Confidence regions of estimated parameters for ecological systems

S. Marsili-Libelli*, S. Guerrizio, N. Checchi
Department of Systems and Computers, University of Florence, Via S. Marta 3, 50139 Firenze, Italy
Received 22 May 2002; received in revised form 8 October 2002; accepted 11 February 2003

Abstract
The parameters of ecological models are usually estimated through numerical search algorithms. Determining confidence boundaries for the parameter values obtained in such a way is a problem of great practical importance. In this paper a method is proposed to estimate such regions in two ways, based on either the Hessian matrix or the Fisher Information Matrix (FIM). There is a conceptual difference in the two approximations: the FIM approach is based on the sensitivity trajectories, whereas the Hessian expansion depends on the shape of the error functional. From a comparison between the two approaches, a discriminating method is obtained to detect inaccurate estimation results. The Hessian and FIM approaches differ by the second derivative terms of the output function. This difference is used to assess the success of the estimation, because the two methods yield the same confidence estimate only if the search terminates at the optimal parameter value. The method is demonstrated with reference to a pair of widely used dynamics: the Monod kinetics and the Richards logistic function applied to algal growth. It is shown that in both cases this method compares favourably with the residual correlation analysis and appears to have more discriminatory power.

Keywords: Parameter estimation; Nonlinear optimisation; Ecological models; Fisher Information Matrix; Covariance; Statistics

1. Introduction
Parameter estimation of dynamical models is a major issue in numerical ecology. Model nonlinearity, data scarcity and non-Gaussian error distribution contribute to the complexity of the problem, which is usually solved by numerical means. Further, determining the “best” parameter values, according to some objective function, is only part of the problem, because it is equally important to attach a measure of confidence to the values thus obtained. Apart from early papers such as Heineken et al. (1967) and scattered but important contributions such as Holmberg (1982) and Donaldson and Schnabel (1987), very rarely has the matter been addressed per se, whereas there are many modelling papers which address this subject as a secondary problem. This justifies to some extent the wide range of methods being used and the lack of generality of the results. The effort of this paper is to bring the problem to the forefront and give it the importance it deserves.

The unifying point in the literature appears to be the concept of seeking a “valid” set of parameters in terms of residuals (van Tongeren, 1995), rather than searching a “true” model (Alewell and Manderscheid, 1998). Sometimes, bounds on parameter values were imposed as an alternative to probabilistic estimation (Norton, 1996). Only in the calibration of the very complex models, such as TERRA (Kercher and Chambers, 2001) describing the terrestrial biogeochemical cycles, the Jacobian matrix was used in the parameter estimation scheme, but not to compute an approximate confidence interval for the estimates.
Conversely, an interesting approach was presented by Håkanson (1996) who proposed a validation procedure involving model residuals and their correlation coefficients. Sensitivity analysis has been used extensively, though not in a homogeneous way. Omlin and Reichert (1999) introduced the sensitivity function as a means of approximating the parameter confidence interval in the context of Bayesian estimation and Omlin et al. (2001) made use of the sensitivity functions to detect identifiability problems of a lake biogeochemical cycle. They use a quadratic functional to estimate a subset of parameters, but their uncertainty is estimated subjectively and grouped into three classes with predetermined range of 5, 20 and 50% of the estimated values.

The influence of model nonlinearities on the range of confidence regions was investigated by Klepper and Bedaux (1997) for simple ecotoxicological models. Stollenwerk et al. (2001) used univariate Gaussian distributions in the context of Fokker–Plank stochastic equations to model phytoplankton time series, but did not attempt to estimate the joint parameters confidence regions. Carrol and Warwick (2001) used it to determine uncertain parameters and identify their ranges in the context of mercury transport models.

Use of linear theory to compute an approximation of the covariance matrix was made by Einola et al. (1998) in the context of Extended Kalman Filter identification of zooplankton populations and more recently by Tang et al. (2001) and Tang and Wang (2002), who used the linear estimation theory as an approximation of the covariance matrix, but failed to discuss the implications of such simplifications nor considered the problem of multiple-dimensional confidence regions.

From this brief survey it appears that parameter uncertainty has been dealt in many different ways and that rarely the confidence intervals have been used to assess the validity of the estimation. It is the purpose of this paper to address this specific problem, and in this sense it may be regarded as a sequel to the previous one (Marsili-Libelli, 1992) presenting a numerical search method for the estimation of model parameters and their confidence intervals.

Unlike the linear case, for which a neat theory already exists (Ljung, 1999), for systems nonlinear in the parameters this is not equally well developed, though many important results are available for algebraic models (Seber and Wild, 1989). This paper considers the numerical computation of confidence regions for the parameters of nonlinear dynamical models, assuming that for their estimation a direct search method is used, based on the flexible polyhedron originally proposed by Nelder and Mead (1964).

In the sequel a modified polyhedron is considered (Marsili-Libelli and Castelli, 1987; Marsili-Libelli, 1992) where the expansion phase in the search direction is optimised.

After introducing the basic definition of the exact confidence regions, two approximation methods are considered: a local linearisation of the output function, leading to the Fisher Information Matrix (FIM; Ljung, 1999; Munak, 1999; Petersen, 2000), and a quadratic expansion of the estimation error functional involving the Hessian matrix (Press et al., 1986). Each method leads to the definition of statistically significant confidence regions. However, a further step is taken proposing a test to determine whether the estimated parameters converge to the “best” parameters which the search algorithm can find. It is shown that the confidence ellipsoids obtained with the Hessian or the Fisher method coincide only when the estimation converges to the true parameters. Otherwise, they yield clearly differing results, indicating an inaccurate estimation. The method is demonstrated with reference to a pair of widely used dynamics: the Monod kinetics, which has been widely used in ecology and biotechnology, and the Richards growth function, which is often applied to plants growth and in this paper is applied to unicellular algal growth. It is shown that in both cases this method is highly sensitive, compares favourably with the residual correlation analysis and appears to have good discriminatory power. The method, which is demonstrated for the Monod and Richards models, can easily be generalised to any structurally identifiable nonlinear model.

1.1. Problem statement

A dynamical autonomous nonlinear system can be described by the general state-output equations:

\[
\begin{align*}
\frac{dx}{dt} &= h(x, p, t) \\
y &= f(x, p)
\end{align*}
\]
where $x \in \mathbb{R}^q$ is the system state, $p \in \mathbb{R}^p$ is the output, $\mathbb{R}^n$ is the vector of its parameters. As a consequence of its nonlinearity, $h(\cdot, \cdot) : \mathbb{R}^n \to \mathbb{R}^q$ and $f(\cdot, t) : \mathbb{R}^q \to \mathbb{R}^q$, in general no analytical solution to model (1) can be found. Parameter estimation is obtained by minimising the weighted quadratic output error functional

$$E(p) = \sum_{i=1}^{N} [y_i - \hat{y}_i]^T V_i^{-1} [y_i - \hat{y}_i]$$  \hspace{1cm} (2)

where $[y_i] \in \mathbb{R}^q$ are the $N \times n_p$ experimental data points, possibly affected by measurement noise, and $[y_i] \in \mathbb{R}^q$ are the model outputs at the same sampling instants $[t_i] \in \mathbb{R}^q$. The diagonal matrices $V_i \in \mathbb{R}^{q \times q}$ are collections of weighting factors to attach more or less importance to each component. They can be specified depending on the estimation problem, or, if statistical information is available in terms of experimental variance $\Sigma_i \in \mathbb{R}^{q \times q}$, it can be assumed that $V_i = \Sigma_i$. In this way, the more a measurement is noise-corrupted the less influence it has in the error functional.

Assuming that model (1) is structurally identifiable in the sense of Pohjanpalo (1978) and Godfrey and Di Stefano (1985), consistent estimation is possible, i.e. the estimated parameters $\hat{p}$ tend to the true parameters $p$. The estimation algorithm applied to the functional Eq. (2) with model (1) as a constraint seeks a parameter vector $\hat{p}$ such that $\hat{p} = \arg \min_p E(p)$, and since no closed-form solution to model (1) is available, the optimisation problem is often solved through a numerical search based on the well-known “flexible polyhedron” or “simplex” (Nelder and Mead, 1964; Hemmelblau, 1972) combined with a numerical simulation algorithm to obtain $[y_i] = 1, \ldots, N$ for each trial of the parameter vector $p$. The algorithm used in this paper is an optimised version of the classical Nelder and Mead flexible polyhedron and is described in details in Marsili-Libelli (1992). However, the aim of this paper is not to address the parameter estimation proper, but to assess the accuracy of the estimated parameters obtained through this numerical search. In fact the drawback of this approach is that the search may terminate far from the true minimum $\hat{p}$ for several reasons, such as inadequate choice of simplex coefficients or initial parameter guess, or trapping in local minima. It is therefore useful to have a method to judge whether the search terminated successfully at the minimum or went astray over the $E(p)$ surface.

1.2. Approximations of estimated parameters confidence regions

Confidence regions are of primary importance because they provide a way to judge the accuracy of parameter estimates. Since the objective functional $E(p)$ represents the “closeness” of the experimental data to the fitted model, it is justifiable to base the confidence region of $\hat{p}$ on the contours of $E(p)$. In the parameter space, such a region has the general form:

$$\{ p : E(p) \leq cE(\hat{p}) \}$$  \hspace{1cm} (3)

for any $c > 1$. This region can be regarded as “exact”, as it is not based on any approximation, although it is difficult to select a value for $c$ with any statistical significance. However, for large $N$ the confidence region based on the $F$ statistics (Seber and Wild, 1989)

$$\{ p : E(p) \leq (1 + \frac{n_p}{N - n_p}) \frac{F_{n_p, N - n_p}}{F_{n_p, N - n_p}} \}$$  \hspace{1cm} (4)

has the required asymptotic confidence level of 100 $(1 - \alpha)\%$ where $F_{n_p, N - n_p}$ is the upper $\alpha$ critical level of the $F_{n_p, N - n_p}$ distribution. The practical difficulty in estimating the region defined by Eq. (4) has been examined by Vanrolleghem and Keesman (1996), who suggested the use of extensive Monte Carlo simulations. More recently, Dochain and Vanrolleghem (2001) proposed a successive contraction method to find the value of $E(p)$ corresponding to the prescribed value of $\alpha$.

1.3. Covariance matrix estimation through a linear approximation

A definition of the confidence region as a function of the parameter covariance matrix $C$ is now introduced. In general such a region can be expressed as (Seber and Wild, 1989; Ljung, 1999)

$$\{ p : (p - \hat{p})^T C^{-1} (p - \hat{p}) \leq \sigma^2 \}$$  \hspace{1cm} (5)

It is well known that for a linear model $y = px + \nu$, with residual noise $\nu \sim N(0, \sigma^2 I)$, $C$ can be obtained in closed form as $C = \sigma^2 (X^T X)^{-1}$ where $X = [x_1, x_2, \ldots, x_q]^T$. 
For nonlinear models there is no exact way to obtain $C$ and the linear approximation may yield a poor estimate of the real confidence region (Donaldson and Schnabel, 1987; Rooney and Biegler, 1999). The previous covariance matrix $C$ obtained for the linear case can be extended to yield the approximate covariance matrix as (Goodwin and Payne, 1977; Munak, 1999; Petersen, 2000; Dochain and Vanrolleghem, 2001)

$$C(J(\hat{p})) = \frac{E(\hat{p})}{N - np} - \sum_{i=1}^{np} (\frac{\partial E}{\partial p_i} |_{\hat{p}}) (\frac{\partial^2 E}{\partial p_i^2} |_{\hat{p}})^{-1}$$

where the term $E(\hat{p})/N - np$ is an unbiased approximation of the residual variance $\sigma^2$ and the $J$ is the Jacobian matrix of model Eq. (1), which can be written column-wise as

$$J = [J_1 J_2 \cdots J_1 J_s]$$

with

$$J_i(\hat{p}) = \frac{\partial y}{\partial p_i} |_{\hat{p}} = \frac{\partial f}{\partial y} |_{\hat{p}} + \frac{\partial f}{\partial x} |_{\hat{p}} + \frac{\partial f}{\partial y_1} |_{\hat{p}}$$

The columns of the Jacobian matrix $J$ are the output sensitivities $S_{pi}$, $S_{pi} = (\partial y/\partial p_i)$ ($s = 1, \ldots, n_p$) with respect to the parameters. They can be generated by model Eq. (1) through the associated state trajectory sensitivity $S_{pi}$, $S_{pi} = (\partial x_i/\partial p_j)$ ($s = 1, \ldots, n_p$) by the linear dynamical model (Perkins, 1972)

$$\frac{dS_{pi}}{dt} = \frac{\partial h}{\partial x_s} |_{\hat{p}} S_{pi} + \frac{\partial h}{\partial y_1} |_{\hat{p}} S_{pi} + \frac{\partial h}{\partial y_s} |_{\hat{p}} S_{pi}(0)$$

Assuming that the measurement noise $w \sim N(0, \sigma^2 I_n)$ is uncorrelated, $C$ given by Eq. (6) is also the inverse of the FIM, defined as FIM $= \sum_{i=1}^{np} (\partial y/\partial p_i)^T (\partial y/\partial p_i)$, in terms of output sensitivities (8), i.e. $C(J(\hat{p})) = C(J(\hat{p}))^{-1}$ (Munak, 1999; Petersen, 2000; Dochain and Vanrolleghem, 2001). In this case it represents the error covariance matrix of the minimum variance unbiased estimator according to the Cramér-Rao theorem (Ljung, 1999).

Substituting $C(J)$ from Eq. (6) into Eq. (5), yields the approximate confidence ellipsoids

$$\{p : (p - \hat{p})^T C(J(\hat{p}))^{-1} (p - \hat{p}) \leq \frac{1}{2} \}$$

1.4. Covariance matrix estimation through the Hessian matrix

For nonlinear models the objective function $E(p)$ given by Eq. (2) is not an exact quadratic form, but for sufficiently small deviations $p - \hat{p}$ it can be approximated by a second-order expansion around the estimated parameters $\hat{p}$. Since in the neighbourhood of the minimum the first-order term is missing $\approx E(p) = (p - \hat{p})^T C(\hat{p})^{-1} (p - \hat{p}) / \sqrt{\text{degrees of freedom}}$

Substituting Eq. (10) into Eq. (4), a result formally similar to the linear case is obtained

$$C_H(p) = \frac{2}{N - np} E(\hat{p}) H(\hat{p})^{-1}$$

with confidence ellipsoids given by

$$\{p : (p - \hat{p})^T C_H^{-1}(p - \hat{p}) \leq \frac{1}{2} \leq n_p F_{\alpha/2, N-n_p}^{-1}$$

In any case, the individual parameter confidence interval $p_i$ can be obtained as

$$\hat{p}_i = \hat{p}_i \pm \sqrt{\text{Student's } t \text{ distribution for the given confidence level } \alpha \text{ and } N - n_p \text{ degrees of freedom.}}$$

2. A parameter estimation validity test

There is an important conceptual difference in the two approximations $C_J$ and $C_H$ and the comparison between the related confidence regions can be used to detect inaccurate estimation results. The FIM approach is based on the sensitivity trajectories, whereas the Hessian expansion depends on the shape of the error functional. The following component-wise
relationship between them can be established as follows:

\[
\begin{align*}
H_{ij} = & \frac{\partial^2 E(\hat{p})}{\partial p_i \partial p_j} = 2 \sum_{i=1}^{N} \left( \left( \frac{d^2 y(x_i, \hat{p})}{dp_i dp_j} \right)^T \frac{d y(x_i, \hat{p})}{dp_i} \right), \\
& - \left( y_i' - y(x_i, \hat{p}) \right)^T \sum_{i=1}^{N} \left( \frac{d^2 y(x_i, \hat{p})}{dp_i dp_j} \right) \frac{d y(x_i, \hat{p})}{dp_i} \\
& = 2 \sum_{i=1}^{N} \left( \frac{d y(x_i, \hat{p})}{dp_i} \right)^T \sum_{i=1}^{N} \left( \frac{d^2 y(x_i, \hat{p})}{dp_i dp_j} \right) \frac{d y(x_i, \hat{p})}{dp_i} \\
& \approx 2 \sum_{i=1}^{N} \sum_{k=1}^{N} s_{ik}(i)^T V^{-1} s_{ik}(i) = 2 \frac{N - n_p}{E(\hat{p})} \text{FIM}_{kx}.
\end{align*}
\]

(14)

Thus, if \((x', \hat{N}(\sigma^2; \hat{V}, \hat{p}))\), \text{FIM} is the expected value of \((x'^2/2)H\) with \(x'^2 = E(\hat{p})/(N - n_p)\) (Seber and Wild, 1989). However, Eq. (14) shows that in general \(H\) and \text{FIM} differ by the term

\[
\sum_{i=1}^{N} \sum_{k=1}^{N} s_{ik}(i)^T V^{-1} s_{ik}(i)
\]

(15)

involving the curvature (see Appendix B) \((d^2 y(x_i, \hat{p})/dp_i dp_j)\) of the surface \(\Omega : (y(t) = \phi(x, \hat{p}) \in \mathcal{P} \subseteq \mathbb{R}^{n_p})\) which acts as a weighting factor for the estimation error \((y_i' - y(x_i, \hat{p}))\). As a consequence, the two approximations in Eq. (14) coincide only when either the curvature or the estimation error is negligible, i.e., when \(\hat{p} \approx \hat{p}\). If for some numerical reason the search terminates far from the minimum, then this term will cause \text{FIM} and \(H\) to differ, revealing an inaccurate estimation result.

\(\Omega \subseteq y \in \mathbb{R}^n\) is a \(n_p\)-dimensional surface in \(\mathbb{R}^{n_p}\). Since \(\Omega\) contains all possible values of \(E[y]\), it is called “expectation surface” (Beale, 1960; Goldberg et al., 1983). \(\Omega\) is also called “solution locus” and can be described in terms of a \(n_p\)-dimensional parameter set \(p \in \mathbb{R}^{n_p}\) and a model function \(\mu = f(x, \hat{p})\) in such a way that \(\Omega = \{p : \mu = f(x, \hat{p})\ i = 1, \ldots, N; \forall p \in \mathcal{P} \subseteq \mathbb{R}^{n_p}\}.

For a nonlinear system the shape of the surface \(\Omega\) may be heavily influenced by the curvature (Donaldson and Schnabel, 1987; Seber and Wild, 1989) so the effect of neglecting this term may be significant. Thus, if the two representations do not agree, this means that the curvature term may significantly amplify the estimation error.

In the neighbourhood of \(\hat{p}\) curvature is composed of a tangent and a normal component with respect to the surface \(\Omega\) (see Appendix B). The former, termed parameter curvature, represents the degree of curvature induced by the choice of parameters, whereas the latter, called intrinsic curvature, measures the degree of distortion of the surface due to the nature of the output function. Around the minimum the intrinsic curvature should therefore become negligible regardless of the value of \(E(\hat{p})\). To assess the practical identifiability of a fixed-structure model parameters, parametric curvature is of secondary importance, because it can be eliminated by a proper re-parametrisation. On the other hand, intrinsic curvature should be accounted for, possibly by introducing corrective factors in evaluating the confidence ellipsoids. Confidence region involving \text{FIM} of prescribed statistical significance can be obtained from Eq. (9) as:

\[
\{p : (p - \hat{p})^T \text{FIM}(p - \hat{p}) \leq r^2_N\}
\]

(16)

where \(r^2_N = n_p F_{n_p, \lambda}^{-1}(1 - \lambda_{\max})\) (Seber and Wild, 1989) and \(\lambda_{\max}\) is the maximum eigenvalue of the \(\hat{B}\) matrix, related to the Hessian matrix by the relationship

\[
I_{s_p} - \hat{B} = \frac{1}{\lambda} \hat{K}^T \hat{H} \hat{K}
\]

(17)

where \(\hat{K}^T \hat{K}\) is the Cholesky decomposition of \(\text{FIM}(\hat{p})\). Through intrinsic curvature the extent of model nonlinearity can be assessed by comparing the maximum and minimum curvature radii \(r_{\max} = (1 - \lambda_{\max})^{-1/2}\) and \(r_{\min} = (1 - \lambda_{\max})^{-1/2}\) with unity. The closer they are to 1 the smaller is the curvature. In this way the relative importance of the curvature term can be appreciated (Bates and Watts, 1980). An efficient numerical procedure for the construction of the covariance matrices \(C_J\) and \(C_H\) is required to
carry out the above analysis. This is summarised in Appendix A.

2.1. Checking parameter validity through correlation analysis

As a comparison to this discriminatory tool the correlation analysis will be considered. In fact, a yardstick for the validity of the estimated parameters is to check whether the estimation residuals are uncorrelated with zero mean. This implies that the correlation terms for any lag \( k \) should fall in the confidence limits of a given statistical significance \( \alpha \), computed as \( \pm \frac{t_{\text{N}-1,\alpha/2}}{\sqrt{N-1}} \) where \( k \) is the lag and \( t_{\text{N}-1,\alpha/2} \) is the two-tails Student’s \( t \)-test for the given confidence level \( \alpha \) and \( N-n_p \) degrees of freedom (Press et al., 1986; Seber and Wild, 1989; Ljung, 1999).

3. Test models

The previous confidence region computation and discrimination method is now demonstrated with the aid of two simple models, which are widely used in ecology: the Monod kinetics, which is at the basis of many ecologically significant processes, and the Richards growth function, which is used to model vegetative growth at any level, from single leaf to whole plant.

3.1. Monod kinetics

This kinetics was selected for its fundamental role in modelling a wide range of microbial processes (Spriet, 1982). In spite of its structural simplicity the Monod kinetics has represented a challenging parameter estimation problem for many years, first in the simple case when both substrate and biomass were observed (Heineken et al., 1967; Swartz and Bremermann, 1975; Nihtilä and Virkkunen, 1977; Holmberg, 1982) and then as a joint state/parameter estimation problem (Holmberg and Ranta, 1982). More complex situations were also considered, when only indirect measurement of dissolved oxygen were used (Marsili-Libelli, 1989) and finally in connection with respirometric studies (Vannolleghem et al., 1995; Petersen, 2000; Dochain and Vannolleghem, 2001).

The Monod kinetics is used here in the following form:

\[
\begin{align*}
\frac{dS}{dt} &= \frac{1}{Y} \frac{\mu_{\text{max}} S X}{K_s + X} \\
\frac{dX}{dt} &= \mu_{\text{max}} S X - K_s X \\
x &= [X]_t^T, \quad y = x
\end{align*}
\]

where \( S \) and \( X \) represent the substrate and the biomass concentration respectively, both expressed as \( \text{mg COD l}^{-1} \). It is assumed that both \( S \) and \( X \) are measured. The performance of the parameter calibration procedure is assessed using a set of \( N \) synthetic data values \( y_i \) \( i = 1, \ldots, N \) generated by the model (18) with nominal parameters \( \tilde{\theta} \) and corrupted with white noise proportional to the output value with characteristics \( \tilde{\nu}_N(0, \sigma^2 I_N) \). For autonomous systems such as (18), no “persistent excitation” (Ljung, 1999) condition on the input can be imposed. On the contrary, identifiability must be related to model structure and initial conditions and a “structural identifiability” test proposed by Pohjanpalo (1978) should be used. In fact, Holmberg (1982) and Holmberg and Ranta (1982) referred to this test for their Monod kinetics identification studies.

For data generation, the following parameter values have been used in the simulations:

\[
\begin{align*}
\tilde{\theta} : & \quad \mu_{\text{max}} = 0.5 \text{ (h}^{-1}) ; \quad K_s = 20 \text{ (mg COD} l^{-1}) ; \quad Y = 0.5 ; \quad K_d = 0.03 \text{ (h}^{-1}) \\
\end{align*}
\]

The initial conditions were set at \( S(0) = 10 \text{ mg COD} l^{-1} \) and \( X(0) = 0.1 \text{ mg COD} l^{-1} \). Holmberg (1982) refers to difficult identifiability when \( S < K_s \) because the data do not cover the whole \( S \) range. In fact, the case considered here (\( K_s/S_o = 2 \)) corresponds to a particularly difficult identifiability condition. Further, it should be considered that when \( K_s/S_o \) decreases, the achievable accuracy for \( \mu_{\text{max}} \) increases but that of \( K_s \) decreases. The sampling interval too has a strong influence on the accuracy. A simulation time of 150 h was assumed with a 3 h sampling time, providing a total of 51 data points which were then corrupted with proportional noise of known characteristics. This kind of noise is often preferred to additive noise with constant variance and has been often used for testing biotechnological models. Holmberg (1982)
has assessed the effect on the identification of different kinds of noise and concluded that experiments with Gaussian additive noise of constant variance do not yield better estimates than those with proportional noise.

A sample set of noisy data used for algorithm testing is shown in Fig. 1. The Monod kinetics (18) was implemented in Matlab 5.3 using a Simulink model, with a Rosenbrock stiff integration method, a maximum step of 0.1 h and an absolute tolerance of $10^{-5}$.

Fig. 1. A sample of synthetic noisy data generated from the Monod kinetics.

Fig. 2. Relationship between optical density and cell count.
3.2. Richards growth kinetics applied to algal growth assays

Another calibration problem is now considered, using laboratory data. Assays of unicellular algae *Selenastrum capricornutum* were grown in a controlled medium with varying nutrient concentrations according to a widely used protocol (Miller et al., 1978). Microalgal growth has been modelled by several authors (Nyholm, 1978; Nyholm and Lingby, 1988; Sterner and Grover, 1998) with Monod kinetics. However, previous studies (Marsili-Libelli et al., 1997, 1998) indicate that the dynamics of algae can be satisfactorily described by the Richards function, originally conceived to model vegetative growth (Causton and Venus, 1981). The dynamics of algal growth in the batch was thus modelled as

\[
\frac{dG}{dt} = Gr \left(1 - \frac{G}{K}\right)
\]

where \(G\) is the algal density (cell cm\(^{-3}\)), \(r\) is the specific growth rate (h\(^{-1}\)), \(n\) is the Richards growth exponent and \(K\) is the carrying capacity (cell cm\(^{-3}\)). The algae were starved before starting the growth batch, in order to deplete the internal nutrient supply, and the initial algal density was measured at the beginning of the batch. In this simple application the pure growth dynamics is considered, without reference to the internal nutrient storage mechanism, termed “cell quota” as introduced by Droop (1983).

During the batch light and CO\(_2\) were always kept well above their limiting values. Growth was indirectly measured using optical density at 670 nm with a UV spectrometer (Perkin-Elmer, Lambda 2).

4. Application of the curvature method to the test cases

The previous theory is now applied to the parameter estimation of the Monod and Richards models (18) and (20) using a simplex-based algorithm (Marsili-Libelli, 1992). In this case the search may stop away from the minimum as a consequence of poor initialisation or improper simplex characteristics.

4.1. Monod kinetics

Two parameter estimation runs were set up using noisy data generated with the Monod model (18) with \(\sigma = 0.05\). Two cases were considered, depending on simplex initialisation: in Case 1 the search was started at a feasible point and terminated correctly at the minimum, whereas in Case 2 an intentionally unsuitable starting point was chosen, in order to obtain a search termination far from the minimum though the terminating condition had been met. The results of Table 1 were obtained. By comparison with the nominal parameters \(\hat{p}\) it can be seen that only in Case 1 the estimation results are sufficiently accurate, whereas in Case 2 completely wrong estimates were obtained.
| True values | 20.000 | 0.500 | 0.500 | 0.0030 |
| Case 1     | 19.0997 | 0.483 | 0.525 | 0.0303 |
| Case 2     | 31.605  | 0.711 | 0.485 | 0.0289 |

When the search was successful (Case 1) the confidence interval of the estimated parameters was computed according to Eq. (13), as shown in Table 2.

It can be seen that $K_s$, with a 95% confidence interval of over 18% of the estimated value, has the largest uncertainty, though much less than that found by Holmberg (1982). On the other hand, $\mu_{\text{max}}$ has a 13% uncertainty and $K_d$ only 3.8%.

The contour plots of Fig. 4 in the $K_s - \mu_{\text{max}}$ space confirm that the estimated values of Table 1 are close to the true parameters only in Case 1, whereas in Case 2 the search stopped far from the minimum, though remaining in the narrow valley.

Comparing the ellipsoids obtained with the two methods, Cases 1 and 2 can be easily discriminated. Fig. 5 shows that in Case 1, where the objective function minimum is reached with sufficient accuracy, the ellipsoids constructed with either the Hessian or the FIM approximation coincide. On the other hand, Fig. 6 shows that in Case 2 the ellipsoids obtained with the Hessian matrix are very different from those computed through the FIM approximation, due to the nonzero curvature terms which tend to amplify the estimation errors.

The curvature radii for the two cases (see Table 3) confirm this analysis: in Case 1 both $r_{\text{min}}$ and $r_{\text{max}}$ are similar, whereas in Case 2 they are significantly different.

### Table 1

<table>
<thead>
<tr>
<th>$K_i$ (mg COD l$^{-1}$)</th>
<th>$\mu_{\text{max}}$ (h$^{-1}$)</th>
<th>F</th>
<th>$K_d$ (h$^{-1}$)</th>
<th>Starting point $E(\hat{p})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True values</td>
<td>20.000</td>
<td>0.500</td>
<td>0.500</td>
<td>0.0030</td>
</tr>
<tr>
<td>Case 1</td>
<td>19.0997</td>
<td>0.483</td>
<td>0.525</td>
<td>0.0303</td>
</tr>
<tr>
<td>Case 2</td>
<td>31.605</td>
<td>0.711</td>
<td>0.485</td>
<td>0.0289</td>
</tr>
</tbody>
</table>

### Table 2

Confidence interval of the Monod growth function estimates

<table>
<thead>
<tr>
<th>$K_i$ (mg COD l$^{-1}$)</th>
<th>$\mu_{\text{max}}$ (h$^{-1}$)</th>
<th>F</th>
<th>$K_d$ (h$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>19.09968 ± 1.72181</td>
<td>0.48298 ± 0.03212</td>
<td>0.5024 ± 0.01139</td>
</tr>
<tr>
<td>Case 2</td>
<td>31.605</td>
<td>0.711</td>
<td>0.485</td>
</tr>
</tbody>
</table>

Fig. 4. Contour plot in the two cases of Table 1. The plus (+) indicates the minimum reached by the optimisation algorithm. Only in Case 1 the true minimum is reached.
Fig. 5. Case 1: Ellipsoids calculated with Hessian (dashed line) and FIM matrix (solid line) coincide indicating convergence to the “true” parameters (+).

Fig. 6. Case 2: Ellipsoids evaluated with Hessian matrix (dashed line) and with FIM matrix (solid line) differ, indicating inaccurate estimated parameters (+).
are very close to 1, implying that the curvature is negligible, whereas in Case 2 the curvature is much more pronounced, indicating that the terminal value of \( \hat{p} \) is far from the minimum \( \hat{p} \).

### 4.2. Sample autocorrelation test

The sampled autocorrelation function (Shumway and Stoffer, 2000) for the two cases of Table 1 is now considered for comparison and shown in Fig. 7. It can be seen that Case 2 has several samples outside the zero autocorrelation. Therefore, in Case 2, the estimated parameters fail to "whiten" the residuals, thus indicating an inaccurate result. On the other hand, even the results for Case 1 are rather inconclusive since there are a couple of borderline cases, e.g. at lag \( k = 3 \) for the substrate and at lag \( k = 1 \) for the biomass. The results of the correlation analysis are thus much less discriminatory and no definite line can be drawn between Cases 1 and 2, since even when convergence to true values is achieved, the residual correlograms do not show a definitely "white" behaviour. Lack of conclusive "whitening" should in this case be traced back to the noise characteristics used to corrupt the model output in Eq. (2) or, more generally, the very structure of the model should be questioned. Also, the limited number of calibration data, as often happens in ecology, make the correlation analysis rather uncertain if not outright unsuitable.

### 4.3. Richards growth function

As with the Monod model, the search algorithm was initialised with three different parameter vectors to produce a possible estimation failure. The three cases are listed in Table 4, where the starting point is indicated, together with the final parameter values and the magnitude of the error functional. Case 1 corresponds to a correct estimation, whereas Cases 2 and 3 are considered inaccurate estimates. Unlike the Monod case, here there is no direct way of checking the estimation accuracy because the "true" parameters...
Fig. 7. Residuals autocorrelation functions for Cases 1 and 2 with 95% zero-limit bounds (dotted lines). Even in Case 1 partially correlated residuals were obtained, with some borderline cases. In Case 2 several correlation samples fall definitely outside the zero-confidence limit.
are not known, and the ellipsoids method is the only way of assessing the reliability of the estimates. Fig. 8 shows that in Case 1 exact coincidence of the Hessian and FIM ellipsoids is obtained, indicating a correct estimation. In fact, in this case the correction factor due to intrinsic curvature does not modify the FIM ellipsoids.

Only when the search was successful (Case 1), the 95% confidence interval of the estimated parameters was computed according to Eq. (13) and shown in Table 5. It can be seen that the exponent $n$ is the most difficult parameter, given its very small magnitude and the possibility to obtain negative values with no ecological significance (Causton and Venus, 1971).
This parameter has a very large 95% confidence interval: over 700% of the estimated value, compared with less than 35% for \( r \) and only 3.5% for \( K \).

Cases 2 and 3 both represent inaccurate estimation results, due to differing causes. In Case 2, shown in Fig. 9, the Hessian ellipsoids are larger than the FIMs. The opposite is true in Case 3, shown in Fig. 10, and therefore the maximum curvature radii are less than one. The curvature radii are grouped in Table 6. It can be seen from the contour plots of Fig. 11 that Case 2 corresponds to a search termination at a point where the gradient is still considerable, far from the minimum. On the other hand, Case 3 represent the opposite situation: the search stopped in a flat region near but not exactly at the minimum. At this point one might wonder whether the correctness of the estimation could be assessed just by inspection of the contour plots. The answer is of course negative, since this visual check would lack the accuracy of the ellipsoids method, which is very sensitive to the contribution of the curvature term.

Given the small number of experimental data, no correlation analysis can be performed. In this case, the proposed method is therefore the only way to check the estimation accuracy.

Table 6  
Curvature radii for the three cases of the Richards growth function

<table>
<thead>
<tr>
<th></th>
<th>( r )</th>
<th></th>
<th>( K )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r_{\text{min}} - r_{\text{max}} )</td>
<td></td>
<td>( r_{\text{max}} - r_{\text{min}} )</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{min}} - K_{\text{max}} )</td>
<td></td>
<td>( K_{\text{max}} - K_{\text{min}} )</td>
</tr>
<tr>
<td>CASE 1</td>
<td>0.9998–1.0003</td>
<td>CASE 1</td>
<td>0.9941–1.0008</td>
</tr>
<tr>
<td>CASE 2</td>
<td>0.9799–1.0004</td>
<td>CASE 2</td>
<td>0.9845–1.1003</td>
</tr>
<tr>
<td>CASE 3</td>
<td>0.9832–1.0004</td>
<td>CASE 3</td>
<td>0.7866–0.9970</td>
</tr>
</tbody>
</table>
Fig. 11. Contour representation of the error functional and the parameters found in the three cases. It can be seen that Case 3 is closer to the minimum (Case 1) and therefore is less influenced by intrinsic curvature. The opposite is true for Case 2.
5. Conclusion

This paper has addressed the problem of estimating the confidence regions of the parameters of nonlinear dynamical model when they are determined through a numerical search algorithm, such as the flexible polyhedron. Two methods to estimate such regions have been considered, based on either the Hessian matrix or the FIM. In the first part of the paper, these quantities are compared and a relationship between them is derived. In the second part of the paper a method is proposed to assess the consistency of the estimation. This test is based on the comparison of the parameter confidence regions obtained with either the Hessian or the FIM approximation. There is a conceptual difference in the two approximations: the FIM approach is based on the sensitivity trajectories, whereas the Hessian expansion depends on the shape of the error functional. This difference is used to assess the success of the estimation, because the two methods yield the same ellipsoids only if the search terminates at the minimum, otherwise they may significantly differ. As a comparison, correlation analysis was considered and it was shown that this may yield inconclusive results, in terms of ‘borderline’ correlation values, even in the case of a valid parameter estimation, whereas inconsistent estimates may not result in substantial residual correlation. By contrast, confidence ellipsoids coincidence or discrepancy is much more pronounced in the case of successful or faulty estimation. In this sense, this method compares favourably with the residual correlation analysis, and it provides superior discriminatory power.

Appendix A. A numerical algorithm for Hessian matrix computation

Computing the covariance matrix through the Hessian matrix involves numerical differentiation of \( E(p) \).

The following algorithm determines the optimal increment for the finite differences approximating the second derivative \( H(p) = (\partial^2 E(p)/\partial p_i \partial p_j)) \) in a robust way, in spite of functional irregularities and data noise. Consider an even collection of central differences with respect to the generic parameter \( p_i \), which in Matlab 5.3 is about \( 2.204 \times 10^{-16} \), and the numerical integration error. Considering three successive central differences with respect to the generic parameter \( p_i \) yields

\[
\begin{align*}
C_1 &= E(\hat{p}_i + h) - E(\hat{p}_i - h) \quad 2h \\
C_2 &= E(\hat{p}_i + (h/2)) - E(\hat{p}_i - (h/2)) \quad h \\
C_3 &= E(\hat{p}_i + (h/4)) - E(\hat{p}_i - (h/4)) \quad h/2 \\
E_{ii} &= E(\hat{p}_i + h) - E(\hat{p}_i) \\
E_{ii} &= E(\hat{p}_i + h/2) - E(\hat{p}_i + h/4) \\
E_{ii} &= E(\hat{p}_i + h/4) - E(\hat{p}_i + h/8) \\
E_{ii} &= E(\hat{p}_i + h/8) - E(\hat{p}_i + h/16) \\
E_{ii} &= E(\hat{p}_i + h/16) - E(\hat{p}_i + h/32) \\
\end{align*}
\]

where \( E(\hat{p}_i) \) and \( E(\hat{p}_i) \) are the third and fifth derivatives of \( E(p) \). The three terms \( C_1, C_2 \) and \( C_3 \) can be combined so that the first derivative of \( E(p) \), indicated as \( E(\hat{p}_i) \) can be computed with an approximation of \( O(h^2) \). Defining the previous quantities with the new symbols

\[
\begin{align*}
E_{ii} &= E(\hat{p}_i + h) - E(\hat{p}_i) \\
E_{ii} &= E(\hat{p}_i + h/2) - E(\hat{p}_i + h/4) \\
E_{ii} &= E(\hat{p}_i + h/4) - E(\hat{p}_i + h/8) \\
E_{ii} &= E(\hat{p}_i + h/8) - E(\hat{p}_i + h/16) \\
E_{ii} &= E(\hat{p}_i + h/16) - E(\hat{p}_i + h/32) \\
\end{align*}
\]

yields an approximation of \( E(\hat{p}_i) \)

\[
\begin{align*}
E(\hat{p}_i) &= \frac{64c_1 - 12c_2 + c_3}{2} + O(h^2) \\
E(\hat{p}_i) &= \frac{128 E_{ii} - 4 E_{ii}}{9} + O(h^2) \\
&\quad + \sum_{i,j=1}^{n} \left( \frac{E_{ij} - E_{ij}}{90} \right) + O(h^3)
\end{align*}
\]

Deriving \( E(\hat{p}_i) \) with respect to the \( p_i \) parameter yields the generic \( i,j \)th element of the Hessian matrix with an approximation of \( O(h^3) \).
For the diagonal elements simpler expressions can be obtained introducing the quantities

\[ E_i, E_j, E_{ij}, E_{ji}, E_{ii}, E_{jj}, E_{ij}^{(2)}, E_{ji}^{(2)} \]

\[ w_i, w_j, w_{ij}, w_{ji}, w_{ii}, w_{jj}, w_{ij}^{(2)}, w_{ji}^{(2)} \]

(\[ A.5, A.6 \])

For the diagonal elements simpler expressions can be obtained introducing the quantities

\[ w_1 = E(\hat{p}_i + \delta h) - 2E(\hat{p}_i) + E(\hat{p}_i - \delta h) \]

\[ w_2 = E(\hat{p}_i + \delta h / 2) - 2E(\hat{p}_i) + E(\hat{p}_i - \delta h / 2) \]

(\[ A.7 \])
to yield
\[ H_{ij} = \frac{4w_2 - w_1}{3} + O(\delta_i^3) \] (A.8)

There are two error sources affecting the numerical computation of \( H \): round-off errors and data noise. Since the Hessian difference is a convex function, a search algorithm can be designed to determine the best choice of \( \delta_k, k = 1, \ldots, n_p \). This search is based on the sequence of matrices \( \hat{H}_i = \hat{H}_{i+1} - \hat{H}_i \) where \( i \) is an iteration counter. At each iteration the following quantities are computed:

\[
\begin{aligned}
& s_k = g_{i-1} + \varepsilon \\
& \delta_{i,k} = \frac{\hat{p}_k}{\varepsilon}, \quad k = 1, \ldots, n_p \\
& \delta H_i = \hat{H}_{i+1} - \hat{H}_i \\
& \eta_i = \|\delta H_i\|_\infty
\end{aligned}
\] (A.9)

The search is terminated if
\[ \eta_i < \tau \quad \text{or} \quad \eta_i > \eta_{i-1} \] (A.10)

where \( \tau \) is the prescribed tolerance. The choice of the optimal increment in computing \( H \) is now illustrated using data obtained from the Monod model (18) with noisy data \( (\sigma = 0.01) \). A plot of the matrix difference \( \eta_i \) as a function of the increment \( g_i \), is depicted in Fig. 12, where the termination criteria Eq. (A.10) has been removed to show the whole co-domain of \( \eta_i \), influenced either by the round-off or the noise effects. In fact, Fig. 12a reveals the occurrences of round-off errors along the downward path whereas Fig. 12b shows the fluctuating behaviour of the approximation when the noise effect prevails and \( \eta \) tends to rise again. If the termination criteria Eq. (A.10) had been in effect, the search would terminate at the first spike in Fig. 12a, as indicated by the arrow. Instead, the search is extended into the round-off region (Fig. 12b) where the fluctuations of the \( \eta \) sequence increases with \( g_i \).

Appendix B. Definition of curvature

The term *curvature* (Beale, 1960; Goldberg et al., 1983; Seber and Wild, 1989) corresponds to the second derivative of the model output functions with respect to parameters

\[
\hat{K} = \left[ \left( \frac{\partial^2 f(x_1, i)}{\partial y_i \partial y_j} \right) \right]_{p=\hat{p}}
\] (B.1)

The advantage of the second-order approximation to the expectation surface \( \Omega \) at the point \( \hat{p} \) depends on the relative magnitude of the quadratic term \( (p - \hat{p})^T \hat{K} (p - \hat{p}) \) compared to the linear term \( \hat{J} (p - \hat{p}) \).

Further, \( \hat{K} \) can be decomposed into two orthogonal components, one being the projection onto the tangent plane \( \hat{K}_T \) and the other \( \hat{K}_N \) being normal to it, such that \( \hat{K} = \hat{K}_T + \hat{K}_N \). This decomposition can be obtained through the projection matrix

\[
\hat{p} = \hat{J} (\hat{J}^T \hat{J})^{-1} \hat{J}^T
\] (B.2)

to obtain

\[
\hat{K}_T = \hat{p} \hat{K} \quad \text{and} \quad \hat{K}_N = (I - \hat{p}) \hat{K}
\] (B.3)
Bates and Watts (1980) also define two measures for comparing each quadratic component with the linear term

\[ W_I = \frac{|(p - \hat{p})^T R_I(p - \hat{p})|}{||(p - \hat{p})||^2} \]

and

\[ W_N = \frac{|(p - \hat{p})^T R_N(p - \hat{p})|}{||(p - \hat{p})||^2} \]

These two terms are called the parameter curvature (tangent) and the intrinsic (normal) curvature, respectively, along the direction of \((p - \hat{p})\).

References


