Practical identifiability analysis of large environmental simulation models

Roland Brun[1] and Peter Reichert

Swiss Federal Institute for Environmental Science and Technology, Dübendorf, Switzerland

Hans R. Künsch

Seminar für Statistik, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

Abstract. Large environmental simulation models are usually overparameterized with respect to given sets of observations. This results in poorly identifiable or nonidentifiable model parameters. For small models, plots of sensitivity functions have proven to be useful for the analysis of parameter identifiability. For models with many parameters, however, near-linear dependence of sensitivity functions can no longer be assessed graphically. In this paper a systematic approach for tackling the parameter identifiability problem of large models based on local sensitivity analysis is presented. The calculation of two identifiability measures that are easy to handle and interpret is suggested. The first accounts for the sensitivity of model results to single parameters, and the second accounts for the degree of near-linear dependence of sensitivity functions of parameter subsets. It is shown how these measures provide identifiability diagnosis for parameter subsets, how they are able to guide the selection of identifiable parameter subsets for parameter estimation, and how they facilitate the interpretation of the correlation matrix of the parameter estimate with respect to parameter identifiability. In addition, we show how potential bias of the parameter estimates, due to a priori fixing of some of the parameters, can be analyzed. Finally, two case studies are presented in order to illustrate the suggested approach.

1. Introduction

The use of large and complex mathematical models is common practice in the environmental sciences. This is mainly due to the increasing knowledge about causal mechanisms within environmental systems and the increasing computer power available. As scientific progress continues and computational costs still decrease, further growth of the models seems to be inevitable [*Beck*, 1999]. A second driving force in this development is the desire to extrapolate from a system under study to similar systems under different driving conditions, for instance, in order to predict the impact of different measures on a similar environmental system. Such an extrapolation is impossible without an attempt to model the causal mechanisms of the system under study at an adequate level.

In his encyclopedic review article, *Beck* [1987] gives an overview of different modeling strategies and model classes applied in the environmental sciences. Basically, he distinguishes between models which try to give an efficient description of the input-output behavior of a system without relying on hypotheses about how the system works internally and models which try to give an internal description of the system. Models of the first class are called black box models. They are usually relatively simple, parsimonious, and identifiable from the observations available. A modeling approach using this type of model

is generally preferred by statisticians and control engineers. The approach is characterized by focusing on what can be learned from the data and by making little use of prior knowledge of the system. A main drawback of this approach is the lack of (mechanistic) interpretability of the model. For this reason, extrapolations to similar systems and different driving conditions are not meaningful. Models of the second class are called mechanistic. They are usually more complex and poorly identifiable. A main reason for the poor identifiability is that "what we would like to know about the internal description of the system ··· is of a substantially higher order than what can be observed about the external description of the system" [*Beck*, 1987, p. 1416]. A modeling approach relying on this model class is usually preferred by natural scientists. It is characterized by an extensive use of causal hypotheses based on current understanding of how processes work, and it is closely linked to a reductionistic world view. From this point of view the description of natural systems requires complex and hence large models. Although the application of mechanistic models for extrapolative tasks is known to be a pretty delicate issue and critical voices even argue that it sometimes is more an exercise in prophecy than in prediction [*Beven*, 1993, 1987], mechanistic models nevertheless provide at least a rational basis for the daily need of extrapolation in the environmental sciences.

Identifiability problems relating to the use of mechanistic models are widely reported and treated in the literature [see, e.g., *Beck*, 1987; *Kleissen et al.*, 1990; *Gupta and Sorooshian*, 1983; *Sorooshian and Gupta*, 1983; *Restrepo and Bras*, 1985]. We briefly summarize different strategies that have been suggested in order to overcome these problems.

---

[1]Also at Seminar für Statistik, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland.

An obvious remedy is model reduction in the sense of restricting the model description to what is observed by the data [see, e.g., *Jakeman and Hornberger*, 1993]. However, this strategy is of limited use because either we stick to a mechanistic model, which in most cases results in possible reductions not being sufficient to achieve identifiability, or we end up with a pure input-output model.

A promising approach to the identification problem is the use of Bayesian methods that incorporate a priori information on the parameters in the form of a prior distribution and do not require the parameters to be identifiable [see, e.g., *Reichert and Omlin*, 1997]. Within the field of water modeling, the Hornberger-Spear algorithm (also called regional sensitivity analysis (RSA) and Monte Carlo filtering method) [*Beck*, 1987; *Hornberger and Spear*, 1981, 1983] and generalized likelihood uncertainty estimation (GLUE) [*Beven*, 1992; *Freer et al.*, 1996] belong to this approach. Current applications of the Bayesian methodology, however, are still suffering from very time-consuming calculations. They are restricted to models with short simulation times and a moderate number of parameters in order to keep computational costs reasonably low. In addition, the problem of choosing appropriate prior distributions for the parameters is often hard to solve.

Another strategy which is widely used is based on the partitioning of the model parameters into a subset of parameters that is to be estimated from given observations and a subset of parameters that is fixed at a priori values. This strategy ensures the remaining model parameters to be identifiable. However, the partitioning of the model parameters is very much subject to the personal judgment of the modeler, a fact which is rarely made fully transparent. Further known drawbacks of this approach are biased estimates and underestimated parameter uncertainties [see, e.g., *van Straten*, 1985].

In this paper, a systematic approach for tackling the identifiability problem for large environmental simulation models is suggested. In order to specify the problem we start with a brief summary of the main issues of model and parameter identifiability in section 2. In section 3 we recall the classical nonlinear least squares estimation of the parameters. We discuss assessment of the fit, the use of the covariance matrix of the estimate, and likelihood contouring techniques. These methods are, however, limited to the case where all parameters can be estimated. In section 4 a strategy for tackling the parameter identifiability problem is presented for cases where the model is overparameterized with respect to given observations. This is nearly always the case with large simulation models. Parameter importance indices which account for the sensitivity of model results to single parameters are introduced in section 4.1, and collinearity indices and related measures which account for the degree of near-linear dependence of sensitivity functions of parameter subsets are introduced in section 4.2. Unless the number of parameters is small, these indices are preferred over the traditional sensitivity function plots. In section 4.3 we first discuss how parameter importance and collinearity indices can guide the selection of identifiable parameter subsets for parameter estimation. It is argued that the process of parameter subset selection is best done by iteratively calculating parameter importance and collinearity indices and performing parameter estimation on selected subsets until convergence is achieved. Subsequently, it is shown how the suggested tools facilitate the assessment of the fit including identification of potential biases, caused by keeping other parameters fixed. Practical issues are discussed in section 4.4. Finally, two small-scale case studies are presented in section 5 in order to illustrate different aspects of the suggested approach and to make comparisons of the different diagnostic tools.

2. Basic Identifiability Issues

Model identifiability analysis consists basically of two problems: the problem of model structure selection and the problem of parameter identification. Both issues have received much attention during the last decade in the environmental science and engineering literature [see, e.g., *Stigter and Beck*, 1994; *Kuczera and Mroczkowski*, 1998]. Important contributions have also been made from the point of view of uncertainty analysis [*Draper*, 1995; *Chatfield*, 1995]. Both *Draper* [1995] and *Chatfield* [1995] agree with *Stigter and Beck* [1994] that in many data analyses, model structure uncertainty is one of the relevant sources of model prediction uncertainty whereas parametric uncertainty often contributes only marginally to the total model prediction uncertainty.

In many applications, however, the problem of parameter identifiability of a given model structure is crucial. This is especially true when working with large environmental simulation models which are thought to summarize current scientific knowledge in a mathematical language. Such models, especially the ones we call "state-of-the-art" models, are built in order to reflect the current consensus of the scientific community about key processes in a specific environmental system [*Beck*, 1987, 1999]. In many case studies, these models are (successfully) applied without questioning the model structure. Adjusting some of the parameters often turns out to be sufficient in order to match reasonably well the data under study. In this context the problem of parameter identifiability of a given model structure arises naturally. It has, however, a slightly different meaning than in classical system identification. When working with large environmental simulation models, it is hardly reasonable to expect to be able to "identify true parameter values" from the data under study. The task is rather to find physically reasonable parameter values that describe the data adequately. In a complex, nonlinear model there are usually many such values, and the goal of identifiability analysis is to obtain some insight about these adequate parameter values. The main tool for this is sensitivity analysis.

For a nonlinear model, parameter sensitivity can be addressed either locally near a given point or over a large region in the space of physically reasonable parameter values. The regional approach introduced within the field of water quality modeling by *Hornberger and Spear* [1981] addresses the latter. It aims to identify those points or regions in the parameter space which lead to model outputs that match the data reasonably well. Some of these points or regions can be far apart, and current implementations of the Hornberger-Spear algorithm include modern statistical tools for density estimation with the aim to provide a flexible description of these regions [*Spear et al.*, 1994; *Spear*, 1997; *Grieb et al.*, 1999]. The approach suffers, however, from the "curse of dimensionality" because it is practically impossible to explore a high-dimensional parameter space without relying on further structural assumptions.

"Global sensitivity analysis," which has gained much attention during the last decade, is another approach to assess sensitivity over the whole parameter space [*Saltelli and Scott*, 1997; *Saltelli*, 1999]. For a review, see *Helton* [1993], and for an application of different methods in an environmental modeling

context, see *Campolongo and Saltelli* [1997]. It is primarily designed to assess the contributions of the model parameters to the variation in the model output. Parameters are varied over a predefined, reasonable region of the parameter space as in the Hornberger-Spear approach. The resulting variability in the model output then is analyzed using regression and analysis of variance-based approaches such as standardized regression, projection pursuit regression (PPR) [*Draper et al.*, 2000], Sobol's sensitivity indices [*Sobol*, 1993; *Saltelli et al.*, 1999], and Fourier amplitude sensitivity tests (FAST) [*Saltelli et al.*, 1999]. The global sensitivity analysis approach is particularly useful when one is primarily interested in quantifying the averaged influence of the parameters on the model output.

Local sensitivity analysis has been applied extensively to large kinetic reaction systems in chemometrics [see, e.g., *Turányi*, 1997]. For a review, see *Turányi* [1990], and for an application to a distributed water quality model, see *Pastres et al.* [1997]. Instead of varying the parameters over a predefined region in parameter space the local sensitivity approach considers the derivatives of the model output with respect to the parameters at a specific location in the parameter space. This approach has proven to be particularly successful in cases where systems are operated around a predefined location in parameter space. In addition, it is promising in cases where there are known parameter values that lead to an acceptable model output and the evaluation of the model is computationally too expensive to carry out a sensible regional analysis. A local sensitivity analysis then provides a comprehensive analysis of the "local region" that leads to a similar, acceptable model output. The local approach is also useful in combination with a regional approach, especially if the parameter space is high dimensional. In high dimensions, even if we compute the model output for 10,000 or 100,000 different parameter values, any two values are still far apart, so the information from a regional sensitivity analysis is necessarily coarse. In these cases, by means of local analyses carried out at a number of parameter values with an acceptable fit for the data, it is still possible to characterize the parameter subspace that leads to a similar, acceptable model output in terms of individual parameter sensitivities and parameter dependencies (see section 4) with reasonable effort.

Local sensitivity analysis has a close relationship to the classical nonlinear parameter estimation problem [*Bates and Watts*, 1988]. In cases where all model parameters can be estimated from the data under study the (local) parameter identifiability problem can be described adequately from the point of view of both local sensitivity analysis and classical parameter estimation. In numerous applications in environmental modeling, however, it is not feasible to estimate all model parameters from the data under study. Then the parameter identifiability problem is slightly different from the classical parameter estimation problem. Nevertheless, there is still an important relationship between local sensitivity analysis and parameter estimation that is worthwhile examining. In order to shed maximum light on this relationship we analyze the (local) parameter identifiability problem in this paper mainly from the perspective of parameter estimation. It is important to note, however, that this is a choice of presentation. The suggested approach is not restricted to cases where parameter estimation is carried out. It can be applied at any interesting point in the parameter space, not only at parameter values obtained by a parameter estimation algorithm.

3. Parameter Estimation and Assessing the Fit

To put the parameter estimation problem into a statistical framework, let us consider the following observation equation:

$$\mathbf{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \mathbf{Z}, \tag{1}$$

where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ is the observation vector and $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\eta_1(\boldsymbol{\theta}), \eta_2(\boldsymbol{\theta}), \ldots, \eta_n(\boldsymbol{\theta})]^T$ is the outcome vector of a deterministic model, evaluated at the same points in time and space as the observations $\mathbf{Y}$. The vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^T$ is the parameter vector of the deterministic model, and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ is the observation error vector. Note that observation equation (1) remains unchanged for different deterministic model classes. All features of the deterministic model except the parameter vector $\boldsymbol{\theta}$ are hidden in the symbol $\boldsymbol{\eta}(\ )$. The model outcome vector $\boldsymbol{\eta}(\boldsymbol{\theta})$ can therefore represent, for example, n calculated outcomes of a set of nonlinear partial differential equations at given locations in space and time, n calculated outcomes of a set of nonlinear ordinary differential equations at given time points, or, simply, n values of a nonlinear function of associated regressor variables.

Estimation of $\boldsymbol{\theta}$ in (1) is often done by weighted least squares. The weighted least squares estimator $\hat{\boldsymbol{\theta}}_{\text{WLS}}$ is defined to be the value minimizing the objective function

$$J = [\mathbf{Y} - \boldsymbol{\eta}(\boldsymbol{\theta})]^T \mathbf{W}[\mathbf{Y} - \boldsymbol{\eta}(\boldsymbol{\theta})] \tag{2}$$

over $\boldsymbol{\theta}$, where $\mathbf{W} = \text{diag}(w_1, w_2, \ldots, w_n)$ is a diagonal matrix of weights. Estimator $\hat{\boldsymbol{\theta}}_{\text{WLS}}$ is also the maximum likelihood estimator if we assume that $\mathbf{Z}$ in (1) is normally distributed with zero mean and $\text{Var}[\mathbf{Z}] = \sigma^2 \mathbf{W}^{-1}$.

The standard assessment of the fit in nonlinear regression is based on the linearization

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \approx \boldsymbol{\eta}(\boldsymbol{\theta}_0) + \left.\frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \boldsymbol{\eta}(\boldsymbol{\theta}_0) + \mathbf{V}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{3}$$

where

$$\mathbf{V} = \left.\frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \tag{4}$$

is the $n \times m$ derivative matrix evaluated at $\boldsymbol{\theta}_0$. This gives the estimated (approximate) covariance matrix of the estimate $\hat{\boldsymbol{\theta}}_{\text{WLS}}$ as

$$\widehat{\text{Var}}[\hat{\boldsymbol{\theta}}_{\text{WLS}}] = \frac{\text{WRSS}}{n - m}(\mathbf{V}^T \mathbf{W} \mathbf{V})^{-1}, \tag{5}$$

with

$$\text{WRSS} = [\mathbf{Y} - \boldsymbol{\eta}(\hat{\boldsymbol{\theta}}_{\text{WLS}})]^T \mathbf{W}[\mathbf{Y} - \boldsymbol{\eta}(\hat{\boldsymbol{\theta}}_{\text{WLS}})]$$

being the minimum weighted residual sum of squares, n being the number of observations, m being the number of parameters, and $\mathbf{V}$ being the derivative matrix evaluated at $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_{\text{WLS}}$. Given $\widehat{\text{Var}}[\hat{\boldsymbol{\theta}}_{\text{WLS}}]$ (approximate) standard errors, marginal confidence intervals and simultaneous confidence ellipsoids can be calculated [*Bates and Watts*, 1988].

Both the covariance matrix of the estimate and likelihood contouring techniques are used as diagnostic tools to detect poor parameter identifiability [e.g., *Kuczera*, 1990]. High off-diagonal absolute elements in the correlation matrix as well as long flat valleys in likelihood (WRSS) contour plots indicate strong dependencies of parameter estimates. In terms of pa-

rameter identifiability this means that the effects of changes in different parameter values on the model output can be self-canceling. In other words, different parameter values can lead to nearly the same model output; that is, the involved parameters are poorly identifiable. Likelihood contouring techniques are also used to check the validity of linear approximation (3) [*Bates and Watts*, 1988].

Despite the undoubted diagnostic merits of the correlation matrix of the estimate and the likelihood contour plots with respect to parameter identifiability, both tools have important shortcomings in the multiparameter case. As will be demonstrated and discussed in the case studies in section 5, the interpretation of the correlation matrix of the estimate with respect to parameter identifiability in the case where more than two parameters are estimated is far from clear. The likelihood contouring approach suffers mainly from two problems. On the one hand, it is not possible to display functions of more than two variables, and on the other hand, it becomes computationally increasingly difficult to evaluate the likelihood on an m-dimensional grid with increasing m. One can plot the likelihood evaluated on a series of two-dimensional grids corresponding to each pair of parameters, but the problem is how to choose the values of the other parameters. The simplest solution is to fix the other parameters at their least squares estimate, which leads to the conditional likelihood function. This is computationally simple but gives only selected cross sections of the contour. As will be demonstrated in the case studies in section 5, conditional likelihood plots can be very misleading with respect to parameter identifiability diagnosis. A more comprehensive view of the global behavior of the m-dimensional likelihood contour is obtained by the two-dimensional profile likelihood function, where the objective function is minimized over all the other parameters for each point of the two-dimensional grid. These profile likelihood plots are strongly preferred for parameter identifiability diagnostics [see *Bates and Watts*, 1988]. However, the evaluation of the two-dimensional profile likelihood on a series of grids can still be computationally expensive because for every point on the grids, $m - 2$ parameters have to be estimated with an iterative algorithm. Especially in cases where the evaluation of $\eta(\theta)$ is computationally expensive, the profile likelihood contouring approach is computationally impossible.

4. Parameter Identifiability Analysis for Overparameterized Models

In section 3 we discussed the parameter estimation problem for identifiable problems. When dealing with models containing practically nonidentifiable parameters, it is often not possible to compute the least squares estimates. A typical approach is to partition θ into two components $(\theta_K^T \theta_{\bar{K}}^T)^T$, where K is a subset of size k from $\{1, 2, \ldots, m\}$ and $\bar{K}$ is the complement of size $m - k$. Only the component θ_K is to be estimated from the data available whereas the other component $\theta_{\bar{K}}$ is fixed at an a priori value. The partitioning of θ is not straightforward. Trying out all possible subsets K is feasible only if m is small and the evaluation of $\eta(\theta)$ is fast. Otherwise a more systematic strategy is needed.

4.1. Screening for Important Parameters: Parameter Importance Indices δ

A natural starting point for the problem of identifying a subset K of k parameters that is potentially identifiable from a

given set of observations is the search for individual parameters that drive the variability in the model output. To assess the individual local parameter importance, we consider the sensitivity of the model output $\eta(\theta)$ to small changes in the parameter values θ at a specific location θ_0. This is given by the sensitivity matrix V defined in (4). Every column v_j, $j = 1, 2, \ldots, m$, represents the change in the model outcome vector $\eta(\theta)$ caused by a small change in θ_j divided by the small change in θ_j at the location θ_0.

The simplest way of calculating V is to use the finite difference approximation. Although this approach is known not to be computationally efficient in many cases, it is widely used because of its simplicity [*Turányi*, 1990]. In addition to simplicity, the calculation of V by means of finite differences is supported by the argument that we are indeed interested in the differences and not in the derivative itself.

In order to obtain dimension-free sensitivity information that enables comparisons we consider the scaled sensitivity matrix $S = \{s_{ij}\}$ with

$$s_{ij} = v_{ij} \frac{\Delta \theta_j}{SC_i}, \qquad i = 1, 2, \cdots, n \qquad j = 1, 2, \cdots, m. \tag{6}$$

Here v_{ij} denotes an element of V, $\Delta \theta_j$ is an a priori measure of the reasonable range of θ_j, and SC_i is a scale factor with the same physical dimension as the corresponding observation, accounting mainly for different scales of different output signals.

The norm of the columns s_j provides an obvious measure of the importance of individual parameters. A large norm $\|s_j\|$ means that a change of $\Delta \theta_j$ in the parameter θ_j has an important effect on the model outcome vector. This makes the parameter θ_j identifiable with the data available if all other parameters are fixed. In order to obtain additional information on the signs and the distribution of the values in each column we recommend the computation of the following five summaries:

$$\delta_j^{msqr} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} s_{ij}^2}, \tag{7}$$

$$\delta_j^{mabs} = \frac{1}{n} \sum_{i=1}^{n} |s_{ij}|, \tag{8}$$

$$\delta_j^{mean} = \frac{1}{n} \sum_{i=1}^{n} s_{ij}, \tag{9}$$

$$\delta_j^{max} = \max_i s_{ij}, \tag{10}$$

$$\delta_j^{min} = \min_i s_{ij}. \tag{11}$$

Large differences between δ_j^{msqr} and δ_j^{mabs} indicate, for example, a high variability or outliers in s_j. Checking δ_j^{max} and δ_j^{min} should help to distinguish between these two cases. The two summaries δ_j^{max} and δ_j^{min} are generally useful for outlier detection and also to know the range of s_j. A comparison of δ_j^{mabs} and δ_j^{mean} shows whether the elements of s_j have all the same sign, and δ_j^{mean} gives information on the sign of the averaged effect a change in a parameter has on the model output.

Ranking the parameters by one of the δ measures in de-

creasing order results in a parameter importance ranking. In the context of weighted least squares estimation of parameter subsets, δ_j^{msqr} is best suited to serve as a ranking criterion.

Note that the δ measures (7)–(11) can be very sensitive to the choice of the $\Delta\theta_j$, the scale factors SC_i, and changes in the experimental layout, respectively. In addition, they depend naturally on θ_0. A suitable choice of θ_0, $\Delta\theta_j$, and SC_i is therefore crucial (see section 4.4).

4.2. Identifiability of Parameter Subsets: Collinearity Index γ

In section 4.1 the tuning importance of individual parameters was discussed. In order to assess the identifiability of a subset K of k parameters ($1 < k \leq m$), however, we have to consider the joint influence of the parameters in K on the model output. In particular, we have to look for compensation effects of changes in the parameter values in K on the model output. This can be done by checking the degree of near-linear dependence within the column subsets S_k of the scaled sensitivity matrix S as defined in (6). If the s_j are (nearly) linearly dependent, a change in the model output caused by a change in a model parameter θ_j in K can be (nearly) compensated by appropriate changes in the other parameters' values in K. This prevents the parameters in K from being uniquely identifiable even if the model output is very sensitive to changes in the individual parameters. In such cases, parameter estimation procedures normally show poor convergence properties or even fail.

Plots of s_j, $j = 1, 2, \ldots, k$, against $i = 1, 2, \ldots, n$ have proven to be valuable diagnostic tools in order to detect near-linear dependencies among the s_j in cases where k is reasonably small. With these plots it is possible to detect near-linear dependencies and insensitive parameters fairly efficiently in cases where there are only a few parameters [Holmberg, 1982; Reichert et al., 1995]. However, if one wants to tackle the parameter identification problem for large parameter subsets, the graphical approach usually fails. With increasing size of the parameter subset and the model outcome vector the graphical output becomes increasingly confusing. In addition, it is very difficult to detect near-linear dependencies among the s_j that involve several parameters. In order to be able to tackle the parameter identifiability problem of large models and to detect multiple near-linear dependencies among the s_j we need diagnostics that are easier to handle and to interpret than the extensive graphical output.

The columns s_j, $j = 1, 2, \ldots, m$, of S are said to be linearly dependent or collinear if there exists a vector $\beta = (\beta_1, \beta_2, \ldots, \beta_m)^T$ with $\|\beta\| \neq 0$ such that $S\beta = 0$. If this equation holds approximately, the columns s_j, $j = 1, 2, \ldots, m$, are said to be nearly linearly dependent or nearly collinear. Much research has been dedicated to collinearity analysis in different fields, including numerics (solving of linear systems of equations) and statistics (linear regression diagnostics). According to the specific needs in the different fields, different collinearity measures were suggested by various authors. The review article by Stewart [1987] summarizes the approaches in numerics and statistics and proposes a unified view of the main issues in these two fields. A brief discussion of the collinearity problem in the context of linear regression diagnostics is given by Weisberg [1990], and a broader discussion is given by Belsley [1991].

A simple but effective approach to measure near collinearity is to look for the linear combination $S\beta$ that has minimal norm

under the constraint $\|\beta\| = 1$. It is well known that this minimum is achieved if β equals the normed eigenvector to the smallest eigenvalue λ_m of $S^T S$. The minimal norm $\|S\beta\|$ under $\|\beta\| = 1$ equals $\sqrt{\lambda_m}$ (for details, see Belsley [1991]). It turns out, however, that this measure is heavily dependent on the norms of the columns of S. Columns with large norms will be more important in determining the eigenvalues than will columns with small norms, and these differences will be reflected as strongly in the eigenvalues as will collinearity [Weisberg, 1990]. Therefore one should standardize the columns first.

Centering the columns is common practice in linear regression diagnostics, but it is not meaningful in nonlinear regression, where there is usually no intercept (see case study in section 5.1, Figure 4). Consequently, we consider the normalized matrix $\bar{S}$ with columns

$$\bar{s}_j = \frac{s_j}{\|s_j\|} \qquad j = 1, 2, \cdots, m. \tag{12}$$

To assess the degree of near-linear dependence of $k \leq m$ columns of $\bar{S}$, we define a collinearity index γ_K as follows:

$$\gamma_K = \frac{1}{\min_{\|\beta\|=1}\|\bar{S}_K\beta\|} = \frac{1}{\sqrt{\lambda_k}}, \tag{13}$$

with $\bar{S}_K$ being a $n \times k$ submatrix of $\bar{S}$ containing those columns that correspond to the parameters in K and λ_k being the smallest eigenvalue of $\bar{S}_K^T\bar{S}_K$.

The above definition has a simple interpretation: A change in the output vector $\eta(\theta)$ caused by a shift of a parameter $\theta_j \in K$ can be compensated, at least in the linear approximation, up to a fraction of 1 divided by the collinearity index γ_K by appropriate changes in the other parameters in K. A collinearity index of 20 therefore means that a change of the calculated results caused by a shift of a parameter $\theta_j \in K$ can be compensated to 5% by appropriate changes in the other parameters in K. A high value of a collinearity index γ_K thus indicates that the parameter set K is poorly identifiable even if the k individual parameters are among the top parameters of the parameter importance ranking. In order to get an overview of the identifiability of different parameter subsets we suggest that γ_K be calculated for all subsets K of the full parameter set M and that γ_K be plotted against the subset size. If M is of size greater than 20, it is computationally convenient to take subsets K from the top 20 parameters of the parameter importance ranking instead of the full set M.

Note that the definition (13) of the collinearity index has a close relationship to the condition number which is used as a measure of collinearity in different fields. The condition number κ_K of $\bar{S}_K^T\bar{S}_K$ is defined as

$$\kappa_K = \sqrt{\frac{\lambda_1}{\lambda_k}} = \frac{\max_{\|\beta\|=1}\|\bar{S}_K\beta\|}{\min_{\|\beta\|=1}\|\bar{S}_K\beta\|}, \tag{14}$$

with λ_1 being the largest and λ_k being the smallest eigenvalues of $\bar{S}_K^T\bar{S}_K$. Our collinearity index compares λ_k with unity, which is the value of $\|\bar{S}_K\beta\|$ if the columns are orthogonal, that is, maximally noncollinear. The value of λ_1 is of little interest in our context. This and the better interpretability of γ_K as discussed above are the reasons for preferring the index γ_K. In practice, however, differences between γ_K and κ_K are usually small since $1 \leq \lambda_1 \leq k$ by a result from linear algebra.

Note that γ_K, unlike the δ measures (7)–(11), does not depend on the choice of the $\Delta\theta_j$ because of normalization of S.

Nevertheless, γ_K still can be very sensitive to the choice of the scale factors SC_i and changes in the experimental layout, respectively, and it depends naturally on θ_0. Suitable choices of SC_i and θ_0 are therefore crucial (see section 4.4).

4.3. Parameter Identifiability Analysis Using Parameter Importance Indices δ and Collinearity Indices γ

In this section we discuss different uses of the indices introduced in sections 4.1 and 4.2 in typical parameter identifiability analyses. First, identifiability diagnosis for a specific parameter subset K is considered. A parameter subset K is said to be (potentially) identifiable if the observed model output is sufficiently sensitive to small changes of all parameters in K on an individual basis and if the collinearity index γ_K does not exceed a critical value. Although it is difficult to give precise criteria, we can nevertheless give some guidelines from our experience. Concerning individual importance, it is important that all parameters in K are in the upper part of the parameter importance ranking and that the δ_j^{msqr} for all parameters in K are approximately on the same order of magnitude. Critical values for γ_K lie in the range of 5–20 according to our experience. As can be seen from $\gamma_K \leq \kappa_K \leq \sqrt{k}\gamma_K$ this range of values is similar to the critical range for κ_K reported by *Belsley* [1991] in an econometric context. According to Belsley, collinearity problems going along with condition numbers below 10 are rare. On the other hand, condition numbers above 100 nearly always lead to severe parameter identifiability problems.

Second, we tackle the problem of choosing a subset K of k parameters out of the full set M of m parameters for parameter estimation. This typically involves several steps. In large applications it has proven to be useful to partition the full parameter set from an a priori point of view into two subsets as a first step: a subset of parameters that should preferably be estimated from the data available and another subset which could not reasonably be estimated from these data (M. Omlin et al., Biogeochemical model of Lake Zürich: Sensitivity, identifiability and uncertainty analysis, submitted to *Ecological Modelling*, 2000; R. Brun et al., Practical identifiability analysis of ASM2d with multicomponent time series data at operational WWTP scale, manuscript in preparation, 2000) (hereinafter referred to as M. Omlin et al., submitted manuscript, 2000, and R. Brun et al., manuscript in preparation, 2000, respectively). Possible reasons for fixing some parameters a priori include the availability of more direct experimentation techniques for these parameters. As a second step the parameter importance ranking is calculated for the subset of parameters that is possibly to be estimated. This step is followed by the calculation of the collinearity index for all possible subsets of the (approximately) top 20 parameters of the parameter importance ranking. Combining the information from the parameter importance ranking and the collinearity analysis leads to the choice of several potential identifiable parameter subsets with collinearity indices below the critical value and containing parameters which are among the top parameters in the parameter importance ranking. For further details of this step, see the case studies by Omlin et al. (submitted manuscript, 2000) and Brun et al. (manuscript in preparation, 2000). After having chosen potentially identifiable subsets, parameter estimation for these subsets is performed. At the new location in parameter space the second step is repeated, and we enter an iterative procedure with possibly several iterations until convergence is achieved.

Third, we consider the assessment of the fit once a subset of parameters has been estimated successfully. Primarily, standard tools like the covariance matrix of the estimate can be used, but we suggest that in addition, the collinearity indices be calculated for all subsets of the estimated parameters at $\hat{\theta}_K$. As will be demonstrated in the case studies, these collinearity indices are very helpful in order to identify the most problematic parameter subsets, and they complete, in a lucid way, the information that is present in the correlation matrix.

An assessment of the fit should include also an analysis of potential bias of the parameter estimates due to fixing some parameters a priori. Under the assumption that $\eta(\theta)$ represents the "true model" structure, estimation of a subset K of parameters leads to biased estimates $\hat{\theta}_K$ unless the fixed parameters $\theta_{\bar{K}}$ are at the true parameter values. As the true values $\theta_{\bar{K}}$ are practically never known, parameter subset estimation will nearly always lead to a problem of biased estimates. The calculation of suitable collinearity indices can help to identify the fixed parameters which are the main (potential) bias sources. A subset $J \subset \bar{K}$ of sensitive, fixed parameters is considered to be an important (potential) bias source if $\gamma_{K \cup J}$ is large whereas γ_K and γ_J are reasonably low. A large $\gamma_{K \cup J}$ means that some of the estimated values $\hat{\theta}_K$ are heavily dependent on the fixed parameter values θ_J. In this case, $\hat{\theta}_K$ has to be understood clearly as a conditional estimate, and this fact has to be taken into account when comparisons with values reported in the literature are made. To assess potential bias sources practically, we suggest calculating collinearity indices for all subsets $K \cup J$, with K being the subset of estimated parameters and $J \subset \bar{K}$ being a subset of sensitive, fixed parameters. If $\bar{K}$ is of large size, the analysis is typically restricted to a subset size of J of 1 or 2 (Omlin et al., submitted manuscript, 2000; Brun et al., manuscript in preparation, 2000).

4.4. Practical Considerations

Three crucial questions of practical relevance remain. First, how to choose a suitable θ_0 to start the parameter identifiability analysis; second, how to choose $\Delta\theta_j$ in (6); and third, how to choose SC_i in (6).

The approach outlined in sections 4.1–4.3 is based on the assumption that there is a known point θ_0 in the parameter space which leads to an acceptable model output. In cases where data are available, θ_0 might be chosen such that the model fits the data reasonably well. If significant prior knowledge about parameter values is available, a suitable θ_0 may be found combining physical reasoning and parameter importance screening information as discussed in section 4.1. In the case where no data are available, θ_0 might be a reasonable "operating point" gained from the literature or a number of points gained from a regional sensitivity analysis.

For a suitable choice of $\Delta\theta_j$, $j = 1, 2, \ldots, m$, we need a prior range of reasonable values for θ_j based on the literature or on expert knowledge. This is similar to specifying a prior distribution in a Bayesian analysis. A typical choice for $\Delta\theta_j$, $j = 1, 2, \ldots, m$, is a quarter of the reasonable range specified for θ_j. Assuming a normal prior distribution for θ_j and assuming that the reasonable range is a 95% prior interval, this choice corresponds to making $\Delta\theta_j$ equal to the prior standard deviation of θ_j. In cases where there is only little prior knowledge available a reasonable choice of $\Delta\theta_j$, $j = 1, 2, \ldots, m$, can be the value θ_0 itself. This choice accounts at least for different scales of the parameters.

Whereas the $\Delta\theta_j$ account for different scales and ranges of possible values of the parameters, the SC_i, $i = 1, 2, \ldots, n$,

account for different scales of different outputs. Usually, SC_i are chosen to be constant for the same output variable. If parameter estimation is performed by weighted least squares, SC_i are best chosen to be $\sqrt{1/w_i}$ whereas the w_i are usually chosen in order to ensure that observations of different output variables (with different scales) have a similar influence on the objective function. Outside the context of weighted least squares, SC_i are best set to a typical or mean value of the corresponding output variable.

5. Case Studies

Two case studies are presented in order to demonstrate different aspects of the suggested approach in particular applications. In order to make the presentation short and clear the case studies involve deterministic models which are over-parameterized with respect to the data available for identification but which are not large in the sense of having tens or even hundreds of parameters. In addition, this should enable a comparison of the suggested measures of identifiability with traditionally applied diagnostic tools that are restricted to models with a small number of parameters (e.g., sensitivity function plots) or to cases where the full parameter set can be estimated (e.g., the correlation matrix of parameter estimates and likelihood contouring). Applications to large deterministic models with many parameters are given by Omlin et al. (submitted manuscript, 2000) and Brun et al. (manuscript in preparation, 2000).

The first case study is dedicated to a common issue in water quality modeling: The modeling of dissolved oxygen in a river. The deterministic model is given by a single partial differential equation. Containing four parameters and a single state variable, the model is of modest complexity only. The aim of this first study is mainly to compare the different diagnostic tools. The second case study deals with microbial laboratory data. It is based on published experimental work and modeling by *Sommer* [1997]. Growth of biomass and simultaneous degradation of two substrates are described by a model of the widespread Monod type. The deterministic model consists of a set of ordinary differential equations. Containing 13 parameters and three state variables, the model is of moderate complexity already. The aim of this study is mainly to show how the suggested measures can guide the selection of a subset of identifiable parameters out of the full parameter set and how they facilitate the assessment of the fit including identification of potential biases caused by keeping some parameters fixed. In addition, this second study demonstrates how the dependence of parameter identifiability on the experimental layout is reflected by the suggested measures.

All calculations including numerical solution of differential equations, sensitivity analysis, and parameter estimation were done with AQUASIM, a computer program designed for simulation and data analysis of aquatic systems [*Reichert*, 1994, 1998].

5.1. Dissolved Oxygen in the River Glatt

5.1.1. River Glatt.
The Glatt catchment is located in the northeastern part of the Swiss Plateau. The river Glatt, outflow of the eutrophic lake Greifensee (river kilometer 0.0), flows through a densely populated area to the river Rhine (confluence at river kilometer 35.5). The river is channelized. In the study reach (river kilometers 26.1–35.1), banks are protected with a stone riprap. Five artificial cascades and boulder ramps

(length approximately equal to 8 m, height 0.9–1.76 m) and small drops every 50 m along the river control channel slope, prevent bed erosion during high flow, and cause high reaeration rates [*Uehlinger et al.*, 2000].

5.1.2. Model and data.
The following oxygen mass balance model is used to describe the O_2 dynamics in the investigated study reach:

$$\frac{\partial C_{O_2}}{\partial t} + v \frac{\partial C_{O_2}}{\partial x} = r_{\text{prod}} + r_{\text{resp}} + r_{\text{ex}}. \tag{15}$$

In (15), C_{O_2} ($gO_2\ m^{-3}$) is the dissolved oxygen concentration, v (md^{-1}) is the mean current velocity, r_{prod} ($gO_2\ d^{-1}\ m^{-3}$) is the oxygen production rate due to gross primary production, r_{resp} ($gO_2\ d^{-1}\ m^{-3}$) is the oxygen consumption rate due to respiration, and r_{ex} ($gO_2\ d^{-1}\ m^{-3}$) is the oxygen transfer rate due to gas exchange between water and the atmosphere along river sections between cascades or ramps. Production, respiration, and gas exchange along river sections are parameterized as follows:

$$r_{\text{prod}} = \frac{1}{d} PI, \tag{16}$$

$$r_{\text{resp}} = -\frac{1}{d} R, \tag{17}$$

$$r_{\text{ex}} = K_{\text{gas}}(C_{O_2}^{\text{sat}} - C_{O_2}). \tag{18}$$

In (16)–(18) d (m) is the mean water depth, P ($W^{-1}\ gO_2\ d^{-1}$) is the production rate constant, I (Wm^{-2}) is the light intensity at the water surface, R ($gO_2\ m^{-2}\ d^{-1}$) is the ecosystem respiration rate per unit surface area of the river bed, K_{gas} (d^{-1}) is the gas exchange constant accounting for gas exchange along river sections, and $C_{O_2}^{\text{sat}}$ is the saturation concentration of C_{O_2}.

Across cascades or ramps, air bubbles enhance reaeration. The reaeration efficiency E_{gas} of a ramp describes the extent that supersaturation or subsaturation will be reduced across the ramp [*Uehlinger et al.*, 2000]:

$$E_{\text{gas}} = \frac{C_{O_2}^{\text{up}} - C_{O_2}^{\text{down}}}{C_{O_2}^{\text{up}} - C_{O_2}^{\text{sat}}}. \tag{19}$$

In (19), $C_{O_2}^{\text{up}}$ is the O_2 concentration upstream of the ramp, and $C_{O_2}^{\text{down}}$ is the O_2 concentration downstream of the ramp.

Dissolved oxygen, temperature, and discharge data used in this study were kindly provided by the Swiss National Hydrological and Geological Survey (NADUF program, *Jakob et al.* [1994]). Instead of light intensity measurements at the river surface of the study reach, global radiation data recorded by the Swiss Meteorological Institute (SMA) at the Zürich Airport ($\approx$8 km south of the study reach) were used for the light intensity I. Data on channel morphology were provided by the Water Protection Authority of the Canton Zürich (AWEL).

5.1.3. Parameter estimation and parameter identifiability analysis.
As the model equations (15)–(19) contain only four parameters that are potentially to be estimated (P, R, K_{gas}, and E_{gas}) and their numerical solution is rather cheap computationally, we start the analysis with parameter estimation. Because there are only observations from a single state (oxygen), weights in (2) are set to 1, and weighted least squares simplifies to ordinary least squares. Following the notation in (2), $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ denotes the O_2 time series data recorded at the end of the study reach, $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\eta_1(\boldsymbol{\theta}),$

Table 1. Parameter Estimates, Absolute and Relative Standard Errors, and Approximate Correlation Matrix of Parameter Estimates for P, R, and K_{gas}, With E_{gas} Being Fixed to 0.5

| Parameter | Estimate | Standard Error | | Correlation Matrix | | |
		Absolute	Relative	K_{gas}	P	R
K_{gas}	48.4	13.8	0.28	1.000		
P	0.104	0.0204	0.20	0.995	1.000	
R	38.7	7.42	0.19	0.993	0.997	1.000

$\eta_2(\theta), \ldots, \eta_n(\theta)]^T$ denotes the predicted O_2 concentrations at kilometer 35.1 at the same time points where O_2 data were recorded, and $\theta = (P, R, K_{gas}, E_{gas})^T$ denotes the parameter vector.

A first attempt to estimate the full parameter vector failed because of lack of convergence of the estimation algorithm. The model is clearly overparameterized with respect to the O_2 time series data available. This is rather obvious because it is not possible to discriminate the two gas exchange processes with oxygen time series data from a single location. We therefore fixed the parameter E_{gas} to a reasonable value of 0.5, using a formula used by *Uehlinger et al.* [2000], and tried to estimate the remaining three parameters. Quick convergence was achieved to the estimates shown in Table 1. Predicted and observed O_2 concentrations are shown in Figure 1.

A first assessment of the fit based on the estimated covariance matrix of the estimates in Table 1 reveals that in spite of quick convergence the parameter subset $\{P, R, K_{gas}\}$ is only poorly identifiable. All elements of the correlation matrix of parameter estimates are above 0.99, and the estimated standard errors are considerably high. The poor identifiability is also shown by the profile likelihood contours in Figure 2. The long flat valleys mean that parameter values can be varied over a large range in the direction of the valley without significant change in the model output. Although profile likelihood contours show a moderate departure from the linear approximation (WRSS increases more slowly in the direction of increasing parameter values, and two ellipses are slightly curved), the linear approximation is nevertheless useful. In particular, the direction of the valley is obtained correctly. We will comment on the conditional likelihood plots below.

To analyze the parameter identifiability problem further, we first produced a sensitivity function plot and calculated δ measures according to (7)–(11). The $\Delta\theta_j$ in (6) was set to θ_j, and SC_i was set to 1 ($gO_2 \ m^{-3}$). Results are shown in Figure 3 and Table 2. The continuous sensitivity functions in Figure 3 give a good overview of the identifiability situation. In order to take into account the present experimental layout we added symbols corresponding to the columns of S. The plot shows that the model output is considerably sensitive to all four parameters. It is most sensitive to P and R, less sensitive to K_{gas}, and even less sensitive to E_{gas}. This finding is confirmed by the parameter importance ranking in Table 2, which basically provides a statistical summary of the information in the sensitivity function plot. It reflects the fact that production, respiration, and reaeration along the river reach downstream of the last cascade have a larger effect on the measured oxygen concentration than the oxygen concentration immediately downstream of the cascade. From both the sensitivity function plot and the parameter importance ranking we can conclude that

every parameter has to be considered as potentially identifiable on an individual basis.

To assess the identifiability of groups of parameters, we made pairwise scatterplots of the columns of S and calculated collinearity indices of all parameter subsets. Results are shown in Figure 4 and Table 3. From Figure 4 we see that the sensitivities of $\{P, E_{gas}\}$, $\{P, K_{gas}\}$, and $\{K_{gas}, E_{gas}\}$ are strongly correlated whereas the sensitivities of the other pairs are not. It is important to note, however, that the correlation does not cause identifiability problems in the cases of $\{P, E_{gas}\}$ and $\{P, K_{gas}\}$. Sensitivities of $\{P, E_{gas}\}$ and $\{P, K_{gas}\}$ are indeed correlated, but they are not (nearly) collinear like the sensitivities of $\{K_{gas}, E_{gas}\}$. The difference is that the regression line relating the two sensitivities passes close to the origin in the case of $\{K_{gas}, E_{gas}\}$, but not in the cases of $\{P, E_{gas}\}$ and $\{P, K_{gas}\}$. The effect of a change in K_{gas} on the model output can therefore be nearly compensated by an appropriate change in E_{gas}. In contrast, the effect of a change in the parameter P on the model output can only be made constant by an appropriate change in K_{gas} and E_{gas}, respectively, but not approximately zero. This is due to the sensitivity of P being equal to zero during nocturnal hours (primary production occurs only when enough light is available) whereas the gas exchange process continues during nighttime (see Figure 3).

The collinearity indices in Table 3 confirm the findings from Figure 4 for pairs of parameters: Columns of S corresponding to the parameters $\{K_{gas}, E_{gas}\}$ are nearly collinear ($\gamma_K = 9.07$), but other pairs of columns are not ($\gamma_K < 2.1$). In addition, Table 3 provides information concerning multiple near-linear dependencies. The parameter subset $\{P, R, E_{gas}\}$ has a collinearity index of approximately 24, and the collinearity index of the full set $\{P, R, K_{gas}, E_{gas}\}$ is almost 60. These high collinearity indices point to the severe identifiability problems we found in the first stage of the analysis. Parameter estimation of the full parameter set failed because of lack of convergence, and estimation of $\{P, R, K_{gas}\}$ resulted in large estimated standard errors and high off-diagonal elements of the estimated correlation matrix (Table 1). The collinearity indices in Table 3 now show that it is the near collinearity of the triple $\{P, R, K_{gas}\}$ that causes the elements of the correlation matrix in Table 1 to be high and not the near collinearity of pairs of sensitivities as might be concluded from a naive interpretation of the correlation matrix.

This fact is also illustrated comparing profile and conditional
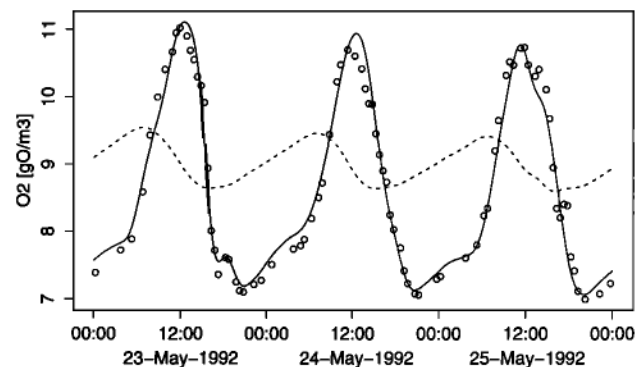
Figure 1. Observed (symbols) and predicted (solid line) O_2 concentrations in the river Glatt at kilometer 35.1. In addition, the calculated oxygen saturation concentration $C_{O_2}^{sat}$ is shown (dashed line).
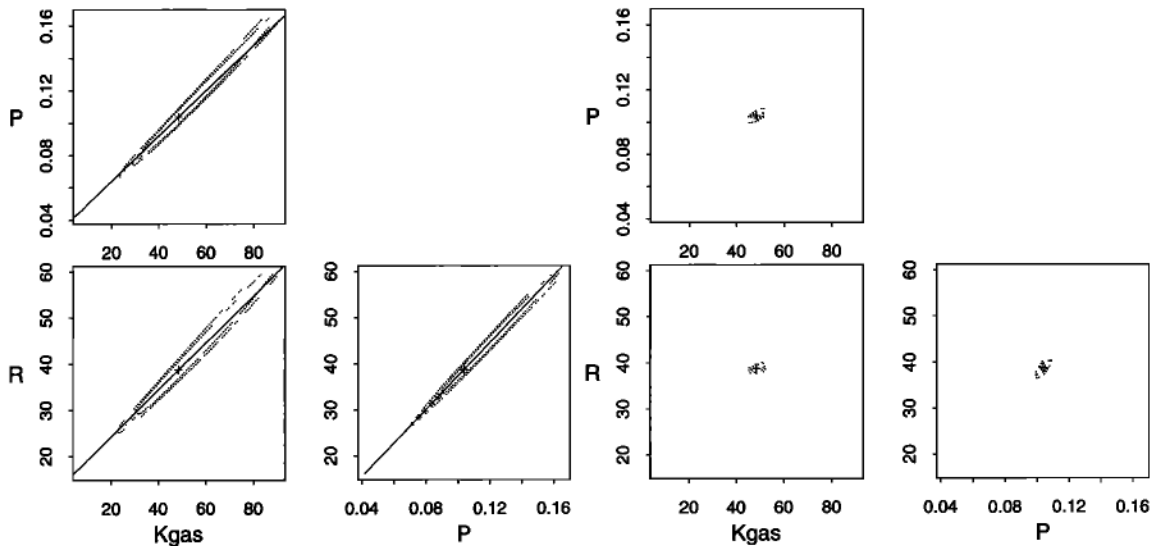
Figure 2. Profile (left-hand side) and conditional (right-hand side) likelihood (WRSS) contour plots (see section 3). The least squares estimate from Table 1 is shown as a cross in all plots. The contours refer to the 80 and 95% confidence regions in the profile likelihood plot and to the same WRSS values in the conditional likelihood plot. The added solid line in the profile likelihood plot is the projection of the eigenvector corresponding to the (smallest) eigenvalue which determines the collinearity index of the parameter subset $\{P, R, K_{gas}\}$ to be high (see Table 3). This eigenvector has the same direction as the main axis of the confidence ellipsoid calculated from linear approximation.

likelihood plots in Figure 2. The long flat valleys of the profile likelihood correspond to the high collinearity index of the parameters subset $\{P, R, K_{gas}\}$ and to the high off-diagonal elements of the correlation matrix in Table 1. They show the poor identifiability of $\{P, R, K_{gas}\}$. The conditional likelihood plot on the right-hand side, however, suggests a different picture. Contours are nearly circular, and there seems to be no parameter identifiability problem. However, this is certainly not true. The nearly circular contours only mean that the identifiability problem is not caused by pairs of parameters, which we know already from the collinearity indices shown in Table 3. It is important to note that the common interpretation of the correlation matrix of the estimate in the sense of pairwise linear dependencies can be quite misleading. Correlation matrices of estimates have to be interpreted in the profile and not in the conditional likelihood sense. As this is rather difficult to do in the multidimensional case, the calculation of

collinearity indices for all parameter subsets can facilitate a correct interpretation of the correlation matrix of the parameter estimate.

In order to illustrate the collinearity index graphically and to demonstrate the poor identifiability of $\{P, R, K_{gas}\}$ in more detail, we performed parameter estimations of $\{P, R\}$ for different values of K_{gas}. Relative changes of estimated parameters and WRSS values are shown in Figure 5. Changing the value of K_{gas} and holding the other parameters fixed at their initial least squares estimate result in a sharply increasing (conditional) cWRSS. This reflects the fact that the model output is fairly sensitive to changes in K_{gas}. However, if $\{P, R\}$ are allowed to change (they are estimated for different values of K_{gas}), changes in cWRSS due to changes in K_{gas} can be compensated to approximately $1/\gamma_K \approx 4\%$, which is shown by the flat (profile) pWRSS line. A high collinearity index therefore means that the one-dimensional profile pWRSS is flat, compensating the possibly steep cWRSS to a large extent. A comparison of Figure 2 and Figure 5 reveals that changes in the estimates of P and R follow nearly exactly the direction of the long flat valleys which is approximately given by the eigenvector corresponding to the (smallest) eigenvalue which determines the collinearity index of the parameter subset $\{P, R, K_{gas}\}$ to be high. We further see that the pWRSS curve is asymmetric. Apart from a departure from the linear approximation, this indicates that the parameters $\{P, R, K_{gas}\}$ are
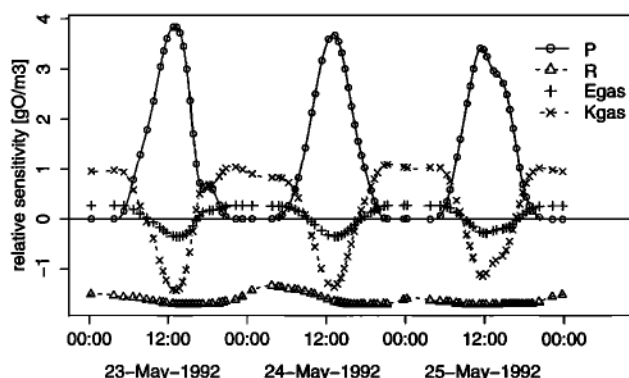

Figure 3. Continuous relative sensitivity functions of the O_2 concentration with respect to the parameters P, R, K_{gas}, and E_{gas} shown as lines. Symbols corresponding to the columns of **S** are added in order to indicate the actual experimental layout.

Table 2. Parameter Importance Ranking[a]

Parameter	δ^{msqr}	δ^{mabs}	δ^{mean}	δ^{max}	δ^{min}
P	1.96	1.45	1.45	3.84	0.00
R	1.63	1.62	−1.62	−1.33	−1.70
K_{gas}	0.84	0.77	0.08	1.09	−1.42
E_{gas}	0.22	0.20	0.02	0.27	−0.35

[a]The δ measures are defined in equations (7)–(11).
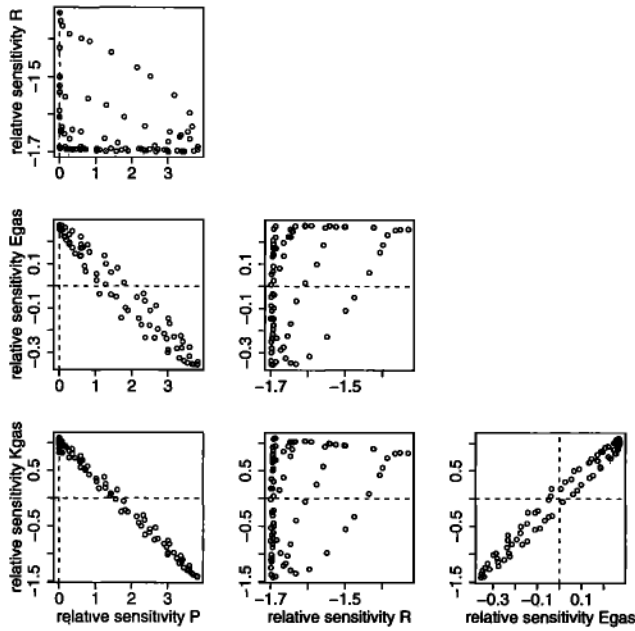
Figure 4. Pairwise scatterplots of columns of sensitivity matrix S. Horizontal and vertical dashed lines are added to show the origin.

better identifiable if gas exchange is small, a fact we already noticed in Figure 2. This is actually true because a smaller gas exchange coefficient leads to a slower approximation of the oxygen concentration to the equilibrium value during the night (which itself is not constant because of the temperature dependence of the saturation concentration, see Figure 1), and this makes the three parameters more identifiable. In highly aerated rivers, however, it is not possible to identify production and respiration rates from oxygen measurements alone.

5.2. Microbial Degradation Experiment

The case study presented in this section is based on published experimental work and modeling including parameter estimation [*Sommer*, 1997].

5.2.1. Experimental setup. The degradation experiment was carried out in a stirred batch reactor under sterile aerobic conditions. The aqueous medium consisted of benzene and toluene, bacteria, growth medium, and distilled water. The biomass was a pure culture, identified as Pseudomonas cepacia, and originated from a ground water sample from a former

gasworks site. Benzene and toluene served as the sole carbon source, and the growth medium supplied the biomass with nitrogen, phosphorus, and other minerals necessary for bacterial growth. Substrate and biomass samples were taken from the liquid phase. The toluene and benzene samples were measured directly on a gas chromatograph, and the biomass concentration was determined measuring the protein content. Further details of the experimental procedure are documented by *Sommer* [1997].

5.2.2. Model, data, and basic simulation. The experiment is described by a modified version of J. Bailey and D. Ollis' model [see *Sommer*, 1997, and references therein]. It consists of three nonlinear first-order differential equations describing the degradation of two individual substrates and growth of biomass. The model is further expanded by a time lag t_{lag} accounting for a lag phase at the beginning of the experiment.

$$\frac{dS_t}{dt} = -k_t \frac{S_t}{K_{st} + S_t + z_b S_b} X \qquad t \geq t_{lag}, \qquad (20)$$

$$\frac{dS_b}{dt} = -k_b \frac{S_b}{K_{sb} + S_b + z_t S_t} X \qquad t \geq t_{lag}, \qquad (21)$$

$$\frac{dX}{dt} = Y_t k_t \frac{S_t}{K_{st} + S_t + z_b S_b} X + Y_b k_b \frac{S_b}{K_{sb} + S_b + z_t S_t} X - bX$$
$$t \geq t_{lag}, \qquad (22)$$

$$\frac{dS_t}{dt} = \frac{dS_b}{dt} = \frac{dX}{dt} = 0 \qquad t < t_{lag}. \qquad (23)$$

States and parameters are explained in Tables 4 and 5, respectively. Degradation and growth processes are formulated with standard Monod kinetics. An inhibition effect of the presence of benzene and toluene on the degradation of toluene and benzene, respectively, is added in (20) and (21). Initial concentrations S_b^{ini}, S_t^{ini}, and X^{ini} are treated as model parameters. Note that the parameters k_t and k_b do not have the usual biological meaning of maximal degradation rates because they

Table 3. Collinearity Indices for All Parameter Subsets

Parameter Subset	γ_K
E_{gas}, K_{gas}	9.07
P, R	2.02
K_{gas}, P	1.57
E_{gas}, P	1.55
K_{gas}, R	1.04
E_{gas}, R	1.04
K_{gas}, P, R	23.81
E_{gas}, P, R	9.13
E_{gas}, K_{gas}, P	9.10
E_{gas}, K_{gas}, R	9.09
E_{gas}, K_{gas}, P, R	59.49
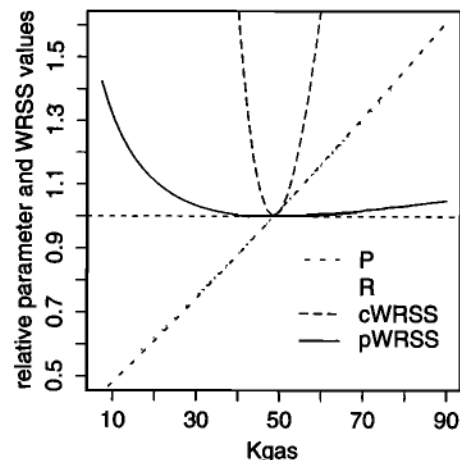
Figure 5. Relative changes of estimated parameters and WRSS values for different values of K_{gas}. Here cWRSS denotes one-dimensional conditional likelihood (WRSS) values, and pWRSS denotes one-dimensional profile likelihood (WRSS) values.

Table 4. Symbols, Units, Scale Factors, Number of Observations, and Description of the State Variables[a]

State	Unit	SC_i	n_{obs}	Description
S_t	mg L^{-1}	0.10	25	concentration of toluene in liquid phase
S_b	mg L^{-1}	0.10	28	concentration of benzene in liquid phase
X	mg L^{-1}	0.35	37	concentration of suspended biomass

[a]SC_i, scale factors; n_{obs}, number of observations.

include effects of yield coefficients and head space factors (two additional factors due to the experimental setup).

The data consist of a total of 90 observations including measurements of all state variables (see also Table 4). It is publicly available from the Institute of Mathematical Modelling, Technical University of Denmark, at http://www.imm.d-tu.dk/documents/ftp/phdliste/phd31data/run4b.inp. In Figure 6 the data are shown together with the deterministic model output for $\theta = \theta_0$ as specified in Table 5. Note that the values θ_0 (except z_b) are fitted values (see Table 8).

5.2.3. Parameter identifiability analysis and parameter estimation. Parameter identifiability analysis and parameter estimation are done for two experimental layouts in order to give a motivating example with regard to possible applications of the suggested approach in the field of experimental design for nonlinear models. The first (layout 1) corresponds to the full data set. This means that we look at a design that takes observations at exactly the same points where data are actually available. The second (layout 2) corresponds to a reduced data set with observations only of the states S_b and S_t at the same points as in layout 1. The second layout is typical for applications where only substrates are measured and not biomass.

In order to provide reasonable estimates of prior parameter uncertainties, parameters were classified into three classes from a biological point of view: accurately known parameters (class 1, relative uncertainty 5%), moderately inaccurately known parameters (class 2, relative uncertainty 20%), and poorly known parameters (class 3, relative uncertainty 50%). Initial concentrations of substrates were assumed to be known accurately (class 1). Maximum degradation rates, decay rate of biomass, yields, and the parameters t_{lag} and X^{ini} were assumed to be known moderately inaccurately (class 2). Finally, half-saturation coefficients and inhibition coefficients were as-

sumed to be known only poorly (class 3) (see Table 5). Because z_b equals zero, an absolute prior uncertainty was assumed directly for this parameter.

In Table 6, results of the parameter importance screening are given. Scale factors were chosen according to Table 4, and prior parameter uncertainties were chosen according to Table 5. The values of δ^{msqr} decrease quasi-continuously down to 10% of the maximum for both layouts. Table 6 shows a very small sensitivity of the model output to the value of the half-saturation coefficient K_{sb}. This is not quite astonishing because typical experimental concentrations are much larger than K_{sb}, and the inhibition term $z_t S_t$ makes the contribution of K_{sb} to the denominator of the Monod factor even less important. Therefore the value of K_{sb} influences the shape of the decreasing benzene concentration only within a small time domain in which the concentration as a function of time bends toward the time axis in Figure 6. The model output is, however, much more sensitive to the value of K_{st}. This is due to the slightly larger value of K_{st} (initial toluene and benzene concentrations are similar), to the absence of the inhibition term (note that $z_b = 0$), and to the effect of toluene as an inhibitor for growth on benzene. Further conclusions can be drawn from the results in Table 6. The δ^{msqr} of Y_b and b is smaller for layout 2 than for layout 1. This indicates that observations of the biomass play a key role in identification of these two parameters. The δ^{max} and δ^{min} show that positive changes in z_b, K_{st}, t_{lag}, z_t, b, and K_{sb} lead to higher substrate concentrations over the whole time interval for layout 2 and that positive changes in k_t, Y_t, X^{ini}, k_b, and Y_b lead to lower substrate concentrations. This is also reflected by the fact that δ^{mabs} and δ^{mean} of these parameters maximally differ in the sign for layout 2. From a biological point of view this finding is obvious, because layout 2 considers only observations of the substrates.

Table 5. Symbols, Units, Values (θ_0), Prior Uncertainty, and Description of Model Parameters[a]

Parameter	Unit	θ_0	Prior Uncertainty Absolute ($\Delta\theta$)	Prior Uncertainty Relative	Description
b	h^{-1}	0.0367	0.00734	0.20	decay coefficient of biomass
k_b	h^{-1}	0.229	0.0458	0.20	maximum degradation rate for benzene
K_{sb}	mg L^{-1}	0.477	0.239	0.50	half-saturation coefficient for benzene
K_{st}	mg L^{-1}	0.863	0.431	0.50	half-saturation coefficient for toluene
k_t	h^{-1}	0.350	0.0699	0.20	maximum degradation rate for toluene
S_b^{ini}	mg L^{-1}	4.72	0.236	0.05	initial concentration of benzene
S_t^{ini}	mg L^{-1}	5.11	0.255	0.05	initial concentration of toluene
t_{lag}	h	2.08	0.417	0.20	time lag
X^{ini}	mg L^{-1}	0.803	0.161	0.20	initial concentration of biomass
Y_b	...	0.639	0.128	0.20	yield coefficient for benzene
Y_t	...	1.14	0.229	0.20	yield coefficient for toluene
z_b	...	0	0.250	...	benzene inhibition coefficient
z_t	...	1.18	0.588	0.50	toluene inhibition coefficient

[a]Parameter values (θ_0), except the value for z_b, are fitted values using the full data set (see fit results in Table 8). Prior parameter uncertainty is given as absolute ($\Delta\theta$) and relative values.
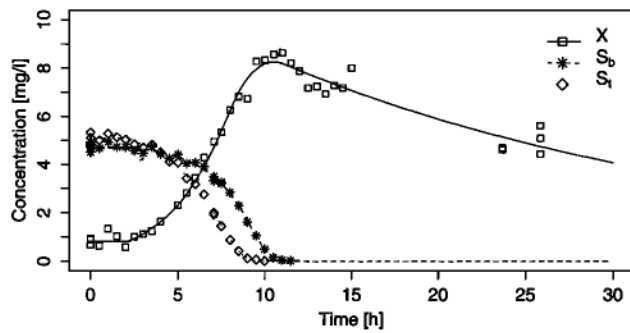
Figure 6. Data (symbols) and deterministic model output (lines) for $\theta = \theta_0$ (see Table 5).

In order to get information on the identifiability of subsets of parameters, collinearity indices γ_K were calculated for all subsets K of the full parameter set M. The calculations were done for both experimental layouts. A graphical representation of the results is given in Figure 7, where the collinearity index versus parameter subset size is plotted. Note that the ordinate is limited to $\gamma_K < 50$ in order to show the information in the interesting range of γ in more detail. The collinearity index for the full set M reaches to 380 in the case of layout 1 and to 2900 in the case of layout 2. This makes clear that the full parameter set is not identifiable either for layout 1 or for layout 2. The plots in Figure 7 show interesting features. From the left-hand-side plot we see that despite the careful design of the experiment there is a pair of parameters with a collinearity index greater than 25. Furthermore, the plot shows that the maximum collinearity index increases quickly with the size of the parameter subset. This means that there are many parameter subsets of moderate size that are practically nonidentifiable. On the other hand, the same plot shows that the minimum collinearity index increases quite slowly with increasing parameter subset size. There is indeed a subset of size 12 with a collinearity index of 20, which is even smaller than the maximum collinearity index of subset size 2. These features lead to the conclusion that with data collected according to layout 1 there is probably only one parameter to be omitted from the full set of 13 parameters in order to be able to perform weighted least squares parameter estimation.

The right-hand plot shows similar features, but both the maximum and the minimum collinearity indices increase faster than in the case of layout 1. This is not astonishing bearing in mind the fact that layout 2 is poorer than layout 1 both quantitatively (53 instead of 90 design points) and qualitatively (two state variables observed instead of three). In the case of layout 2 there are three parameter subsets of size 2 that show a collinearity index greater than 20. On the other hand, there is a subset of size 8 with a collinearity index smaller than 20. Subsets with size greater than 9 do not appear in the plot; they all have collinearity indices greater than 50. The conclusion we draw from this is that it should be possible to estimate a subset of no more than 8 parameters by weighted least squares with data collected according to layout 2.

In Table 7, collinearity indices for specific parameter subsets are shown. In Figure 7 the points corresponding to these subsets are labeled with the number of the subset. Considering the model equations (20)–(22), the high collinearity indices of subsets 4 and 8 are evident: Higher degradation rates k_t and k_b can nearly be compensated by an appropriate adjustment of K_{st} and z_b and K_{sb} and z_t, respectively. The particularly high collinearity index of subset 4 makes clear that every parameter set including this subset is nonidentifiable either with layout 1 or with layout 2. Considering the collinearity indices of subsets 1–3, it seems best to omit parameter z_b from parameter estimation. A look at the collinearity indices of subsets 9 and 10 confirms this decision: Dropping z_b in set 9 decreases the collinearity index dramatically for both layout 1 and layout 2. The resulting collinearity index is only slightly higher than the collinearity index of subset 8. It is interesting to note that collinearity indices of subsets 5–8 are higher than the collinearity indices of the corresponding subsets 1–4. This finding is not evident considering the model equations (20)–(22). However, it becomes clear when looking at Figure 6. Degradation of S_t is faster than degradation of S_b. Many design points for S_t show a nearly constant S_b. This makes the identification of z_b very difficult. Collinearity indices of subsets 11 and 12 show that initial concentrations of S_b and S_t and the time lag parameter t_{lag} are well-identifiable parameters for both layout 1 and layout 2. The reason is that both experimental layouts provide measurements of S_b and S_t at the initial state of the experiment and during the lag phase (see Figure 6). Whereas there is no significant difference in the collinearity indices between layout 1 and layout 2 regarding subsets 1–10, there are large

Table 6. Parameter Importance Ranking for Layout 1 and Layout 2[a]

Parameter	Layout 1					Layout 2				
	δ^{msqr}	δ^{mabs}	δ^{mean}	δ^{max}	δ^{min}	δ^{msqr}	δ^{mabs}	δ^{mean}	δ^{max}	δ^{min}
z_b	5.16	2.91	2.02	15.48	-6.04	6.48	4.03	4.03	15.48	0.00
k_t	4.35	2.47	-1.72	4.97	-12.48	5.46	3.43	-3.43	0.00	-12.48
Y_t	3.35	2.19	-0.50	5.84	-10.50	3.74	2.29	-2.29	0.00	-10.50
X^{ini}	2.31	1.44	-0.83	2.72	-6.42	2.86	1.93	-1.93	0.00	-6.42
K_{st}	2.23	1.25	0.86	6.83	-2.64	2.80	1.72	1.72	6.83	0.00
t_{lag}	2.03	1.31	0.93	5.25	-1.95	2.56	1.83	1.83	5.25	0.00
k_b	1.80	0.88	-0.65	1.34	-7.70	2.30	1.26	-1.26	0.00	-7.70
z_t	1.61	0.85	0.63	5.89	-1.30	2.06	1.22	1.22	5.89	0.00
S_t^{ini}	1.17	0.73	0.73	2.55	-0.07	1.48	1.00	1.00	2.55	-0.07
S_b^{ini}	1.10	0.65	0.62	2.36	-0.20	1.42	0.99	0.93	2.36	-0.20
Y_b	0.91	0.62	0.05	1.77	-2.49	0.82	0.49	-0.49	0.00	-2.49
b	0.66	0.43	-0.12	1.27	-1.88	0.44	0.26	0.26	1.27	0.00
K_{sb}	0.50	0.20	0.15	2.61	-0.45	0.64	0.29	0.29	2.61	0.00

[a]The δ measures are defined in equations (7)–(11).
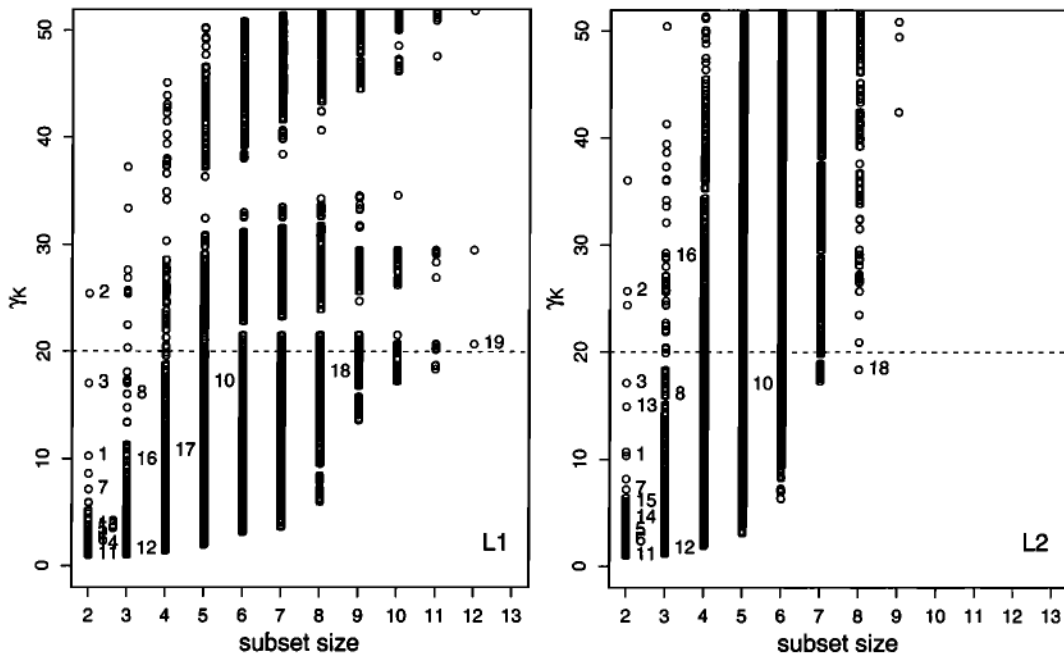
Figure 7. Collinearity indices for all parameter subsets for layout 1 (L1) and layout 2 (L2). The numbers in the plots refer to the parameter subsets defined in Table 7. The horizontal dashed lines indicate the critical γ in this particular application.

differences regarding subsets 13–17. For instance, subset 17 seems to be identifiable for layout 1 but not for layout 2. The nonidentifiability of subset 17 for layout 2 reflects the fact that layout 2 does not provide information about the biomass X. Lacking this information, the parameters of subset 17 cannot reasonably be estimated from a biological point of view.

Combining the information gained from Figure 7 and Table 7 leads to the conclusion that subset 19 (for layout 1) and subset 18 (for layout 2) are the subsets that combine maximal subset size with a collinearity index which still should allow parameter estimation without convergence problems. In order to demonstrate this, we performed weighted least squares parameter estimation according to (2) for subset 19 with full data

and for subset 18 with data according to layout 2. Weights were chosen as $w_i = 1/(SC_i)^2$ with scale factors SC_i as given in Table 4. Parameters that are not estimated for layout 2 are set to the estimates obtained for layout 1. Convergence was achieved in both cases only after quite many iterations. This clearly indicates that a collinearity index of approximatively 20 is already critical for parameter identifiability in this application.

Results of parameter estimation are given in Table 8 (subset 19 and layout 1) and Table 9 (subset 18 and layout 2). Both Table 8 and Table 9 show high standard errors for K_{sb}, z_t, and K_{st}, indicating that these three parameters are only poorly identifiable. This is, on the one hand, due to the high collinear-

Table 7. Collinearity Index for Specific Parameter Subsets for Layout 1 and Layout 2

Subset	Parameters	Layout 1 γ_K	Layout 2 γ_K
1	K_{st}, k_t	10.27	10.35
2	K_{st}, z_b	25.42	25.71
3	k_t, z_b	17.01	17.09
4	K_{st}, k_t, z_b	152.02	149.51
5	K_{sb}, k_b	3.32	3.34
6	K_{sb}, z_t	2.38	2.39
7	k_b, z_t	7.16	7.18
8	K_{sb}, k_b, z_t	16.02	16.00
9	K_{st}, k_t, z_b, K_{sb}, k_b, z_t	376.18	378.01
10	K_{st}, k_t, K_{sb}, k_b, z_t	17.11	16.93
11	S_b^{ini}, S_t^{ini}	1.04	1.02
12	S_b^{ini}, S_t^{ini}, t_{lag}	1.51	1.62
13	Y_b, Y_t	3.80	14.89
14	X^{ini}, Y_b	1.98	4.50
15	X^{ini}, Y_t	3.58	6.02
16	X^{ini}, Y_b, Y_t	9.85	28.94
17	b, X^{ini}, Y_b, Y_t	10.71	144.23
18	K_{st}, k_t, K_{sb}, k_b, z_t, S_b^{ini}, S_t^{ini}, t_{lag}	18.04	18.38
19	K_{st}, k_t, K_{sb}, k_b, z_t, S_b^{ini}, S_t^{ini}, t_{lag}, b, X^{ini}, Y_b, Y_t	20.72	2454.05

Table 8. Parameter Estimates, Absolute and Relative Standard Errors, and Approximate Correlation Matrix of Parameter Estimates for Parameter Subset 19 Estimated With Full Data

| Parameter | Estimate | Standard Error | | Correlation Matrix | | | | | | | | | | | |
		Absolute	Relative	b	k_b	K_{sb}	K_{st}	k_t	S_b^{ini}	S_t^{ini}	t_{lag}	X^{ini}	Y_b	Y_t	z_t
b	0.0367	0.00273	0.07	1.00											
k_b	0.229	0.0426	0.19	0.34	1.00										
K_{sb}	0.477	0.325	0.68	0.40	0.91[a]	1.00									
K_{st}	0.863	0.325	0.38	−0.27	−0.52	−0.61	1.00								
k_t	0.350	0.0384	0.11	−0.17	−0.31	−0.37	0.93[a]	1.00							
S_b^{ini}	4.72	0.0338	0.01	−0.08	0.15	0.19	−0.08	−0.16	1.00						
S_t^{ini}	5.11	0.0439	0.01	0.11	0.00	−0.07	0.06	−0.01	0.16	1.00					
t_{lag}	2.08	0.264	0.13	0.05	−0.31	−0.28	0.23	0.33	−0.16	−0.54	1.00				
X^{ini}	0.803	0.118	0.15	0.08	−0.11	−0.14	−0.36	−0.40	0.05	−0.21	0.61	1.00			
Y_b	0.639	0.127	0.20	0.51	0.51	0.48	−0.27	−0.05	−0.04	0.06	0.22	0.31	1.00		
Y_t	1.14	0.106	0.09	−0.26	−0.44	−0.36	0.26	0.07	−0.03	−0.06	−0.32	−0.51	−0.92[a]	1.00	
z_t	1.18	0.428	0.36	0.26	0.93[a]	0.77	−0.52	−0.39	0.03	0.12	−0.40	−0.05	0.34	−0.32	1.00

[a]Absolute off-diagonal elements in the correlation matrix exceeding a value of 0.9.

ity indices of subsets 1 and 8 as shown in Table 7. On the other hand, it also reflects the relatively low sensitivity of the model output to changes in K_{sb} (see Table 6). Furthermore, it is interesting to note that despite the small sensitivity of the model output to changes in b (see Table 6) the standard error for b in Table 8 is reasonably low. The reason for this is that there exists a time domain (after all substrate is consumed) where the model output exclusively depends on this parameter. This makes the sensitivity function of b nearly orthogonal to the other sensitivity functions, which is shown by low collinearity indices for all pairs of parameters containing b.

A closer look at the correlation matrix in Table 8 reveals that there are four off-diagonal elements exceeding a value of 0.9. To make clear which dependencies of parameter sensitivities are shown by these elements, we consider the collinearity indices in Table 7. Parameter subset $\{K_{st}, k_t\}$ (subset 1) has a collinearity index of approximately 10 in Table 7 and a correlation of 0.93 in Table 8. This indicates that the high value in the correlation matrix is mainly caused by a pairwise near-linear dependence of the sensitivities of $\{K_{st}, k_t\}$. Pairwise interpretation of the other off-diagonal elements exceeding 0.9, however, would be misleading. Considering the high elements of $\{K_{sb}, k_b\}$ and $\{k_b, z_t\}$ in Table 8, we see from Table 7 that the corresponding collinearity indices are only small (3.3) and moderate (7.2), respectively. On the other hand, we find a high collinearity index for the triple $\{K_{sb}, k_b, z_t\}$ (16.0). This makes clear that the high correlations of $\{K_{sb}, k_b\}$ and $\{k_b, z_t\}$ in Tables 8 and 9 show the dependence of the triple $\{K_{sb}, k_b, z_t\}$ and not the dependence of the pairs $\{K_{sb}, k_b\}$ and $\{k_b, z_t\}$. The situation is similar in the case of $\{Y_b, Y_t\}$. The collinearity index in Table 7 (3.8) clearly indicates that the pairwise interpretation is not adequate. The high absolute element of $\{Y_b, Y_t\}$ points to the strong dependence of the triple $\{X^{ini}, Y_b, Y_t\}$ (collinearity index of 10.7) instead.

Fixing some parameters and estimating the others leads to biased estimates under the assumption that the model is correctly specified, unless the fixed parameters are fixed at their (unknown) true values. Exploring potential bias problems is especially important if the model output is very sensitive to changes in the fixed parameter values and if the collinearity index increases strongly when a fixed parameter is added to the estimated parameter subset. As we know from Table 6, the model output is very sensitive to changes in the fixed parameter z_b. In addition, the collinearity index increases dramatically (from 21 to 380) if we add z_b to the estimated parameter subset. This leads to the expectation that setting z_b to another value would change at least some of the parameter estimates significantly.

In order to study the influence of the value of z_b on the estimates of the other parameters we performed additional parameter estimations of subset 19 for different values of z_b using full data. Relative changes of estimated parameters and WRSS values are shown in Figure 8. To keep Figure 8 simple, lines for parameter estimates which change less than 10% over the range considered for z_b were suppressed. From Figure 8 it can be seen that fixing z_b, for instance, at 0.2 increases the

Table 9. Parameter Estimates, Absolute and Relative Standard Errors, and Approximate Correlation Matrix of Parameter Estimates for Parameter Subset 18 Estimated With Data According to Layout 2

| Parameter | Estimate | Standard Error | | Correlation Matrix | | | | | | | |
		Absolute	Relative	k_b	K_{sb}	K_{st}	k_t	S_b^{ini}	S_t^{ini}	t_{lag}	z_t
k_b	0.242	0.0619	0.26	1.00							
K_{sb}	0.616	0.453	0.74	0.96[a]	1.00						
K_{st}	0.727	0.279	0.38	0.19	0.10	1.00					
k_t	0.337	0.0325	0.10	0.09	0.03	0.95[a]	1.00				
S_b^{ini}	4.72	0.0368	0.01	−0.22	−0.18	−0.31	−0.42	1.00			
S_t^{ini}	5.10	0.0435	0.01	0.04	0.03	−0.17	−0.32	0.10	1.00		
t_{lag}	2.07	0.218	0.11	−0.19	−0.18	0.57	0.78	−0.30	−0.62	1.00	
z_t	1.28	0.649	0.51	0.97[a]	0.90[a]	0.17	0.06	−0.27	0.14	−0.28	1.00

[a]Absolute off-diagonal elements in the correlation matrix exceeding a value of 0.9.
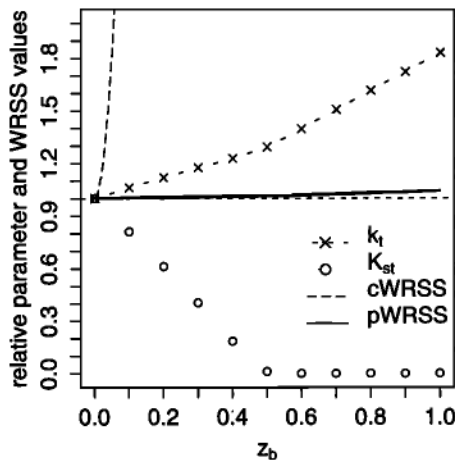
Figure 8. Relative changes of estimated parameters and WRSS values for different values of z_b. Here cWRSS denotes one-dimensional conditional likelihood (WRSS) values, and pWRSS denotes one-dimensional profile likelihood (WRSS) values.

estimate of k_t by approximately 12% and decreases the estimate of K_{st} by approximately 40%. The other parameters do not change significantly. At the same time the conditional WRSS increases 1200% (this is out of plot range), whereas the profile WRSS increases only 0.6%. The large change in the conditional WRSS reflects the high δ^{msqr} of z_b (see Table 6), and the flat profile WRSS reflects the high collinearity index of the full parameter set ($\gamma = 380$). The sharp increase in conditional WRSS can be compensated to a very large extent by appropriate changes in other parameters. In addition, the γ values in Table 7 show which parameter estimates are expected to change, compensating for changes in z_b. Parameter subset 9 shows a collinearity index of 376, which is nearly as high as the collinearity index of the full set. This means that we have to consider only the parameters belonging to subset 9 as candidates that might change significantly. The other parameter estimates are not expected to change because they are not suited to compensate for changes in z_b. A closer look at parameter subset 4 reveals that this triple already has a collinearity index of 152. Estimates of K_{sb} and k_t are therefore expected to change significantly when z_b is set to another value. This is exactly what we see in Figure 8. The conclusion that can be drawn is that K_{sb} and k_t have to be seen clearly as conditional estimates (conditional on $z_b = 0$) whereas the other parameter estimates are less dependent on the value of z_b.

6. Summary and Conclusions

As models in the environmental sciences continue to grow further, assessing practical parameter identifiability for large simulation models becomes increasingly important. In this paper we presented two diagnostic tools suitable to cope with the challenge posed by large models with many parameters and time-consuming numerical model evaluations. The sensitivity measures δ analyze the identifiability of parameters on an individual basis, and the collinearity indices γ analyze the compensation effects occurring in the case of joint estimation of parameter subsets. The use of these tools in the following four tasks was demonstrated: identifiability diagnosis for arbitrary parameter subsets, the selection of identifiable parameter sub-

sets for parameter estimation, the correct interpretation of the correlation matrix of parameter estimates with respect to parameter identifiability, and the exploration of potential bias problems due to fixing some parameters a priori. As the diagnostics δ and γ do not depend on the observations but only on the experimental layout, they are also suitable for experimental design purposes, a potential application that so far has only been sketched.

The principal conclusion of this paper is that it is worthwhile to calculate δ and γ indices routinely in the course of any parameter identification procedure of overparameterized models and to carry out analyses as suggested and summarized above. This is especially true for the so-called "state-of-the-art" models. Careful parameter identifiability analyses for such models with typical data sets and at typical "operating points" will give a deeper understanding of parameter identifiability problems due to overparameterization. In addition, comparisons of applications of the same model to different data sets will be easier if parameter identifiability is discussed in more detail, including, for example, potential bias problems. Finally, we see an important potential of the suggested tools in future applications to experimental design based on large, nonlinear simulation models.

References

Bates, D. M., and D. G. Watts, *Nonlinear Regression Analysis and its Applications*, John Wiley, New York, 1988.

Beck, M. B., Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, 23(8), 1393–1442, 1987.

Beck, M. B., Coping with ever larger problems, models, and data bases, *Water Sci. Technol.*, 39(4), 1–11, 1999.

Belsley, D. A., *Conditioning Diagnostics: Collinearity and Weak Data in Regression*, John Wiley, New York, 1991.

Beven, K., Towards a new paradigm in hydrology, in *Water for the Future: Hydrology in Perspective*, edited by J. C. Rodda and N. C. Matalas, *IAHS Publ.*, 164, 393–403, 1987.

Beven, K., Future of distributed modeling, *Hydrol. Processes*, 6(3), 253–254, 1992.

Beven, K., Prophecy, reality and uncertainty in distributed hydrological modeling, *Adv. Water Resour.*, 16(1), 41–51, 1993.

Campolongo, F., and A. Saltelli, Sensitivity analysis of an environmental model: An application of different analysis methods, *Reliab. Eng. Syst. Safety*, 57(1), 49–69, 1997.

Chatfield, C., Model uncertainty, data mining and statistical inference (with discussion), *J. R. Stat. Soc., Ser. A*, 158(3), 419–466, 1995.

Draper, D., Assessment and propagation of model uncertainty (with discussion), *J. R. Stat. Soc., Ser. B*, 57(1), 45–97, 1995.

Draper, D., A. Saltelli, S. Tarantola, and P. Prado, Scenario and parameteric sensitivity and uncertainty analyses in nuclear waste disposal risk assessment: The case of GESAMAC, in *Mathematical and Statistical Methods for Sensitivity Analysis*, edited by A. Saltelli, K. Chan, and M. Scott, chap. 13, pp. 275–292, John Wiley, New York, 2000.

Freer, J., K. Beven, and B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, 32(7), 2161–2173, 1996.

Grieb, T. M., R. J. M. Hudson, N. Shang, R. C. Spear, S. A. Gherini, and R. A. Goldstein, Examination of model uncertainty and param-

eter interaction in a global carbon cycling model (GLOCO), *Environ. Int.*, 25(6-7), 787–803, 1999.

Gupta, V. K., and S. Sorooshian, Uniqueness and observability of conceptual rainfall runoff model parameters: The percolation process examined, *Water Resour. Res.*, 19(1), 269–276, 1983.

Helton, J. C., Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliab. Eng. Syst. Safety*, 42(2-3), 327–367, 1993.

Holmberg, A., On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities, *Math. Biosci.*, 62, 23–43, 1982.

Hornberger, G. M., and R. C. Spear, An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, 12, 7–18, 1981.

Hornberger, G. M., and R. C. Spear, An approach to the analysis of behavior and sensitivity in environmental systems, in *Uncertainty and Forecasting of Water Quality*, edited by M. Beck and G. van Straten, pp. 101–116, Springer-Verlag, New York, 1983.

Jakeman, A. J., and G. M. Hornberger, How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29(8), 2637–2649, 1993.

Jakob, A., J. Zobrist, J. S. Davies, P. Liechti, and L. Sigg, NADUF: Langzeitbeobachtung des chemisch-physikalischen Gewässerzustandes, *Gas Wasser Abwasser*, 74, 171–186, 1994.

Kleissen, F. M., M. B. Beck, and H. S. Wheater, The identifiability of conceptual hydrochemical models, *Water Resour. Res.*, 26(12), 2979–2992, 1990.

Kuczera, G., Assessing hydrologic model nonlinearity using response surface plots, *J. Hydrol.*, 118(1-4), 143–161, 1990.

Kuczera, G., and M. Mroczkowski, Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, 34(6), 1481–1489, 1998.

Pastres, R., D. Franco, G. Pecenik, C. Solidoro, and C. Dejak, Local sensitivity analysis of a distributed parameters water quality model, *Reliab. Eng. Syst. Safety*, 57(1), 21–30, 1997.

Reichert, P., Aquasim: A tool for simulation and data-analysis of aquatic systems, *Water Sci. Technol.*, 30(2), 21–30, 1994.

Reichert, P., AQUASIM 2.0: User manual, technical report, Swiss Fed. Inst. for Environ. Sci. and Technol., Dübendorf, Switzerland, 1998.

Reichert, P., and M. Omlin, On the usefulness of overparameterized ecological models, *Ecol. Modell.*, 95(2-3), 289–299, 1997.

Reichert, P., R. von Schulthess, and D. Wild, The use of AQUASIM for estimating parameters of activated sludge models, *Water Sci. Technol.*, 31(2), 135–147, 1995.

Restrepo, P. J., and R. L. Bras, A view of maximum-likelihood estimation with large conceptual hydrologic models, *Appl. Math. Comput.*, 17(5048), 375–403, 1985.

Saltelli, A., Sensitivity analysis: Could better methods be used?, *J. Geophys. Res.*, 104(D3), 3789–3793, 1999.

Saltelli, A., and M. Scott, Guest editorial: The role of sensitivity analysis in the corroboration of models and its link to model structural and parametric uncertainty, *Reliab. Eng. Syst. Safety*, 57(1), 1–4, 1997.

Saltelli, A., S. Tarantola, and K. P. S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, 41(1), 39–56, 1999.

Sobol, I. M., Sensitivity estimates for nonlinear mathematical models, *Math. Modell. Comput. Exp.*, 1(4), 407–414, 1993.

Sommer, H. M., Variability in microbiological degradation experiments: Analysis and case study, Ph.D. thesis, Inst. of Math. Modell., Tech. Univ. of Den., Lyngby, 1997.

Sorooshian, S., and V. K. Gupta, Automatic calibration of conceptual rainfall runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, 19(1), 260–268, 1983.

Spear, R. C., Large simulation models: Calibration uniqueness and goodness of fit, *Environ. Modell. Software*, 12(2-3), 219–228, 1997.

Spear, R. C., T. M. Grieb, and N. Shang, Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, 30(11), 3159–3169, 1994.

Stewart, G. W., Collinearity and least squares regression (with discussion), *Stat. Sci.*, 2, 68–100, 1987.

Stigter, J. D., and M. B. Beck, A new approach to the identification of model structure, *Environmetrics*, 5, 315–333, 1994.

Turányi, T., Sensitivity analysis of complex kinetic systems: Tools and applications, *J. Math. Chem.*, 5(3), 203–248, 1990.

Turányi, T., Applications of sensitivity analysis to combustion chemistry, *Reliab. Eng. Syst. Safety*, 57(1), 41–48, 1997.

Uehlinger, U., C. König, and P. Reichert, Variability of photosynthesis-irradiance curves and ecosystem respiration in a small river, *Freshwater Biol.*, 44(3), 493–507, 2000.

van Straten, G., Analytical methods for parameter space delimitation and application to shallow lake phytoplankton dynamics modeling, *Appl. Math. Comput.*, 17(4), 459–482, 1985.

Weisberg, S., *Applied Linear Regression*, 2nd ed., John Wiley, New York, 1990.

R. Brun and H. R. Künsch, Seminar für Statistik, ETH Zürich, ETH Zentrum, CH-8092 Zürich, Switzerland. (brun@stat.math.ethz.ch; kuensch@stat.math.ethz.ch)

P. Reichert, EAWAG, Postfach 611, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland. (peter.reichert@eawag.ch)