ISSN 1360–1725

THE UNIVERSITY of MANCHESTER

UMIST

Sensitivity Analysis of Parameters in Modelling With Delay-Differential Equations

Christopher T.H. Baker & Fathalla A. Rihan

Numerical Analysis Report No. 349

September 1999

Manchester Centre for Computational Mathematics
Numerical Analysis Reports

DEPARTMENTS OF MATHEMATICS

Reports available from:
Department of Mathematics
University of Manchester
Manchester M13 9PL
England

And over the World-Wide Web from URLs:
`http://www.ma.man.ac.uk/MCCM/MCCM.html`
`ftp://ftp.ma.man.ac.uk/pub/narep`

Sensitivity Analysis of Parameters in Modelling With Delay-Differential Equations

Christopher T.H. Baker * and Fathalla A. Rihan*

September 7, 1999; released on the WWW on October 8th 1999

Abstract

Many problems in bioscience for which observations are reported in the literature can be modelled by suitable functional differential equations incorporating a delay, parameterized by parameters p_1,p_2,...,p_L. Given such observations (which usually contain error or 'noise'), we may determine the parameters by optimizing a measure of best fit. It is often desirable to have information about "sensitivity" aspects of the problem. For example, the user may wish to estimate the effect of perturbing the parameters on the solution. In data-fitting, it may be important to know the effect of small changes of the data on the parameter estimates. In addition, one might wish to determine the effect of nonlinearity of the model solutions.

Our aim in this paper is to produce a new method to estimate (i) the sensitivity of the state variables to the parameter estimates $\{p_i\}$, (ii) the sensitivity of the parameter estimates to the observations and (iii) the nonlinearity effects for delay differential models. The sensitivity of the parameter estimate to the observation is *low* if the sensitivity of the state variable to the parameter estimate is *high*. Sensitivity coefficients are used to determine the covariance matrix of parameter estimates and hence to determine the standard deviations. Numerical results, based on growth of *E.coli* colonies, are used to illustrate the results.

A revised version of this Technical Report will be submitted for publication.

Keywords Sensitivity analysis, parameter estimates, neutral delay differential equation, time-lag, nonlinearity effect.

1 Introduction

Delay differential equations (DDEs) are increasingly used in numerous application areas that include population dynamics (taking into account the gestation and the maturation time), infectious diseases (accounting for the incubation periods), physiological and pharmaceutical kinetics (modelling, for example, hematopoiesis and respiration, where the delays are due, respectively, to cell maturation and blood transport between the lung and brain, etc.), chemical and enzyme kinetics (such as mixing reactants), biological immune response (in which the antibody production by the T-cell population depends on the antigenic stimulation at an earlier time), the navigational control of ships and aircraft (with, respectively, large and short lags), and more general control problems. We refer to [4, 5, 6, 9, 18] for more examples in biomathematics. The object of a sensitivity analysis is to determine systematically the effect of uncertain parameters on system solutions and the effect of the noisy data on the certainty to which parameters may be determined; see also [14, 16].

*Mathematics Department, The Victoria University of Manchester, Manchester M13 9PL, ENGLAND. The second-named author acknowledges support from the Egyptian Government.

Consider the system of delay differential equations, parameterized by $\mathbf{p} \in \mathbb{R}^L$:

$$\begin{aligned} \mathbf{y}'(t,\mathbf{p}) &= \mathbf{f}(t,\mathbf{y}(t),\mathbf{y}(t-\tau),\mathbf{p}), \quad t \geq 0, \\ \mathbf{y}(t,\mathbf{p}) &= \psi(t,\mathbf{p}), \quad t \leq 0. \end{aligned} \tag{1.1}$$

In (1.1), the dependent variable is $\mathbf{y}(t,\mathbf{p}) = (y_1(t,\mathbf{p}), y_2(t,\mathbf{p}), \ldots, y_M(t,\mathbf{p}))^{\mathrm{T}}$ and the parameters are given by $\mathbf{p} = (p_1, p_2, \ldots, p_L)^{\mathrm{T}} \in \mathbb{R}^L$; τ is the *time-lag*, to be identified as a parameter. Our problem, given data $\{t_j; \mathfrak{y}_j\}_{j=1}^N$ (incorporating measurements $\mathfrak{y}_j$ of the solution $\mathbf{y}(t,\mathbf{p})$ at time t_j; where $N \geq L$), is to find the parameter $\mathbf{p} = \widehat{\mathbf{p}}$ for which the function $\mathbf{y}(t;\mathbf{p})$ provides a 'best' fit, at arguments $t = t_j$, to the given set $\{\mathfrak{y}_j\}_{j=1}^N$. (If the model is correct, $\mathfrak{y}_j$ represents an observed value of $\mathbf{y}(t_j,\mathbf{p})$.) When determining the best fit we suppose that the unknown parameter $\widehat{\mathbf{p}}$ is estimated by minimizing the objective function [7]:

$$\Phi(\widehat{\mathbf{p}}) \leq \Phi(\mathbf{p}) := \sum_{j=1}^{N} \left[\mathbf{y}(t_j,\mathbf{p}) - \mathfrak{y}_j\right]^2. \tag{1.2}$$

Other objective functions are sometimes employed, but will not be considered in this paper. We refer to the above as the *least squares* (LS) approach, and to $\widehat{\mathbf{p}}$ as the *LS* estimator.

Determining $\widehat{\mathbf{p}}$ relies, in general, upon numerical techniques, and these are essential to the approach presented in this paper. In particular, we rely upon robust and accurate solvers of systems of (neutral) delay differential equations. It was explained in [2] how the properties of the solutions of delay (and neutral delay) equations can result in poor continuity of Φ; see Section 4.

1.1 Sensitivity issues and nonlinear bias in the model

Of considerable importance in assessing the model (1.1) is the sensitivity of the model solution $\mathbf{y}(t,\mathbf{p})$ to changes in the parameter $\mathbf{p}$ or the sensitivity of the best fit $\widehat{\mathbf{p}}$ to changes in the data $\{t_j; \mathfrak{y}_j\}_{j=1}^N$. A knowledge of how the solution can vary with respect to small change in the data or the parameters can yield insights into the model behaviour and can assist the modelling process. For example, (*i*) if it can be seen that a particular parameter p_j has no effect on the solution, it may be possible to eliminate it from the modelling process; (*ii*) it might be found that a particular parameter $p^{\#}$ is affected by a small change in data and another not affected. In view of the preceding remarks, we desire to compute *the sensitivity of the state variable* $\mathbf{y}(t,\mathbf{p})$ *to the parameter estimates* $\mathbf{p}$ and to estimate *the sensitivity of the parameter estimate* $\mathbf{p}$ *to the observations* $\mathfrak{y}_j$.

In general, the parameter estimation problem for (1.1) using (1.2) is an example of nonlinear regression. Nonlinear regression models differ from linear regression models in that, given the usual assumption of an independent and identically distributed normal stochastic term, linear models give rise to unbiased, normally distributed, minimum variance estimators, whereas nonlinear regression models have these properties only asymptotically (when the sample size becomes very large). Thus it is also desired to estimate *the nonlinearity effect* in the models. Let us amplify this further.

When the predictions are governed by models using ordinary differential or delay-differential equations as models, then the *LS* approach generally leads to a nonlinear minimization problem. Difficulties may arise from the fact that nonlinear regression models differ in general from the linear regression models in that the *LS* parameter estimates can be *biased, non-normally distributed*, and have *variance* exceeding the minimum possible variance [26, p.13]. These characters differ from model to model, so it is necessary to estimate the nonlinearity effect for every model we use. The percentage *bias* in the parameter estimates is a good indicator to the quantitative effect of nonlinearity; see [8, 26]. We shall consider this further in Section 3.

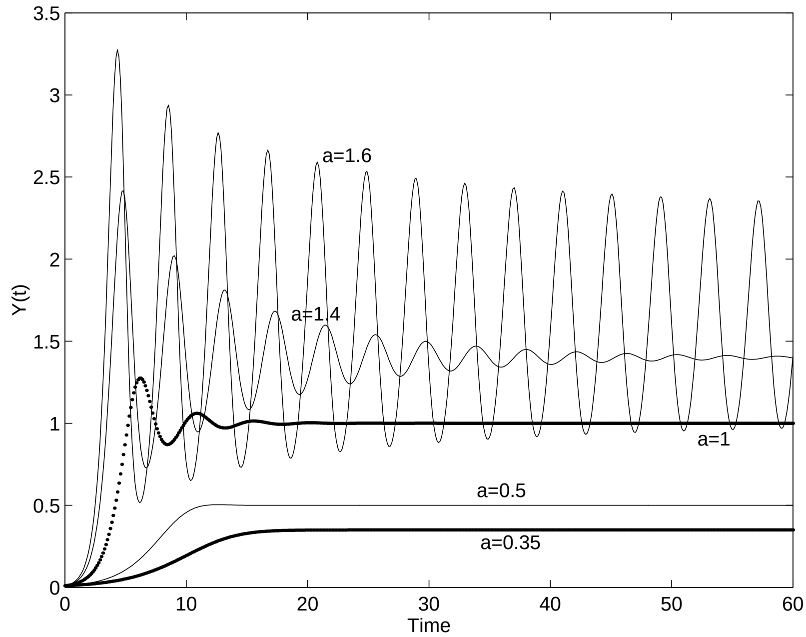

Figure 1.1: Intersection of solutions to the different DDEs can cause non-unique best fit in certain data.

1.2 Uniqueness of best fit

For a given set $\{t_j\}_{j=1}^{N}$ and an arbitrary function f in (1.1), there is no reason to suppose that there exists a *unique* minimizer $\widehat{\mathbf{p}}$ of $\Phi(\mathbf{p})$. Indeed, it is easy to find examples for non-unique best fit models; one requires only to find solutions for two different parameters that agree at the points $t_1, t_2, \ldots, t_N$. In FIGURE 1.1 we give an example of such a scenario; plotting the graphs of solutions corresponding to small initial functions for the equation

$$y'(t) = y(t)\Big[a - y(t-1)\Big], \quad t > 0, \tag{1.3}$$

where $\mathbf{p} = [a]$, and $1 \le a \le 1.6$ (say) demonstrates that solutions for different parameters may pass through a common set of values. If the data correspond to the points of intersection, $\mathbf{p}$ is not uniquely determined.

The question of what happens as $N \to \infty$ under varying assumptions is of theoretical interest but could only be answered with precise assumptions on $\{t_i\}$ and f. Even when there exists a unique $\widehat{\mathbf{p}}$, the success of iterative methods for determining its value may depend upon a sufficiently close starting approximation. Graphical displays of $\Phi(\mathbf{p})$ for a particular model (see Section 5, FIGURES 5.3 & 5.4) provide some insight.

2 A Special Case: *A Linear Neutral-Delay Differential Equation*

We consider here a linear scalar neutral-delay differential model of the form:

$$y'(t,\mathbf{p}) \quad = \quad \rho_0 y(t,\mathbf{p}) + \rho_1 y(t-\tau,\mathbf{p}) + \rho_2 y'(t-\tau,\mathbf{p}), \quad t \ge 0, \tag{2.1}$$

where the right-hand side depends upon the parameters $\{\rho_0, \rho_1, \rho_2, \tau\}$. Accompanying this equation is the initial condition

$$y(t,\mathbf{p}) \quad = \quad \psi(t,\mathbf{p}), \quad t \le 0,$$

and in our numerical examples $\psi(t, \mathbf{p})$ will depend upon a parameter ρ_3 and on τ. To cover the complete set of parameters, we take $\mathbf{p} = [\rho_0, \rho_1, \rho_2, \rho_3, \tau]^{\mathrm{T}}$ as the parameter vector.

The model will allow us to demonstrate the salient features of our discussion, and we address the modifications required in the case of a more general model (1.1) in Section 6 towards the end of the paper. We first examine the sensitivity of $y(t, \mathbf{p})$ to $\mathbf{p}$.

2.1 Sensitivity of $y(t, \mathbf{p})$ to the parameter $\mathbf{p}$

In the case of a scalar $y(t, \mathbf{p})$, the partial derivatives $\dfrac{\partial y}{\partial p_i}$ measure the local sensitivity of the solution with respect to changes in the parameters p_i. The sensitivity coefficients that we wish to compute are the functions $s_i(t, \mathbf{p}) \equiv \left\{\frac{\partial}{\partial p_i}\right\} y(t, \mathbf{p})$ such that

$$y(t, \mathbf{p} + \delta\mathbf{p}) - \left\{y(t, \mathbf{p}) + \sum_i \delta p_i s_i(t, \mathbf{p})\right\} \quad \text{is} \quad \mathcal{O}(||\delta\mathbf{p}||^2) \quad \text{as} \quad ||\delta\mathbf{p}|| \to 0 \tag{2.2}$$

(or is at least $o(1)||\delta\mathbf{p}||$, in the case of reduced differentiability). The sensitivity coefficients $s_i(t, \mathbf{p})$ are evaluated at the optimal parameters $\mathbf{p} := \widehat{\mathbf{p}} = [\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\tau}]^{\mathrm{T}}$. We therefore seek the components of *the first order sensitivity coefficients*, evaluated at $(t, \widehat{\mathbf{p}})$, of the vector

$$\left[\frac{\partial}{\partial \rho_0} y, \frac{\partial}{\partial \rho_1} y, \frac{\partial}{\partial \rho_2} y, \frac{\partial}{\partial \rho_3} y, \frac{\partial}{\partial \tau} y\right]^{\mathrm{T}} = \left[s_1(t, \mathbf{p}), s_2(t, \mathbf{p}), s_3(t, \mathbf{p}), s_4(t, \mathbf{p}), s_5(t, \mathbf{p})\right]^{\mathrm{T}}.$$

2.1.1 Computing the sensitivity coefficients

The sensitivity coefficients of the model (2.1) can be computed by solving a system of neutral-delay differential equations (NDDEs). This system comes from differentiating the model (2.1) with respect to the parameters $\{\rho_0, \rho_1, \rho_2, \rho_3, \tau\}$. We introduce the variable $z(t, \mathbf{p}) = y'(t, \mathbf{p})$ and we repeat equation (2.1) for y to give us (in the present model) a system of equations which can be expressed as

$$\begin{aligned} \mathbf{u}'(t, \mathbf{p}) &= \mathbf{A}\mathbf{u}(t, \mathbf{p}) + \mathbf{B}\mathbf{u}(t - \tau, \mathbf{p}) + \mathbf{C}\mathbf{u}'(t - \tau, \mathbf{p}), \quad t > 0, \\ \mathbf{u}(t, \mathbf{p}) &= \Psi(t, \mathbf{p}), \quad t \le 0, \end{aligned} \tag{2.3}$$

where

$$\mathbf{A} = \begin{bmatrix} \rho_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \rho_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \rho_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \rho_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho_1 & \rho_1 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \rho_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \rho_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho_2 & \rho_2 \end{bmatrix}, \quad \mathbf{u}(t, \mathbf{p}) = \begin{bmatrix} y(t, \mathbf{p}) \\ s_1(t, \mathbf{p}) \\ s_2(t, \mathbf{p}) \\ s_3(t, \mathbf{p}) \\ s_4(t, \mathbf{p}) \\ z(t, \mathbf{p}) \\ s_5(t, \mathbf{p}) \end{bmatrix}; \quad \Psi(t, \mathbf{p}) = \begin{bmatrix} \psi(t, \mathbf{p}) \\ \frac{\partial}{\partial \rho_0}\psi(t, \mathbf{p}) \\ \frac{\partial}{\partial \rho_1}\psi(t, \mathbf{p}) \\ \frac{\partial}{\partial \rho_2}\psi(t, \mathbf{p}) \\ \frac{\partial}{\partial \rho_3}\psi(t, \mathbf{p}) \\ \psi'(t, \mathbf{p}) \\ \frac{\partial}{\partial \tau}\psi(t, \mathbf{p}) \end{bmatrix}.$$

(and some terms $\frac{\partial}{\partial p_i}\psi(t, \mathbf{p})$ are non-vanishing in the case that the initial function ψ depends non-trivially upon ρ_0, ρ_1, ρ_2, ρ_3 or τ).

System (2.3) may be solved by software for NDDEs, discussed in Section 4.

2.1.2 Second order sensitivity coefficients

In some applications, the sensitivity of the parameter estimates $\mathbf{p}$ to the observed data $\mathfrak{y}_j$ might be characterized by considering the *second order sensitivity coefficients*, $r_{ij}(t, \mathbf{p})$

$$r_{ij}(t, \mathbf{p}) := \frac{\partial}{\partial p_j} s_i(t, \mathbf{p}) = \frac{\partial^2}{\partial p_j \partial p_i} y(t, \mathbf{p}), \quad (i, j = 1, \ldots, L_p)$$

again evaluated at $(t, \widehat{\mathbf{p}})$.

These values, r_{ij} (the sensitivity coefficients of second order), measure the sensitivity of the coefficients $s_i(t, \mathbf{p})$ to perturbations in the parameters:

$$s_i(t, \mathbf{p} + \delta\mathbf{p}) = s_i(t, \mathbf{p}) + \sum_j r_{ij}(t, \mathbf{p})\delta p_j \ + \ o(1)||\delta\mathbf{p}||.$$

They are therefore of some intrinsic interest in assessing the reliability of the values s_i.

The coefficients r_{ij} can again be derived via the system of neutral delay differential equations (2.3). For a 5-parameter system, there are 25 coefficients and each requires a delay differential equation as part of a coupled system of delay differential equations that also includes equations for y and for $z := y'$. However, the number of equations could be reduced by assuming sufficient differentiability, $r_{ij}(t, \mathbf{p}) = r_{ji}(t, \mathbf{p})$. For the present model, we give the appropriate equations in the appendix.

2.2 Sensitivity of the optimum parameter $\widehat{\mathbf{p}}$ to perturbations in the data

To compute $\frac{\partial \widehat{\mathbf{p}}}{\partial \mathfrak{y}_j}$, the sensitivity of the parameter estimate $\widehat{\mathbf{p}}$ to the observed data $\mathfrak{y}_j$, assume that the objective function

$$\Phi(\mathbf{p}) \equiv \Phi(\mathbf{p}, \mathfrak{y}) := \sum_i \Big[y(t_i, \mathbf{p}) - \mathfrak{y}_i\Big]^2 \tag{2.4}$$

is smooth as a function of $\mathbf{p}$ in the neighbourhood of the optimal parameter $\widehat{\mathbf{p}}$. Then we have

$$\frac{\partial}{\partial p_k} \Phi(\mathbf{p}, \mathfrak{y}) = 2 \sum_i \Big[y(t_i, \mathbf{p}) - \mathfrak{y}_i\Big] \frac{\partial y(t_i, \mathbf{p})}{\partial p_k}, \tag{2.5}$$

$$\frac{\partial^2}{\partial p_l \partial p_k} \Phi(\mathbf{p}, \mathfrak{y}) = 2 \sum_i \frac{\partial y(t_i, \mathbf{p})}{\partial p_l} \frac{\partial y(t_i, \mathbf{p})}{\partial p_k} + 2 \sum_i \Big[y(t_i, \mathbf{p}) - \mathfrak{y}_i\Big] \frac{\partial^2 y(t_i, \mathbf{p})}{\partial p_l \partial p_k}. \tag{2.6}$$

To minimize the objective function (2.4), the right hand side of equation (2.5) vanishes at $\mathbf{p} = \widehat{\mathbf{p}}$ (where $\widehat{\mathbf{p}} \equiv \widehat{\mathbf{p}}(\mathfrak{y})$), so

$$\sum_i [y(t_i, \widehat{\mathbf{p}}(\mathfrak{y})) - \mathfrak{y}_i] s_k(t_i, \widehat{\mathbf{p}}(\mathfrak{y})) \quad = \quad 0. \tag{2.7}$$

Now, the left hand side of equation (2.7) is a function of $\widehat{\mathbf{p}}(\mathfrak{y})$ and $\mathfrak{y}$; differentiating both sides with respect to $\mathfrak{y}_j$ yields, for $k = 1, \ldots, L_p$,

$$\sum_{i=1}^{N} \sum_{l=1}^{L_p} \Big[s_k(t_i, \widehat{\mathbf{p}}) s_l(t_i, \widehat{\mathbf{p}}) + [y(t_i, \widehat{\mathbf{p}}) - \mathfrak{y}_i]\, r_{lk}(t_i, \widehat{\mathbf{p}})\Big] \frac{\partial \hat{p}_l}{\partial \mathfrak{y}_j} = s_k(t_j, \widehat{\mathbf{p}}). \tag{2.8}$$

If we assume that $y(t_i, \widehat{\mathbf{p}})$ is close to the observed value $\mathfrak{y}_i$, so that the term $[y(t_i, \widehat{\mathbf{p}}) - \mathfrak{y}_i]$ in the left hand side of equation (2.8) can be neglected, then the above system can be approximated by

$$\sum_{i=1}^{N} \sum_{l=1}^{L_p} s_k(t_i, \widehat{\mathbf{p}}) s_l(t_i, \widehat{\mathbf{p}}) \frac{\partial \hat{p}_l}{\partial \mathfrak{y}_j} \approx s_k(t_j, \widehat{\mathbf{p}}), \qquad k = 1, \ldots, L_p,$$

or

$$\sum_{i=1}^{N} s_k(t_i,\widehat{\mathbf{p}})\Big(\sum_{l=1}^{L_p} s_l(t_i,\widehat{\mathbf{p}})\frac{\partial \hat{p}_l}{\partial \mathfrak{y}_j}\Big) \approx s_k(t_j,\widehat{\mathbf{p}}), \qquad k=1,\ldots,L_p. \tag{2.9}$$

This equation can be written in a compact form

$$\Big[\sum_{i=1}^{N}\mathbf{s}(t_i,\widehat{\mathbf{p}})\mathbf{s}^{\mathrm{T}}(t_i,\widehat{\mathbf{p}})\Big]\frac{\partial \widehat{\mathbf{p}}}{\partial \mathfrak{y}_j} \approx \mathbf{s}(t_j,\widehat{\mathbf{p}}). \tag{2.10}$$

Then the sensitivity of the best fit parameter estimate $\widehat{\mathbf{p}}$ to observations $\mathfrak{y}_j (j=1,2,\ldots,N)$ can be estimated by

$$\frac{\partial \widehat{\mathbf{p}}}{\partial \mathfrak{y}_j} \approx \Big[\mathfrak{B}(\widehat{\mathbf{p}})\Big]^{-1}\mathbf{s}(t_j,\widehat{\mathbf{p}}), \tag{2.11}$$

where $\mathbf{s}$ is $L_p \times 1$ vector, given by (2.3), and $\mathfrak{B}(\widehat{\mathbf{p}}) := \Big[\sum_{i=1}^{N}\mathbf{s}(t_i,\widehat{\mathbf{p}})\mathbf{s}^T(t_i,\widehat{\mathbf{p}})\Big]$ is an $L_p \times L_p$ nonsingular matrix.

A desirable property of the model is that the sensitivity of the parameter estimate to the observation, $\frac{\partial \widehat{\mathbf{p}}}{\partial \mathfrak{y}_j}$, should be small in order to minimize the effects of observation noise on the parameter estimate. Equation (2.11) suggests that increasing $\mathbf{s}(t,\widehat{\mathbf{p}})$ (the sensitivity of the state variable with respect to the unknown parameter) decreases the sensitivity of the parameter estimate to observation.

2.3 Standard Deviation of Parameter Estimates

We can use the sensitivity coefficients $(s_i,\ i=1,\ldots,L)$ to determine the covariance matrix $[\varsigma_{ij}]$, of the estimates, as follows:

$$\begin{bmatrix} \varsigma_{11} & \varsigma_{12} & \cdots & \varsigma_{1L} \\ \varsigma_{21} & \varsigma_{22} & \cdots & \varsigma_{2L} \\ \varsigma_{31} & \varsigma_{32} & \cdots & \varsigma_{3L} \\ .. & .. & \cdots & .. \\ \varsigma_{R1} & \varsigma_{R2} & \cdots & \varsigma_{LL} \end{bmatrix} = 2\frac{\Phi(\widehat{\mathbf{p}})}{N-L}\left[H(\widehat{\mathbf{p}})\right]^{-1},$$

where $(N-L)$ is the number of degree of freedom and $H(\widehat{\mathbf{p}})$ is the Hessian matrix of the objective function $\Phi(\widehat{\mathbf{p}})$. Using the notation $\frac{\partial}{\partial \mathbf{p}}$ and $\frac{\partial}{\partial \mathbf{p}^T}$ in §§ 7.1, the Hessian matrix can be written in the form

$$H(\widehat{\mathbf{p}}) = \left[\frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}^{\mathrm{T}}}\Phi(\widehat{\mathbf{p}})\right].$$

This matrix can be approximated, in terms of (2.6) and using the sensitivity coefficients, as:

$$H(\widehat{\mathbf{p}}) \approx \widetilde{H}(\widehat{\mathbf{p}}) := 2\left[\sum_{k=1}^{N} s_i(\xi_k,\widehat{\mathbf{p}})s_j(\xi_k,\widehat{\mathbf{p}})\right]_{i,j=1,\ldots,L}.$$

The standard deviations for the parameter estimates are the quantities $\sigma_i \equiv \sigma(\hat{p}_i) = \sqrt{\varsigma_{ii}}$ $(i = 1,\ldots,L)$.

3 Indications of Bias

We remarked earlier that percentage *bias* in the parameter estimates is a good indicator of the quantitative effect of nonlinearity [26]. To examine the *biases* of the parameter estimates we proceed as follows:

- Perturb the solution, of the model, corresponding to the best-fit parameters $\widehat{\mathbf{p}}$ with normally distributed random errors of zero mean and variance (see [7]),

$$\sigma^2 = \frac{\Phi(\widehat{\mathbf{p}})}{N-L}.$$

- Find new best-fit parameters $\widetilde{\mathbf{p}}$ to the perturbed data.

- Repeat this process, perhaps, 500 or preferably 1000 times, to generate a statistically significant estimate of the mean value of $\widetilde{\mathbf{p}}$.

- If the *relative biases* satisfy the relation,

$$|\widehat{\mathbf{p}} - mean\{\widetilde{\mathbf{p}}\}| < 0.01|\widehat{\mathbf{p}}|,$$

  then the *LS* estimates are not significantly biased and the effect of non-linearity is not significant and the experimenter can have confidence in the parameter estimates, and their standard deviations (see TABLE 5.2).

In other words, if the *LS* estimator of parameters of a nonlinear model is only *slightly biased* with a distribution of which close to that of a normal distribution and with a variance only slightly in excess of the minimum variance bound, it seems reasonable to consider the estimator as behaving *close to a linear*. If, on the other hand, the *LS* estimator is *badly biased*, with distribution far from normal and variance greatly in excess of the minimum variance bound, the nonlinear model might be far from the linear model in behaviour. For more details about the nonlinearity effects in parameter estimations, we may refer to [1], [7], [8], [12] and [26].

4 Solvers of DDEs and Parameter Estimation for DDE Models

There is now an extensive body of expertise in the mathematical community concerned with the numerical solution of DDEs: Fourth-order Runge-Kutta methods and two point Hermite interpolation polynomials have been used by Neves [20] and Neves and Thompson [21]. Algorithms based on fourth- and seventh-order Runge-Kutta-Fehlberg methods together with Hermite interpolation polynomials were presented by Oberle and Pesch [22]. Thompson [28] has developed numerical methods which are based on a continuously embedded Runge-Kutta method of Sarafyan [29]. Our numerical work has been based upon the use of an explicit Runge-Kutta method. This method is based on the Dormand & Prince fifth-order Runge-Kutta method for ODEs [15] due to Shampine [27] and fifth-order Hermite interpolant [22].

There are currently a number of general purpose codes for solving initial value problems for DDEs. An important feature of such codes is that they aim to produce a solution to within a given accuracy for a wide range of requested tolerances. Paul [24] has developed such a code that based on the feature of [23] and [25]. The resulting code is uniformly fifth-order accurate for ODEs, DDEs and neutral differential equations (NDDEs). Another code based on continuously embedded sixth-order Runge-Kutta methods for the solution of functional differential equations has been proposed by Corwin *et al.* [13].

The task of parameter estimation is one of minimizing a suitable objective function $\Phi(\mathbf{p})$, for example one given by (1.2), based on the unknown parameters and observed data. In the case of parameter estimation for DDEs, this can include not only estimating parameters appearing in the DDEs but also estimating the position of the initial point, the initial function and the delayed arguments.

For example, consider the problem of estimating the parameters $\rho_0,\ \rho_1,\ \rho_2,\ \rho_3$ and τ in the model (2.1). The optimum parameter $\widehat{\mathbf{p}} \equiv [\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\tau}]^{\mathrm{T}}$ is taken to be the value such that

$$\Phi(\widehat{\mathbf{p}}) \leq \Phi(\mathbf{p})$$

for all physically meaningful values of $\mathbf{p}$ and $\widehat{\mathbf{p}}$.

Given a set of experimental data, $\{\mathfrak{y}_j\}_{j=1}^N$, the technique for finding the best-fit parameter values for a given mathematical model and objective function involves solving the model equations using the current values of the parameters in order to compute $\Phi(\mathbf{p})$. The parameter values are then adjusted (by the minimization routine, for example `E04UPF` in[1] the `NAG` library, `LMDIF` from[2] `NETLIB` and `FMINS` in `MATLAB`) so as to reduce the value of the objective function. However, in order to find the *global* best-fit parameter values, the initial estimate of the parameter values should be sufficiently close to the global minimum. Thus, good starting estimates for the parameter values can be of great assistance, both in speeding up the minimization process and finding the global minimum.

4.1 Some problems with parameter estimation in DDEs

One obvious difficulty with such procedures (from both the practical and the theoretical viewpoint) is that solutions of DDEs are not, in general differentiable, with respect to variation of the delay. In addition, discontinuities can arise in the solution of a DDE. Such discontinuities, when they arise from the initial point $t_0(\mathbf{p})$ and the initial function $\psi(t, \mathbf{p})$, may propagate into $\Phi(\mathbf{p})$ via the solution $y(t, \mathbf{p})$ if it has a jump at one of the data points $\{\zeta_i\}$; see [3]. Therefore parameter estimation in DDEs mainly depends on:

- differentiability of the solution $y(t; \mathbf{p})$ with respect to the parameter $\mathbf{p}$,
- the existence and uniqueness of the solution $y(t, \mathbf{p})$ that depends on the initial function ψ and the parameter $\mathbf{p}$,
- existence and position of the jump discontinuity points,
- the statistical nature of the observed data-points $\{\zeta_i, \mathfrak{y}_i\}_{i=1}^N$.

From the fact that:

$$\left(\frac{\partial \Phi(\zeta_i; \mathbf{p})}{\partial p_l}\right)_{\pm} = 2\sum_{i=1}^{N} [y(\zeta_i; \mathbf{p}) - \mathfrak{y}_i] \left(\frac{\partial y(\zeta_i; \mathbf{p})}{\partial p_l}\right)_{\pm}, \tag{4.1}$$

$$\begin{aligned}\left(\frac{\partial^2 \Phi(\zeta_i; \mathbf{p})}{\partial p_l \partial p_m}\right)_{\pm\pm} = 2\sum_{i=1}^{N} \Bigg[& \left(\frac{\partial y(\zeta_i; \mathbf{p})}{\partial p_l}\right)_{\pm} \left(\frac{\partial y(\zeta_i; \mathbf{p})}{\partial p_m}\right)_{\pm} \\ & + [y(\zeta_i; \mathbf{p}) - \mathfrak{y}_i] \left(\frac{\partial^2 y(\zeta_i; \mathbf{p})}{\partial p_l \partial p_m}\right)_{\pm\pm}\Bigg].\end{aligned} \tag{4.2}$$

It is clear from equations (4.1) and (4.2) that, unless $\mathfrak{y}_i = y(\zeta_i; \mathbf{p})$, jumps can arise in the first or the second partial derivative of $\Phi(\mathbf{p})$, with respect to p_l, if the first or the second partial derivatives of $y(t, \mathbf{p})$ with respect to p_l has a jump at $t = \zeta_i$ (one of the data-points). These jumps can propagate into the second derivative of $\Phi(\mathbf{p})$ if the first derivative of $y(t; \mathbf{p})$, with respect to p_l, has a jump at

[1] E04UPF is designed to minimize an arbitrary smooth sum of squares function subject to constraints (which may include simple bounds on the variables, linear constraints and smooth nonlinear constraints) using a sequential quadratic programming (SQP) method.

[2] LMDIF is an unconstrained minimization routine based on the Levenberg-Marquardt algorithm.

one of the data-points $t = \zeta_i$ even when $\mathfrak{y}_i = y(\zeta_i; \mathbf{p})$. For more discussion about these issues we refer to [2].

The connection between jumps in the derivatives of $y(t;\mathbf{p})$ with respect to t and the partial derivatives of $y(t;\mathbf{p})$ with respect to some p_l can also be seen in the sensitivity coefficient system (2.3). It is clear that the jumps in this system are intimately related to the jumps in the derivatives $y(t,\mathbf{p})$ with respect to t in the delay differential system. Such jumps can spread forward along the integration interval. Location of these jumps is determined by $t-\tau$, additional discussion has been given in [3].

5 Numerical Example: Growth of *Escherichia Coli (E.coli)* Colonies

A synchronous culture of *E.coli* that exhibits prolonged step-like growth [19, Fig.4] (experimental data given in TABLE 5.1) can be modelled by a hierarchy of models (2.1) (where we include successively more parameters), with the initial function $\psi(t)$ given by a distribution curve [5]:

$$y(t) := \psi(t) = \frac{2\times 2.25}{\tau_{cell}}(N\rho_3/\rho_1)E(\frac{2t}{\tau_{cell}}+1), \quad t \le 0, \tag{5.1}$$

where

$$E(t) = \begin{cases} \exp(\frac{-1}{1-t^2}) & for\ |t| \le 1, \\ 0 & for\ |t| \ge 1. \end{cases}$$

We can then perform the *sensitivity analysis* for the hierarchy in order to estimate the effect of adding new components. The cases that we consider are:

(a) fit $\tau_{culture} = \frac{\ln(2)}{\rho_0}$ (exponential growth, with initial value N) in the model: $y'(t) = \rho_0 y(t)$,

(b) fit $\tau := \tau_{cell} = \frac{1}{\rho_1}$, *with* $\rho_0 = \rho_2 = 0$, *and* $\rho_3 = 1$ (time-lag growth) in the model: $y'(t) = \rho_1 y(t-\tau)$,

(c) fit $\tau := \tau_{cell}$, ρ_0, ρ_1, *with* $\rho_2 = 0$, *and* $\rho_3 = 1$ in the model: $y'(t) = \rho_0 y(t) + \rho_1 y(t-\tau)$,

(d) fit $\tau := \tau_{cell}$, ρ_0, ρ_1, ρ_2, *with* $\rho_3 = 1$ in the model: $y'(t) = \rho_0 y(t) + \rho_1 y(t-\tau) + \rho_2 y'(t-\tau)$,

(e) fit $\tau := \tau_{cell}$, ρ_0, ρ_1, ρ_2, *and* ρ_3, with $\psi(t)$ given by (5.1).
The corresponding graphs are shown in Figure 5.2.

The biological meaning of the parameters is as follows:

$\tau > 0$	:	the average cell-division time;
$-\rho_0 \ge 0$	:	the rate of cell-death in the culture;
$\rho_1 \ge 0$	:	the rate of commitment to the cell-division process;
$0 \le \rho_2 \le 2$	:	the gradual dispersion of synchronization of cell-division;
		($\rho_2 = 2$ implies that synchronization is presented);
$0 \le \rho_3 \le 1$	:	the fraction of cells dividing over the first step.

The term 'synchronous' refers to the fact that the cells in the culture are homogeneous and synchronized [11]. Thus features of the model are: (*i*) All the cells have the same division time. (*ii*) All the cells divide simultaneously. (*iii*) There is prolonged initial step-like growth, as shown in FIGURE 5.1. (*iv*) The initial number, N, of *E.coli* colonies is unknown which can be specified as a parameter to be estimated or can be estimated by backwards continuation of the data (in this case $N = 99$). The parameter ρ_2 has a natural interpretation, so the neutral delay term is attractive for qualitative reasons. Procedures of the above sections are applied to the concerned model (2.1).

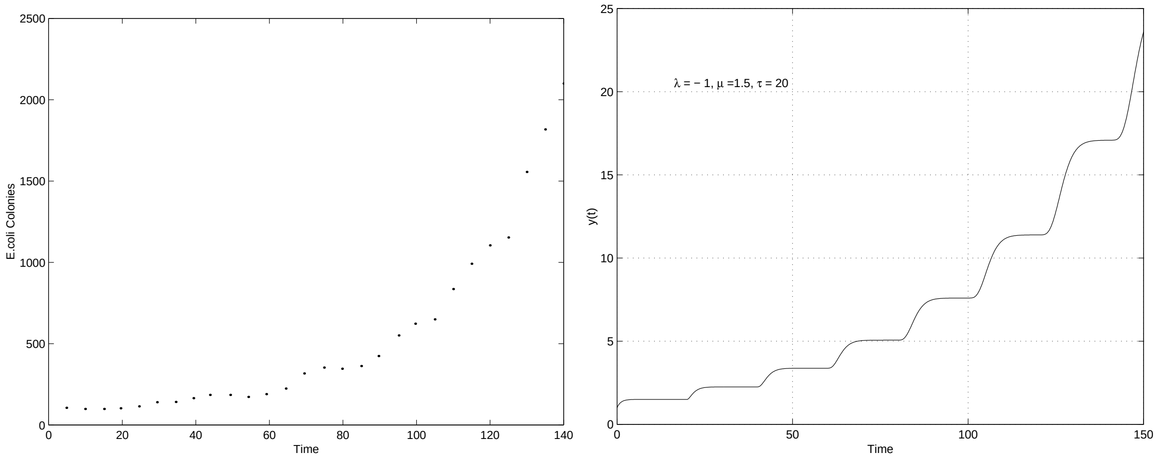

Figure 5.1: Data for synchronous *E.coli* growth and a graph of the solution of the DDE: $y'(t) = \lambda y(t) + \mu y(t-\tau)$.

Numerical values of parameter estimates, standard deviations, nonlinear bias, and sensitivity of the state variable to the parameter estimates and the parameter estimates to the observations are given in TABLES 2,3, for the observed data given in TABLE 5.1. Local uniqueness of the best-fit for the *least square* approach, and the qualitative effect of nonlinearity are shown through the FIGURES 5.3 & 5.4.

We now turn to an interpretation of our numerical results:

The results obtained show that there is a considerable "qualitative improvement" (indicated by a graph of the best-fit solution and the experimental data; see FIGURE 2) and "quantitative improvement" (indicated by the size of $\|Err\|_2$ and the standard deviations of the best-fit parameter values; see TABLES 2) when we add more parameters in our model. The fact that $\rho_2 < 2$ indicates that the initially synchronized cell population becomes desynchronized over time. We note also that *neutral* DDEs provide better qualitative and quantitative consistency with the step-like growth patterns than ODEs or DDEs with constant time-lag. In addition, using delay differential model (2.1) in cell growth, gives direct estimates of some relevant growth parameters of synchronous cultures such as: the cell-doubling time, the fraction of cells that are dividing, the rate of commitment of cells to cell division, the degree of synchronization of cells in the population, and the death rate of cells. Whereas using ODE model, in cell growth, only provides an indirect estimate of the culture-doubling time $\tau_{culture} = \ln(2)/\rho_0$.

In TABLE 5.2, we note that the parameters are slightly biased, so that the model-data combination is not badly nonlinear. In FIGURE 5.3, the closeness of the graph of $\Phi(\mathbf{p})$ to a parabola indicates the small degree of nonlinearity of the model-data set combination in one parameter time-lag model. FIGURE 5.4 shows the pairwise plots of the parameters. The regularity behaviour of the contours (such as ellipses) indicates whether the model-data combinations are not badly nonlinear in five parameters time-lag model. In addition, the contours, in these Figures, indicates the degree of the closeness of those models to a linear regression behaviour.

The sensitivity coefficients of the state variable to the parameter estimates should be large enough and the sensitivity of the parameter estimates to the observation should be small in order to minimize the effect of the observation noise on the parameter estimates. Results are obtained in TABLE 5.3 which show that the sensitivity of the solution to the parameter estimates is high and

Observed data of E.coli colonie growth										
Time(mins)	4.85	9.96	15.1	19.6	24.6	29.5	34.6	39.4	43.9	49.2
Cells per ml	10.6	99	99	103	115	140	142	165	185	185
Time(mins)	54.3	59.2	64.5	69.5	74.9	79.8	85.0	89.7	95.2	99.7
Cells per ml	173	190	224	317	353	346	363	424	551	623
Time(mins)	105.0	110.0	115.0	120.0	125.0	130.0	135.0	140.0		
Cells per ml	650	836	992	1105	1153	1556	1818	2100		

Table 5.1: Observed data of E.coli colonie growth [19, Fig.4].

Parameter(s), STD, Errors and NLB, for E.coli growth models						
Model	ρ_0	ρ_1	ρ_2	ρ_3	τ	$\|Err\|_2$
1 param ODE	-	-	-	-	33.5863	724.86
$\sigma(.)$	-	-	-	-	0.3768	
$NLB(.)$	-	-	-	-	0.0028%	
1 param DDE	-	-	-	-	25.0616	571.80
$\sigma(.)$	-	-	-	-	0. 0.2993	
$NLB(.)$	-	-	-	-	0. 0.0136%	
3 param DDE	-0.1156	0.30886	-	-	27.7165	258.82
$\sigma(.)$	0.0293	0.0595	-	-	0.3438	
$NLB(.)$	0.9426%	0.9939%	-	-	0.0314%	
4 param NDDE	-0.0257	0.0504	1.6847	-	20.2719	160.16
$\sigma(.)$	0.0038	0.0082	0.0467	-	0.0868	
$NLB(.)$	0.7022%	0.3304%	0.0338%	-	0.1043%	
5 param NDDE	-0.0057	0.0131	1.8407	0.1600	20.2229	129.88
$\sigma(.)$	0.0051	0.0101	0.0783	0.1107	0.0486	
$NLB(.)$	0.0708%	0.0506%	0.0062%	0.0384%	0.0420%	

Table 5.2: Parameter estimates , STD, Errors and their nonlinear biases (NLB) for E.coli growth models.

the sensitivity of the parameter estimates to the observations is low. This reflects that the model is correct and the parameters used have significant effect.

6 Generalizations of the Model

We here indicate how the mathematics for the special model (2.1) is modified to cover the general non-linear case (1.1). We introduce some notation in the Appendix used to define the sensitivity coefficients.

6.1 Sensitivity of state variables to parameter estimates

The relevant sensitivity quantities are, respectively, the matrices

$$\left\{\frac{\partial}{\partial \mathfrak{y}_j}\right\}^{\mathrm{T}} \mathbf{p} := \left[\frac{\partial p_i}{\partial \mathfrak{y}_j}\right]_{i=1,\cdots,L} \in \mathbb{R}^{L\times M}. \tag{6.1}$$

Sensitivity coefficients of E.coli colonies										
par.	ρ_0		ρ_1		ρ_2		ρ_3		τ	
	$\partial y/\partial\rho_0$	$\partial\rho_0/\partial\eta_j$	$\partial y/\partial\rho_1$	$\partial\rho_1/\partial\eta_j$	$\partial y/\partial\rho_2$	$\partial\rho_2/\partial\eta_j$	$\partial y/\partial\rho_3$	$\partial\rho_3/\partial\eta_j$	$\partial y/\partial\tau$	$\partial\tau/\partial\eta_j$
Time	One param. ODE model									
34.6	-	-	-	-	-	-	-	-	8	0.0000
69.5	-	-	-	-	-	-	-	-	7	0.0004
105	-	-	-	-	-	-	-	-	20	0.0015
140	-	-	-	-	-	-	-	-	56	0.0043
Time	One param. DDE model									
34.6	-	-	-	-	-	-	-	-	16	0.0000
69.5	-	-	-	-	-	-	-	-	18	0.0001
105	-	-	-	-	-	-	-	-	75	0.0003
140	-	-	-	-	-	-	-	-	233	0.0011
Time	Three params. DDE model									
34.6	1099	0.0000	383	0.0000	-	-	-	-	11	0.0014
69.5	5062	0.0000	1863	0.0000	-	-	-	-	10	0.0018
105	19379	0.0000	7369	0.0000	-	-	-	-	30	0.0024
140	65723	0.0000	25557	0.0002	-	-	-	-	53	0.0041
Time	Five params. Neutral DDE model									
34.6	7140	0.0000	1640	0.0000	10	0.0005	280	0.0002	10	0.0002
69.5	40190	0.0000	15960	0.0000	110	0.0003	870	0.0001	20	0.0002
105	166780	0.0000	75000	0.0000	640	0.0005	2920	0.0000	80	0.0001
140	617650	0.0000	289710	0.0000	3910	0.0006	10380	0.0003	40	0.0017

Table 5.3: Absolute values of sensitivity coefficients of the state variables to the parameter estimates and the parameter estimates to the observations for E.coli colonies growth models.

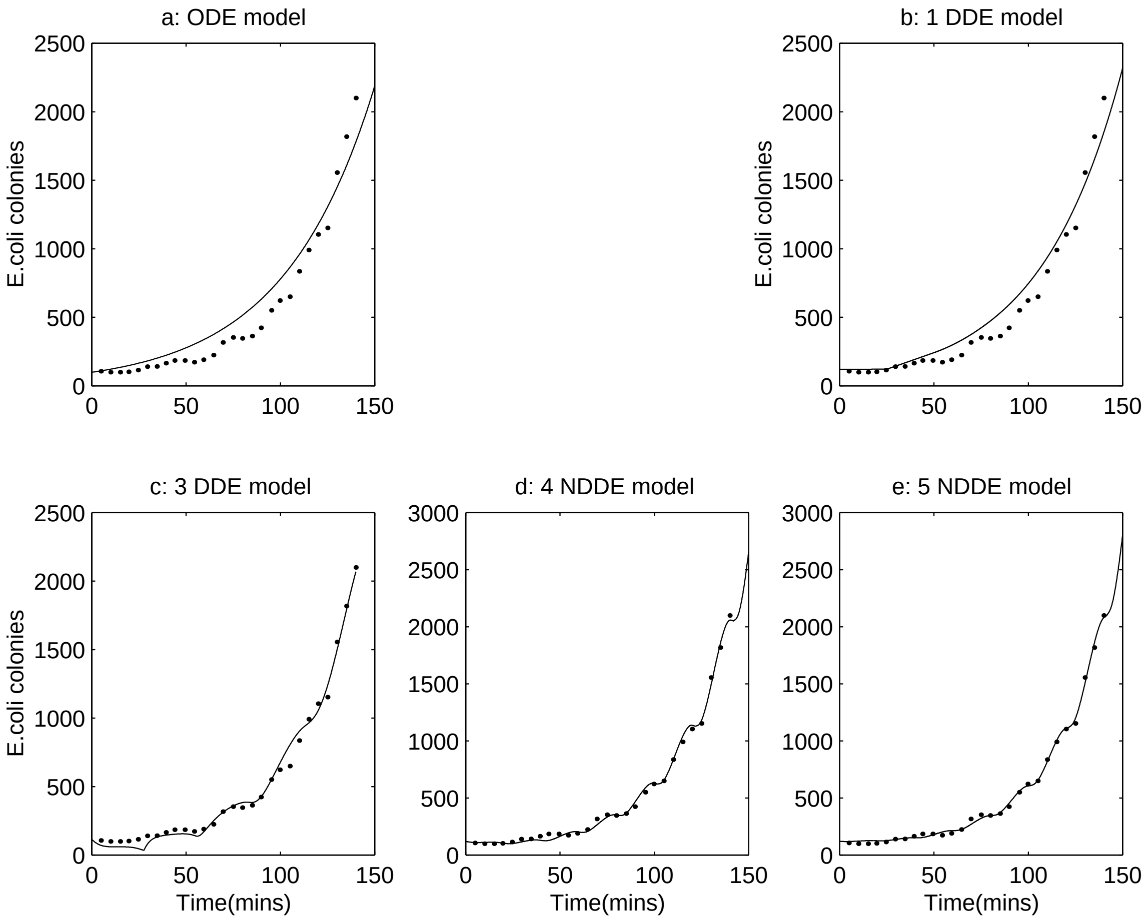

Figure 5.2: A synchronous culture of $E.ecoli$ K12λ F^{-} cells was prepared by loading 2×10^{10} cells from an exponential culture into a 15ml tube. The cells were then centrifuged at 2500g for 20 minutes and the top 2% of cells suspended in fresh growth medium. The graphs represent: (a) the exponential growth model of one parameter $\tau_{culture} = \frac{\ln(2)}{\rho_0}$, (b) the time-lag model with one parameter $\tau_{cell} = \frac{1}{\rho_1}$, (c) the DDE model with three parameters, $\rho_0,\ \rho_1$ and τ_{cell}, (d) the NDDE model with four parameters, $\rho_0,\ \rho_1$, $\tau_{cell},\ \rho_2$ and (e) the NDDE model with five parameters, $\rho_0,\ \rho_1,\ \tau_{cell},\ \rho_2,\ \rho_3$.

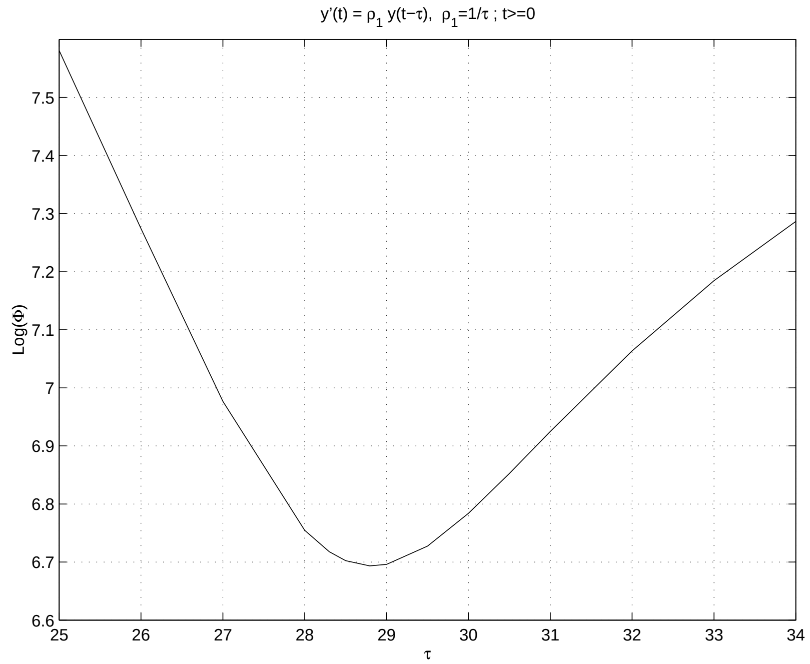

Figure 5.3: Local uniqueness of the best fit and the depending of Φ on τ (for one parameter time-lag model).

which we shall discuss in the next section, and, as we shall discuss now, the matrix of *sensitivity coefficients*

$$\mathbf{S}(t,\mathbf{p}) \equiv \left\{\frac{\partial}{\partial \mathbf{p}}\right\}^{\mathrm{T}} \mathbf{y}(t,\mathbf{p}) := \left[\frac{\partial y_i(t,\mathbf{p})}{\partial p_j}\right]_{\substack{i=1,\cdots,M\\ j=1,\cdots,L}} \in \mathbb{R}^{M\times L}. \tag{6.2}$$

Assuming appropriate differentiability of $\mathbf{y}(t,\mathbf{p})$ with respect to $\mathbf{p}$,

$$\mathbf{y}(t,\mathbf{p}+\delta\mathbf{p}) = \mathbf{y}(t,\mathbf{p}) + \sum_{j=1}^{L} \frac{\partial \mathbf{y}(t,\mathbf{p})}{\partial p_j}\delta p_j + O(\|\delta\mathbf{p}\|^2), \quad \text{or}$$

$$\mathbf{y}(t,\mathbf{p}+\delta\mathbf{p}) = \mathbf{y}(t,\mathbf{p}) + \mathbf{S}(t,\mathbf{p})\delta\mathbf{p} + O(\|\delta\mathbf{p}\|^2).$$

Thus the $M\times L$ matrix $\mathbf{S}(t,\mathbf{p})$ may be regarded as the local sensitivity [17] of the solution $\mathbf{y}(t,\mathbf{p})$ to small changes in $\mathbf{p}$.

Applying $\dfrac{\partial}{\partial \mathbf{p}}$ to (1.1) yields, in the case τ is not one of the parameters $\{p_\ell\}$, the variational equation is

$$\mathbf{S}'(t) = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t,\mathbf{y}(t),\mathbf{y}(t-\tau);\mathbf{p})\right]\mathbf{S}(t) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{y}_\tau}(t,\mathbf{y}(t),\mathbf{y}(t-\tau);\mathbf{p})\right]\mathbf{S}(t-\tau) + \mathbf{B}(t).$$

Alternatively, we write

$$\mathbf{S}'(t) \quad = \quad \mathbf{J}(t)\mathbf{S}(t) + \mathbf{J}_\tau(t)\mathbf{S}(t-\tau) + \mathbf{B}(t), \quad t \geq 0, \tag{6.3}$$

where:
$\mathbf{S}(t)$ is the $M\times L$ sensitivity coefficient matrix ($\mathbf{S}_{ij} \equiv \frac{\partial y_i}{\partial p_j}$), $J(t)$ is the $M\times M$ Jacobian matrix ($\mathbf{J}_{ij} \equiv \frac{\partial f_i}{\partial y_j}$), $\mathbf{B}(t)$ is an $M\times L$ matrix of partial derivatives ($\mathbf{B}_{ij} \equiv \frac{\partial f_i}{\partial p_j}$). Note that the i-th column in $\mathbf{S}(t)$, $\mathbf{s}_i(t) \equiv \mathbf{s}_i(t,\mathbf{p}) := \left[\frac{\partial y_1(t,\mathbf{p})}{\partial p_i}, \frac{\partial y_2(t,\mathbf{p})}{\partial p_i}, \ldots, \frac{\partial y_M(t,\mathbf{p})}{\partial p_i}\right]^{\mathrm{T}}$, is the sensitivity solution vector for the model parameter p_i.

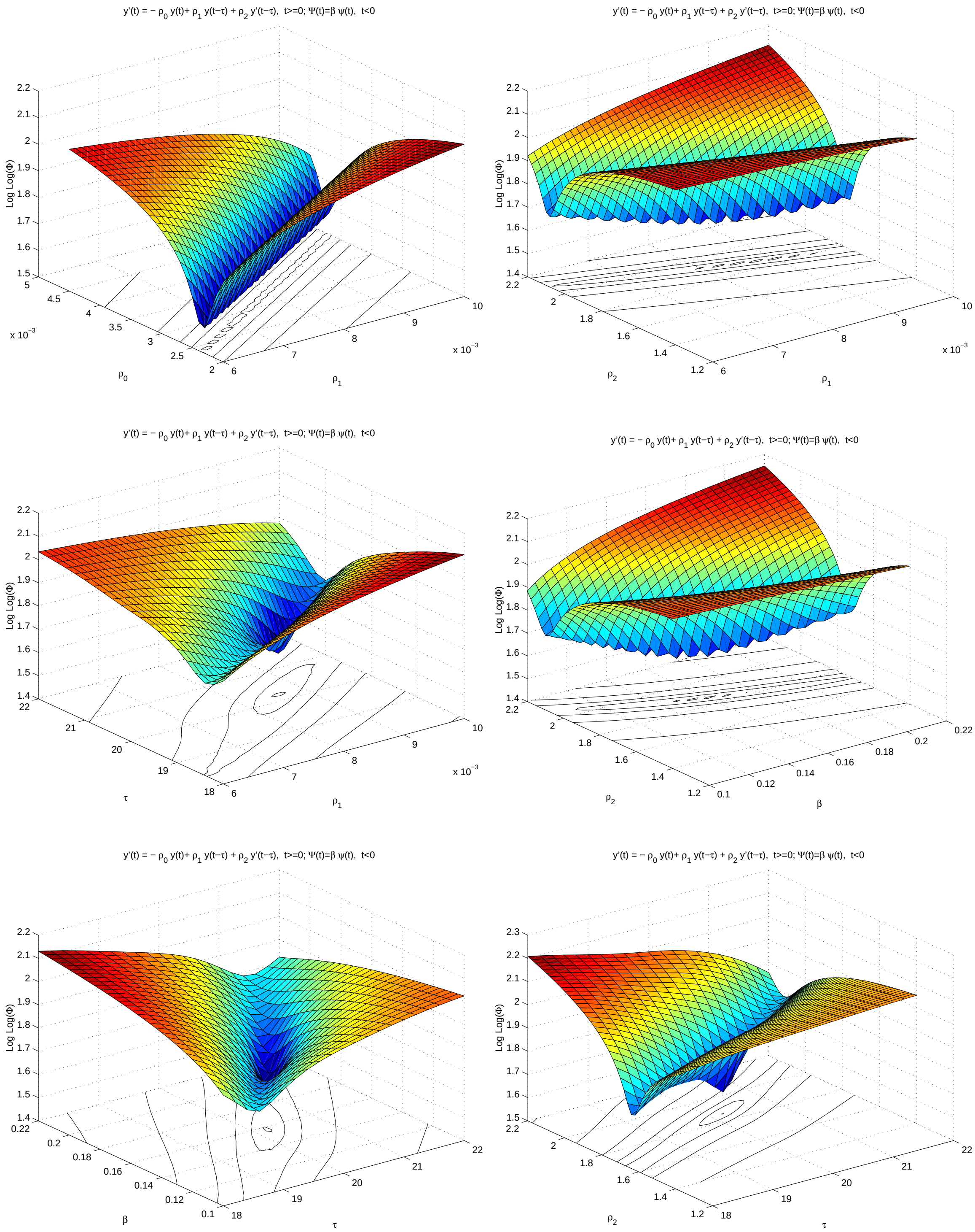

Figure 5.4: Local uniqueness of the best fit and the depending of $\Phi(\mathbf{p})$ on the pairwise parameter estimates. For each graph, contours indicate the correlation of the parameter with each other and the inference region of least square estimate. The closeness of the contour to the ellipse, the small degree of nonlinearity of the model to the data.

In the case that τ is a component, p_j, of $\mathbf{p}$, the matrix $\mathbf{B}(t)$ in (6.3) should be changed so that,

$$\mathbf{B}_{ij}(t) \equiv \frac{\partial f_i}{\partial p_j} := \frac{\partial f_i}{\partial y(\sigma_j(t))}\frac{dy(\sigma_j(t))}{d\sigma_j(t)}\frac{d\sigma_j(t)}{dp_j} = -\frac{\partial f_i}{\partial y(\sigma_j(t))}y'(t-p_j),$$

where $\sigma_j(t) = t - p_j$, and the other components are as before.

The *second order sensitivity coefficients* give information on the sensitivity of $\mathbf{S}(t,\mathbf{p})$ to $\mathbf{p}$ and are defined by the $(LM) \times L$ matrix, $\mathbf{R}(t) \equiv \frac{\partial}{\partial \mathbf{p}}\mathbf{S}(t)$, whose ith row is

$$\mathbf{r}_i(t) \equiv \mathbf{r}_i(t,\mathbf{p}) := \left[\frac{\partial \mathbf{s}_1(t,\mathbf{p})}{\partial p_i}, \frac{\partial \mathbf{s}_2(t,\mathbf{p})}{\partial p_i}, \ldots, \frac{\partial \mathbf{s}_L(t,\mathbf{p})}{\partial p_i}\right].$$

A further differentiation, applying $\dfrac{\partial}{\partial \mathbf{p}}$ to (6.3) gives, in terms of *Kronecker products*, the result

$$\begin{aligned}\mathbf{R}'(t) \quad &= \quad \mathbf{A}(t)\mathbf{S}(t) + \Big(\mathbf{I}_L \otimes \mathbf{J}(t)\Big)\mathbf{R}(t) + \mathbf{A}_\tau(t)\mathbf{S}(t-\tau) + \\ &\quad \Big(\mathbf{I}_L \otimes \mathbf{J}_\tau(t)\Big)\mathbf{R}(t-\tau) + \mathbf{K}(t), \quad t \geq 0. \qquad (6.4)\end{aligned}$$

In (6.4), $\mathbf{R}(t)$ is an $(LM) \times L$ sensitivity coefficient matrix of second order ($\mathbf{R}_{ij} \equiv \frac{\partial \mathbf{s}_i}{\partial p_j}$), $\mathbf{A}(t)$ is an $(LM) \times M$ partial derivatives of the Jacobian matrix ($\mathbf{A}_{ij} \equiv \frac{\partial \mathbf{J}_i(t)}{\partial p_j},\ where,\ \mathbf{J}_i = \frac{\partial f_k}{\partial y_i}$), and $\mathbf{K}(t)$ is an $(LM) \times L$ partial derivative matrix ($\mathbf{K}_{ij} \equiv \frac{\partial \mathbf{b}_i}{\partial p_j},\ where\ \mathbf{b}_i = \frac{\partial f_k}{\partial p_i}$).

To find the sensitivity coefficient matrices, $\mathbf{S}$ and $\mathbf{R}$, we need to solve the (neutral) delay differential systems (6.3) and (6.4) simultaneously with the system (1.1), with respectively, associated initial functions

$$\mathbf{S}(t,\mathbf{p}) = \frac{\partial}{\partial \mathbf{p}}\psi(t,\mathbf{p}), \text{ and } \mathbf{R}(t,\mathbf{p}) = \frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}^{\mathrm{T}}}\psi(t,\mathbf{p}) \quad \text{for } t \leq 0. \qquad (6.5)$$

It should be noted here that jumps can propagate into systems (6.3) and (6.4) if the derivative of $y(t,\mathbf{p})$ has a jump and the time lag term τ is considered as as a parameter to be estimated; see [2].

6.2 Sensitivity of parameter estimates to observations

To compute the sensitivity of $\widehat{\mathbf{p}}$ to $\mathfrak{y}_j$, the objective function $\Phi(\mathbf{p})$ is minimized by finding where its derivatives vanishes:

$$\frac{\partial}{\partial \mathbf{p}}\Phi(\widehat{\mathbf{p}}) = 2\sum_{j=1}^{N}\mathbf{S}^{\mathrm{T}}(t_j,\widehat{\mathbf{p}})\left[\mathbf{y}(t_j,\widehat{\mathbf{p}}) - \mathfrak{y}_j\right] = 0, \qquad (6.6)$$

where $\mathbf{S}(t,\widehat{\mathbf{p}})$ is defined via equation (6.3).

For simplicity, denote

$$\Phi_p := \frac{\partial}{\partial \mathbf{p}}\Phi, \quad \Phi_{pp} := \frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}^{\mathrm{T}}}\Phi.$$

Now Φ_p is a function of the parameter estimates $\widehat{\mathbf{p}}$ and the observations $\mathfrak{y}_j$, so that

$$\Phi_p(\widehat{\mathbf{p}}, \mathfrak{y}_1, \mathfrak{y}_2, \ldots, \mathfrak{y}_N) = 0, \qquad (6.7)$$

and $\widehat{\mathbf{p}} = \widehat{\mathbf{p}}(\mathfrak{y}_1, \mathfrak{y}_2, \ldots, \mathfrak{y}_N)$.

Differentiating equation (6.7) with respect to $\mathfrak{y}_j$ yields,

$$\Phi_{pp}\frac{\partial \widehat{\mathbf{p}}}{\partial \mathfrak{y}_j} + \Phi_{\mathfrak{y}_j p} = 0, \qquad (6.8)$$

where, from (6.6),

$$\Phi_{pp} = 2\sum_{j=1}^{N}\mathbf{S}^{\mathrm{T}}(t_j,\widehat{\mathbf{p}})\mathbf{S}(t_j,\widehat{\mathbf{p}}) + 2\sum_{j=1}^{N}\mathbf{R}^{\mathrm{T}}(t_j,\widehat{\mathbf{p}})\Big(\mathbf{I}_L \otimes \big[\mathbf{y}(t_j,\widehat{\mathbf{p}}) - \mathfrak{y}_j\big]\Big), \tag{6.9}$$

$$\Phi_{\mathfrak{y}_j p} = -2\mathbf{S}(t_j,\widehat{\mathbf{p}}), \tag{6.10}$$

and $\mathbf{R} := \dfrac{\partial^2}{\partial\mathbf{p}\partial\mathbf{p}^{\mathrm{T}}}\mathbf{y}(t,\widehat{\mathbf{p}})$ is defined by (6.4).
Solving equation (6.8) for the $\partial\widehat{\mathbf{p}}/\partial\mathfrak{y}_j$, we have

$$\begin{aligned}\frac{\partial\widehat{\mathbf{p}}}{\partial\mathfrak{y}_j} &= -\Big[\Phi_{pp}\Big]^{-1}\Phi_{\mathfrak{y}_j p}\\ &= \Big[\sum_{j=1}^{N}\mathbf{S}^{\mathrm{T}}(t_j,\widehat{\mathbf{p}})\mathbf{S}(t_j,\widehat{\mathbf{p}}) + \sum_{j=1}^{N}\mathbf{R}^{\mathrm{T}}(t_j,\widehat{\mathbf{p}})\Big(\mathbf{I}_L \otimes \big[\mathbf{y}(t_j,\widehat{\mathbf{p}}) - \mathfrak{y}_j\big]\Big)\Big]^{-1}\mathbf{S}(t_j,\widehat{\mathbf{p}}) \quad .(6.11)\end{aligned}$$

Assume that $\mathbf{y}(t_i,\widehat{\mathbf{p}})$ is close to the observed value $\mathfrak{y}_i$, so that the second term in the right hand side of equation (6.11) can be neglected. Then

$$\frac{\partial\widehat{\mathbf{p}}}{\partial\mathfrak{y}_j} \approx \Big[\sum_{i=1}^{N}\mathbf{S}^{\mathrm{T}}(t_i,\widehat{\mathbf{p}})\mathbf{S}(t_i,\widehat{\mathbf{p}})\Big]^{-1}\mathbf{S}(t_j,\widehat{\mathbf{p}}). \tag{6.12}$$

Equation (6.12) shows that increasing $\mathbf{S}(t,\widehat{\mathbf{p}}) \equiv \frac{\partial\mathbf{y}(t,\widehat{\mathbf{p}})}{\partial\mathbf{p}}$ (the sensitivity of the state variable with respect to the unknown parameter) decreases the sensitivity, $\frac{\partial\widehat{\mathbf{p}}}{\partial\mathfrak{y}_j}$, of the parameter estimate to observation. A desirable property of the model (1.1) is that the sensitivity of the parameter estimate to the observation should be small in order to minimize the effects of observation noise on the parameter estimate.

In practice solving the variational equations like (6.3) and (6.4) can represent a major computing problem. The M dimensional system (6.3) requires the Jacobian of the model equation (1.1) and at the same time equation (6.3) must be solved simultaneously with (1.1). This is a challenging problem when M and L are large and when the model equations are stiff.

Acknowledgments

We are grateful to professor Gennadii Bocharov for his comments on an earlier version of this paper.

7 Appendices

7.1 Some notation

Suppose that, for each $t \geq t_0$, $a(t,\mathbf{q})$ is a scalar quantity dependent upon the column vector $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{a}(t,\mathbf{q}) \in \mathbb{R}^n$ is a (column) vector dependent upon $\mathbf{q}$, and $\mathcal{A}(t,\mathbf{q}) \in \mathbb{R}^{n\times m}$ and $\mathcal{B}(t,\mathbf{q}) \in \mathbb{R}^{m\times l}$ are also matrices dependent upon $\mathbf{q}$. We define the *gradient* as a column vector,

$$\frac{\partial}{\partial\mathbf{q}}a(t,\mathbf{q}) := \Big[\frac{\partial}{\partial q_1}a(t,\mathbf{q}), \frac{\partial}{\partial q_2}a(t,\mathbf{q}), \cdots, \frac{\partial}{\partial q_m}a(t,\mathbf{q})\Big]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}. \tag{7.1}$$

The corresponding row vector is written as

$$\Big\{\frac{\partial}{\partial\mathbf{q}}\Big\}^{\mathrm{T}}a(t,\mathbf{q}) := \Big[\frac{\partial}{\partial q_1}a(t,\mathbf{q}), \frac{\partial}{\partial q_2}a(t,\mathbf{q}), \cdots, \frac{\partial}{\partial q_m}a(t,\mathbf{q})\Big] \in \mathbb{R}^{1\times m}. \tag{7.2}$$

In a similar manner we define the matrix

$$\left\{\frac{\partial}{\partial \mathbf{q}}\right\}^{\mathrm{T}} \mathbf{a}(t,\mathbf{q}) := \Big[\frac{\partial}{\partial q_1}\mathbf{a}(t,\mathbf{q}), \frac{\partial}{\partial q_2}\mathbf{a}(t,\mathbf{q}), \cdots, \frac{\partial}{\partial q_m}\mathbf{a}(t,\mathbf{q})\Big] \equiv \mathcal{A}(t,\mathbf{q}) \ \in \mathbb{R}^{n\times m}, \tag{7.3}$$

which allows us to define the *Hessian matrix* as

$$\left\{\frac{\partial^2}{\partial \mathbf{q}\partial \mathbf{q}^{\mathrm{T}}}\right\} a(t,\mathbf{q}) \quad := \quad \left\{\frac{\partial}{\partial \mathbf{q}}\right\}\left\{\frac{\partial}{\partial \mathbf{q}}\right\}^{\mathrm{T}} a(t,\mathbf{q}). \tag{7.4}$$

We may also define, in terms of (7.3),

$$\left\{\frac{\partial^2}{\partial \mathbf{q}\partial \mathbf{q}^{\mathrm{T}}}\right\} \mathbf{a}(t,\mathbf{q}) \quad := \quad \frac{\partial}{\partial \mathbf{q}}\mathcal{A}(t,\mathbf{q}) \ \in \mathbb{R}^{mn\times m}, \tag{7.5}$$

$$\frac{\partial(\mathcal{A}\mathcal{B})}{\partial \mathbf{q}} \quad = \quad \frac{\partial \mathcal{A}}{\partial \mathbf{q}}\mathcal{B} + \Big(\mathbf{I}_m \otimes \mathcal{A}\Big)\frac{\partial \mathcal{B}}{\partial \mathbf{q}} \ \in \mathbb{R}^{mn\times l}, \tag{7.6}$$

where I_m is an $m \times m$ identity matrix, and $\mathbf{q}$ is an $m-$component column vector; see [10].

7.2 The second order sensitivity coefficients for (2.1)

We here present the variational equations for the sensitivity coefficients, r_{ij}, of second order for the model (2.1). They comprise a set of coupled neutral delay differential equations:

$$\begin{aligned}
\frac{d}{dt}r_{11}(t,\mathbf{p}) &= s_1(t,\mathbf{p}) + \rho_0 r_{11}(t,\mathbf{p}) + \rho_1 r_{11}(t-\tau,\mathbf{p}) + \rho_2 r'_{11}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{12}(t,\mathbf{p}) &= s_2(t,\mathbf{p}) + \rho_0 r_{12}(t,\mathbf{p}) + s_1(t-\tau,\mathbf{p}) + \rho_1 r_{12}(t-\tau,\mathbf{p}) + \rho_2 r'_{12}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{13}(t,\mathbf{p}) &= s_3(t,\mathbf{p}) + \rho_0 r_{13}(t,\mathbf{p}) + \rho_1 r_{13}(t-\tau,\mathbf{p}) + s'_1(t-\tau,\mathbf{p}) + \rho_2 r'_{13}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{14}(t,\mathbf{p}) &= \rho_0 r_{14}(t,\mathbf{p}) + \rho_1 r_{14}(t-\tau,\mathbf{p}) + \rho_2 r'_{14}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{15}(t,\mathbf{p}) &= s_5(t,\mathbf{p}) + \rho_0 r_{15}(t,\mathbf{p}) + \rho_1 r_{15}(t-\tau,\mathbf{p}) - \rho_1 s'_1(t-\tau,\mathbf{p}) +\\
&\quad \rho_2 r'_{15}(t-\tau,\mathbf{p}) - \rho_2 s''_1(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{22}(t,\mathbf{p}) &= \rho_0 r_{22}(t,\mathbf{p}) + s_2(t-\tau,\mathbf{p}) + \rho_1 r_{22}(t-\tau,\mathbf{p}) + \rho_2 r'_{22}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{23}(t,\mathbf{p}) &= \rho_0 r_{23}(t,\mathbf{p}) + s_3(t-\tau,\mathbf{p}) + \rho_1 r_{23}(t-\tau,\mathbf{p}) + s'_2(t-\tau,\mathbf{p}) + \rho_2 r'_{23}(t-\tau,\mathbf{p})\\
\frac{d}{dt}r_{24}(t,\mathbf{p}) &= \rho_0 r_{24}(t,\mathbf{p}) + \rho_1 r_{24}(t-\tau,\mathbf{p}) + \rho_2 r'_{24}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{25}(t,\mathbf{p}) &= \rho_0 r_{25}(t,\mathbf{p}) + s_5(t-\tau,\mathbf{p}) + \rho_1 r_{25}(t-\tau,\mathbf{p}) - s'_2(t-\tau,\mathbf{p}) -\\
&\quad \rho_1 s'_2(t-\tau,\mathbf{p}) + \rho_2 r'_{25}(t-\tau,\mathbf{p}) - \rho_2 s''_2(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{33}(t,\mathbf{p}) &= \rho_0 r_{33}(t,\mathbf{p}) + \rho_1 r_{33}(t-\tau,\mathbf{p}) + s'_3(t-\tau,\mathbf{p}) + \rho_2 r'_{33}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{34}(t,\mathbf{p}) &= \rho_0 r_{34}(t,\mathbf{p}) + \rho_1 r_{34}(t-\tau,\mathbf{p}) + \rho_2 r'_{34}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{35}(t,\mathbf{p}) &= \rho_0 r_{35}(t,\mathbf{p}) + \rho_1 r_{35}(t-\tau,\mathbf{p}) - \rho_1 s'_3(t-\tau,\mathbf{p}) +\\
&\quad s'_5(t-\tau,\mathbf{p}) + \rho_2 r'_{25}(t-\tau,\mathbf{p}) - y''(t-\tau,\mathbf{p}) - \rho_2 s''_3(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{44}(t,\mathbf{p}) &= \rho_0 r_{44}(t,\mathbf{p}) + \rho_1 r_{44}(t-\tau,\mathbf{p}) + \rho_2 r'_{44}(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{45}(t,\mathbf{p}) &= \rho_0 r_{45}(t,\mathbf{p}) + \rho_1 r_{45}(t-\tau,\mathbf{p}) - \rho_1 s'_4(t-\tau,\mathbf{p})\\
&\quad + \rho_2 r'_{45}(t-\tau,\mathbf{p}) - \rho_2 s''_4(t-\tau,\mathbf{p}),\\
\frac{d}{dt}r_{55}(t,\mathbf{p}) &= \rho_0 r_{55}(t,\mathbf{p}) + \rho_1 r_{55}(t-\tau,\mathbf{p}) + \rho_1 z'(t-\tau,\mathbf{p}) - 2\rho_1 s'_5(t-\tau,\mathbf{p}),\\
&\quad +\rho_2 r'_{55}(t-\tau,\mathbf{p}) + \rho_2 z''(t-\tau,\mathbf{p}) - 2\rho_2 s''_5(t-\tau,\mathbf{p}).
\end{aligned}$$

The associated initial functions are

$$r_{ij}(t, \mathbf{p}) = \frac{\partial^2 \psi}{\partial p_i \partial p_j} \quad \text{for} \quad t \leq 0.$$

References

[1] D.M. Ates and D.G. Watis, Relative Curvature of Nonlinearity, *J.R. Statist.* **42**(1) (1980) 1-25.

[2] C.T.H. Baker and C.A.H. Paul, Pitfalls in Parameter Estimation for Delay Differential Equations, *SIAM J. Sci. Comput.* **18** (1997) 305-314.

[3] C.T.H. Baker, C.A.H. Paul and D.R. Willé, Issues in the Numerical Solution of Evolutionary Delay Differential Equations, *Adv. Comp. Math.* **3** (1995) 171-196.

[4] C.T.H. Baker, G.A. Bocharov and C.A.H. Paul, Mathematical Modelling of the Interleukin-2 T-Cell System: A Comparative Study of Approaches Based on Ordinary and Delay Differential Equations. *J. theor. Medicine* **2** (1997) 117-128.

[5] C.T.H. Baker, G.A. Bocharov, C.A.H. Paul and F.A. Rihan, Modelling and Analysis of Time-Lags in Cell Proliferation, *J. Math. Biol.* **37**(4) (1998) 341-371.

[6] C.T.H. Baker, G.A. Bocharov and F.A. Rihan, A Report on the Use of Delay Differential Equations in Numerical Modelling in the Biosciences, *MCCM Tech. Rep.,* **343** (1999), Manchester, ISSN 1360-1725.

[7] Y. Bard, *Nonlinear Parameter Estimation* (Academic Press, 1974).

[8] D.M. Bates and D.G. Watts, *Nonlinear Regression Analysis and its Applications* (New York: John Wiley & Sons, 1988).

[9] G.A. Bocharov and A.A. Romanyukha, Mathematical Model of Antiviral Immune-Response-III - Influenza-A Virus Infection, *J. theor. Biol.* **167** (1994) 323-360.

[10] J.W. Brewer, Kronecker Products and Matrix Calculus in System Theory, *IEEE Tranc. on Circuits and Systems* **25**(9) (1978) 772-781.

[11] I.L. Cameron and G.M. Padilla, *Cell Synchrony Studies in Biosynthetic Regulation* (New York: Academic Press, 1966).

[12] J. Chambers and M. Hill, Fitting Nonlinear Models: Numerical Techniques, *Biometrika* **60** (1973) 1-13.

[13] S.P. Corwin, D. Sarafyan and S. Thompson, DKLAG6: A Code Based on Continuously Embedded Sixth-Order Runge-Kutta Methods for the Solution of State-Dependent Functional Differential Equations, *Appl. Numer. Math.* **24** (1997) 317-330.

[14] R.P. Dickinson and R.J. Gelinas, Sensitivity Analysis of Ordinary Differential Equation System–a Direct Method, *J. Comput. Phys.* **21** (1976) 123-143.

[15] J.R., Dormand and P.J. Prince, Family of embedded Runge-Kutta formulae, *J. Coump. Appl. Math.* **6** (1980) 19–26.

[16] L. Eno, J.G. Beumee and H. Rabitz, Sensitivity Analysis of Experimental Data, *Appl. Math. Comput.* **16** (1985) 153-163.

[17] J.R. Leis and M.A. Kramer, The Simultaneous Solution and Sensitivity Analysis of Systems of Described by Ordinary Differential Equations, *ACM. Trans. Math. Software* **14** (1988) 45-60.

[18] G.I. Marchuk, A.A. Romanyukha and G.A. Bocharov, Mathematical Model of Antiviral Immune-Response. II.Parameter Identification for Acute Viral Hepatitis B, *J. theor. Biol.* **151** (1991) 41-70.

[19] J.M. Mitchison and W.S. Vincent, *A Method of Making Synchronous Cell Cultures by Density Gradient Centrifugation*, In: I.L. Cameron et. al: Cell synchrony. Studies in biosyncthetic regulation (New York: Academic Press 1966).

[20] K.W. Neves, Automatic Integration of Functional Differential Equations: an approach, *ACM Trans. Math. Software* **1** (1975) 369-371.

[21] K.W. Neves and S. Thompson, Software for Numerical Solution of Systems of Functional Differential Equations With State Dependent Delays, *Appl. Numer. Math.* **9** (1992) 385-401.

[22] H.J. Oberle and H.J. Pesch, Numerical Treatment of Delay Differential Equations by Hermite Interpolation, *Numer. Math.* **37** (1981) 235-255.

[23] C.A.H. Paul, *Runge-Kutta Methods for Functional Differential Equations* (PhD. Thesis. University of Manchester, 1992).

[24] C.A.H. Paul, A User-Guide to Archi - An Explicit Runge-Kutta Code for Solving Delay and Neutral Differential Equations and Parameter Estimation Problems, *MCCM report No. 283 (ISSN 1360-1725), University of Manchester* (revised edition, 1997).

[25] C.A.H. Paul and C.T.H. Baker, Explicit Runge-Kutta Methods for Numerical Solution of Singular Delay Differential Equations, *Numerical Analysis Report MCCM report No. 212, University of Manchester, 1992).*

[26] D.A. Ratkowsky, *Nonlinear Regression Modeling, A Unified Practical Approach* (New York: Marcel Dekker 1983).

[27] L.F. Shampine, Quadrature and Runge-Kutta Formulas, *Appl. Math. Comput.* **2** (1976) 161-171.

[28] S. Thompson, Stepsize Control for Delay Differential Equations Using Continuously Embedded Runge-Kutta Methods of Sarafyan, *J. Comput. Appl. Math.* **31**(2) (1990) 267-275.

[29] S. Thompson, Rootfinding and Interpolation With Runge-Kutta-Sarafyan Methods, *Trans. Soc. Computer Simulation* **2**(3) (1985) 207-218.