$$a_1 > 0, \text{ and } a_2 = a_1^2/x_0, \Rightarrow I_{\text{opt}} = I(0). \quad (\text{A.20})$$

565 Injecting I_{opt} in (A.18) provides μ_{\max} .

566 Once I_{opt} is identified, evaluating (A.18) at $I(0) \neq I_{\text{opt}}$ provides α

$$\alpha = \frac{a_1 \mu_{\max}}{I(0)(\mu_{\max} x_0 - a_1)} \left(\frac{I(0)}{I_{\text{opt}}} - 1 \right)^2. \quad (\text{A.21})$$

567 The model is therefore structurally identifiable.

568 *Appendix A.2. Practical identifiability*

569 Parameter estimation of Haldane and Monod type kinetics is known to be hampered
 570 by practical identifiability problems due to strong correlation between its parameters. To
 571 represent the effect of light on microalgae growth, the Haldane kinetics is often used

$$\phi_I = \tilde{\mu} \frac{I}{I + K_{sI} + I^2/K_{iI}}, \quad (\text{A.22})$$

572 where $\tilde{\mu}$ is the specific growth rate, K_{sI} is the light affinity constant and K_{iI} is the
 573 inhibition constant. By applying the first-order optimality condition, the optimal light
 574 intensity for growth is $I_{\text{opt}} = \sqrt{K_{sI} K_{iI}}$. The nominal values for the Haldane model used in
 575 the present study were $\tilde{\mu} = 1.18 \text{ d}^{-1}$, $K_{sI} = 150 \mu\text{E m}^{-2}\text{s}^{-1}$ and $K_{iI} = 2000 \mu\text{E m}^{-2}\text{s}^{-1}$.

576 In this work, instead of the standard Haldane kinetics, we made use of the param-
 577 eterized kinetics (A.7), which has the same shape than the Haldane kinetics but offers
 578 certain advantages in terms of practical identifiability properties. On the basis of good
 579 quality nominal parameters, it was previously shown that adequate inputs allow to iden-
 580 tify the optimal conditions for growth I_{opt} and T_{opt} , which derive automatically on the

581 identification of μ_{\max} . Such property allows to weaken the correlation between μ_{\max} and
582 α .

583 For illustration, we performed a D-optimal protocol consisted of five experiments with
584 ten equidistant sampling times for both standard Haldane and the parameterized kinetics.

585 For the Haldane kinetics, the correlation matrix of the parameters was:

$$\begin{array}{cccc} \tilde{\mu} & 1.0 & & \\ K_{sI} & 0.96 & 1.0 & \\ K_{iI} & -0.98 & -0.92 & 1.0 \end{array}$$

586 For the parameterized kinetics, the correlation matrix of the parameters was:

$$\begin{array}{cccc} \mu_{\max} & 1.0 & & \\ \alpha & -0.53 & 1.0 & \\ I_{\text{opt}} & -0.25 & -0.10 & 1.0 \end{array}$$

587 As observed, the Haldane kinetics exhibits a stronger parameter correlation than the
588 parameterized kinetics. Therefore, in terms of practical identifiability, the parameterized
589 kinetics is preferred over the standard Haldane kinetics.