# Pseudo maximum likelihood estimation for differential equations

Nadia Lalam [a], Chris A. J. Klaassen [b, a]

[a]EURANDOM
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[b]Korteweg–de Vries Institute for Mathematics, University of Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

**Summary**

We consider a set of deterministic differential equations describing the temporal evolution of some system of interest, and containing an unknown finite-dimensional parameter to infer. The observations of the solution of the set of differential equations are assumed to be stochastically disturbed by two sorts of uncertainties: the state variables of the system are measured with errors, and they are not measured at the intended time, but at an unknown point in time. We present maximum likelihood based estimators of the parameter value and prove that these estimators are consistent, and even $\sqrt{N}$-consistent provided the statistical model is identifiable. The results are applied to the analysis of a gene regulatory network modelling the early development of a Drosophila embryo.

*Key words*: Consistency, Gene regulatory network, $M$-estimator, Ordinary differential equations, Parameter identification, Statistical modelling.
*2000 Mathematics Subject Classification*: Primary: 62F12, Secondary: 34A34, 65L99, 62P10.

# 1 Introduction

The formalism that we will study consists of a set of coupled ordinary differential equations (ODE's) accounting for the temporal evolution of some state variables describing a specific phenomenon. For example, this modelling is widely used in chemical engineering (Cutlip and Shacham, 1999), ecology (Turchin, 2001), epidemiology (Bailey, 1975), and physics (Bellman, 2003).

We consider state variables $g_k(t;\theta)$, $1 \leq k \leq K$, satisfying a system of differential equations where $\theta$ is the unknown finite-dimensional parameter of interest. Denote by $g(t;\theta)$ the vector $(g_k(t;\theta),\ 1 \leq k \leq K)$. We assume that the state variables $g_k(t;\theta)$ satisfy the following differential equations for all $1 \leq k \leq K$:

$$\frac{dg_k(t;\theta)}{dt} = \sigma_k(g(t;\theta), u(t), \theta),\ \text{for all } t \geq t_0, \tag{1}$$

with $g_k(t_0;\theta)$ independent of $\theta$ at the starting time $t_0$, and with $u(\cdot)$ an input function. The goal is estimation of the true parameter value $\theta_0$ arising in (1) from observations of the solution of the model. Note that statistical inference on the parameter is also called parameter identification in the differential equations literature (Voss *et al.*, 2004). This is traditionally done by computing the least squares estimators (Eason, 1976, Baake *et al.*, 1992, Mueller *et al.*, 2004, Li *et al.*, 2005) resulting from the minimization of the sum of squared differences between the observed data and the model. When the observation errors of the state variables are Gaussian, the least squares estimators coincide with the maximum likelihood estimators that are well-known to have desirable asymptotic properties such as consistency and efficiency (Cramér, 1946), the asymptotics consisting in letting the sample size increase to infinity. Least squares estimation yields useful results if time is observed exactly, or at least if it is observed with negligible errors. Typically however, this is not the case and often one misses the intended observation time by an unknown random amount of time. This is the situation we will study in this paper and we will present a general likelihood based method here. This so-called pseudo maximum likelihood method reduces to the least squares method whenever one realizes the intended observation times exactly.

We will introduce in section 2 the statistical setting allowing to model the observations of the solution of (1) that are supposed to be performed independently at different time-points. The proposed statistical model accounts for the errors inherent to the observation of the state variables, and for the errors inherent to the time at which the measurements are made. We will define in section 3 the pseudo maximum likelihood procedure that we propose to obtain asymptotically consistent estimators. The methodology is described in section 3.1, and an illustrative example of application is given in section 3.2. We will conclude with a discussion of the presented results in section 4. Proofs are deferred to Appendices.

## 2 Statistical model

We assume that observing an "individual" gives information on $g_k(t;\theta_0)$, for all $k$ in a subset of $\{1,\ldots,K\}$, and for one point in time only. We define a statistical model for observations on a group of "individuals" divided into $d$ subgroups with $d$ fixed. The data in a subgroup correspond to the observations collected for the same subset of $\{1,\ldots,K\}$ and with the same intended time-point. The data in each subgroup $i$, $1 \leq i \leq d$, are viewed as realizations of independent and identically distributed (i.i.d.) random variables $X_{ij}$, $1 \leq j \leq n_i$. Here $n_i$ is the number of observations in subgroup $i$. By $N = \sum_{i=1}^{d} n_i$ we denote the total number of observations. The assumption of independence of these random variables relies on the fact that each measurement comes from a different "individual", and each "individual" is observed at a single time-point. Each $X_{ij}$ is a table $(X_{ijk})_k$ where $k$ indexes the subset $\kappa_i$ of $\{1,\ldots,K\}$ for which the values of $(g_k(t_i;\theta_0))_k$ are observed at some time-point $t_i$.
Write

$$g_\theta = (g_k(t;\theta) : \, t \geq t_0, \, 1 \leq k \leq K)$$

for the solution of system (1) of ODE's under parameter value $\theta$ and with given initial conditions. Note that $X_{ij}$ from "individual" $j$, $1 \leq j \leq n_i$, of subgroup $i$, $1 \leq i \leq d$, contains only limited information about $g_\theta$, namely only about $g_k(t_i;\theta)$, $k \in \kappa_i$. Time $t_i$ and subset $\kappa_i$ characterize subgroup $i$. Observing "individual" $j$ in subgroup $i$ yields measurements $x_{ijk}$ of $g_k(t_i;\theta)$. We view these measurements $x_{ijk}$ as realizations of the random variables $X_{ijk}$, where in principle

$$\begin{aligned} X_{ijk} &= g_k(t_i + \tau_\delta \delta_{ij};\theta) + \tau_\varepsilon \varepsilon_{ijk}, \\ & k \in \kappa_i, \, 1 \leq j \leq n_i, \, 1 \leq i \leq d, \end{aligned} \tag{2}$$

holds with $\varepsilon_{ijk}$ and $\delta_{ij}$ random errors with variance 1. The unknown quantities $\tau_\varepsilon$ and $\tau_\delta$ are the standard deviations of the random errors in the state variable observation and the time at which the state variable is observed respectively.
We stress the fact that in this model, we do not observe the realized time $t_i + \tau_\delta \delta_{ij}$. We only know $t_i$, the intended time for observation and measurement. So, the complete set of data is

$$(x_{ijk}, t_i : \, 1 \leq j \leq n_i, \, k \in \kappa_i, \, 1 \leq i \leq d). \tag{3}$$

Actually, we will use the linear approximation

$$g_k(t_i + \tau_\delta \delta_{ij};\theta) \approx g_k(t_i;\theta) + \sigma_k(t_i;\theta)\tau_\delta \delta_{ij} \tag{4}$$

with $\sigma_k(t_i;\theta) = \sigma_k(g(t_i;\theta), u(t_i), \theta)$ the right-hand side of (1) at time $t_i$. Together (2) and (4) yield our statistical model

$$X_{ijk} = g_k(t_i;\theta) + \sigma_k(t_i;\theta)\tau_\delta \delta_{ij} + \tau_\varepsilon \varepsilon_{ijk}, \tag{5}$$

where all $\varepsilon_{ijk}$ are assumed to be i.i.d. with unknown density function $f(\cdot)$ with variance 1, and all $\delta_{ij}$ are assumed to be i.i.d. with unknown density function $p(\cdot)$ with variance 1. We also assume that $\varepsilon_{ijk}$ and $\delta_{ij}$ are independent and centered with mean 0.

Within the model given by (1) and (5), we aim at estimating the true value of the unknown parameter $\gamma = (\theta, \tau_\varepsilon^2, \tau_\delta^2)$ based on the data (3). Here $\gamma$ is assumed to belong to a set $\Gamma \subset \Theta \times \mathbb{R}_+^2$, but $\theta$ remains the main parameter of interest. In case one is not sure about the assumption of the error $\delta_{ij}$ having mean 0, one might consider to view $t_i$ as unknown and add these $t_i$ as additional nuisance parameters to $\gamma$. Another possible intermediary case is that the intended time is perturbed by a systematic bias considered as an additional nuisance parameter to $\gamma$.

# 3 Pseudo maximum likelihood estimation

## 3.1 Methodology and results

We will use a likelihood based method to estimate the parameter values. The joint likelihood of the random variables $(X_{ijk})_{i,j,k}$ satisfying (5) is equal to

$$\prod_{i=1}^{d}\prod_{j=1}^{n_i}\int_{\mathbb{R}}[\prod_{k\in\kappa_i}\frac{1}{\tau_\varepsilon}f(\frac{X_{ijk} - g_k(t_i;\theta) - \sigma_k(t_i;\theta)\tau_\delta y}{\tau_\varepsilon})]p(y)dy. \tag{6}$$

Since the densities $f(\cdot)$ and $p(\cdot)$ are unknown, we propose to estimate the true parameter value of $\gamma$, denoted by $\gamma_0 = (\theta_0, \tau_{\varepsilon_0}^2, \tau_{\delta_0}^2)$, by using the pseudo maximum likelihood methodology. This approach consists in replacing the densities $f(\cdot)$ and $p(\cdot)$ in the above relationship by the standard normal density function $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, and in subsequently maximizing the resulting pseudo likelihood function. This estimation technique is classically used in econometrics (Greene, 2003).

Recall that $N = \sum_{i=1}^{d} n_i$ denotes the total number of observations. Straightforward but tedious computation shows that the pseudo maximum likelihood estimator (PMLE) $\widehat{\gamma}_N = (\widehat{\theta}_N^{PMLE}, \widehat{\tau}_{\varepsilon_N}^{2\,PMLE}, \widehat{\tau}_{\delta_N}^{2\,PMLE})$ of the true parameter value $\gamma_0$ minimizes the following pseudo likelihood based cost function $M_N(\cdot)$ which is $-2/N$ times the joint log-density of all the $X_{ijk}$ if the densities $f(\cdot)$ and $p(\cdot)$ were standard Gaussian in (6) (with the constant term $\sum_{i=1}^{d} n_i|\kappa_i|\log(2\pi)/N$ not taken into account, $|\kappa_i|$ being the cardinality of $\kappa_i$):

$$M_N(\gamma) = \frac{1}{N} \sum_{i=1}^{d} \sum_{j=1}^{n_i} m_i(X_{ij}, \gamma) \tag{7}$$

with

$$\begin{aligned} m_i(X_{ij}, \gamma) &= (|\kappa_i| - 1) \log \tau_\varepsilon^2 + \log(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k \in \kappa_i} \sigma_k^2(t_i; \theta)) \\ &+ \frac{1}{\tau_\varepsilon^2} \sum_{k \in \kappa_i} (X_{ijk} - g_k(t_i; \theta))^2 - \frac{\tau_\delta^2 [\sum_{k \in \kappa_i} (X_{ijk} - g_k(t_i;\theta))\sigma_k(t_i;\theta)]^2}{\tau_\varepsilon^2(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta))}. \end{aligned} \tag{8}$$

*Remark 1 If the time measurement is assumed to be performed without error, that is $\tau_\delta = 0$, then the PMLE $\widehat{\theta}_N^{PMLE}$ reduces to the ordinary least squares estimator.*

We will prove that $\widehat{\gamma}_N$ is consistent and typically even $\sqrt{N}$-consistent as the total number of observations $N$ tends to infinity.
Assume that

$$\lim_{N\to\infty} \frac{n_i}{N} = p_i > 0, \text{ for all } 1 \le i \le d, \tag{9}$$

holds. By the strong law of large numbers, $M_N(\gamma)$ converges a.s. under the true parameter value $\gamma_0$ to

$$\begin{aligned} M(\gamma;\gamma_0) = \sum_{i=1}^{d} p_i \Bigg\{ &(|\kappa_i| - 1)\log\tau_\varepsilon^2 + \log(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)) \\ &+ \frac{1}{\tau_\varepsilon^2} \sum_{k\in\kappa_i} [\tau_{\varepsilon_0}^2 + \tau_{\delta_0}^2 \sigma_k^2(t_i;\theta_0) + (g_k(t_i;\theta_0) - g_k(t_i;\theta))^2] \\ &- \frac{\tau_\delta^2}{\tau_\varepsilon^2(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i}\sigma_k^2(t_i;\theta))} \Bigg[ \tau_{\varepsilon_0}^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta) + \tau_{\delta_0}^2 [\sum_{k\in\kappa_i} \sigma_k(t_i;\theta_0)\sigma_k(t_i;\theta)]^2 \\ &+ [\sum_{k\in\kappa_i} (g_k(t_i;\theta_0) - g_k(t_i;\theta))\sigma_k(t_i;\theta)]^2 \Bigg] \Bigg\}. \end{aligned} \tag{10}$$

**Proposition 3.1** *The parameter value $\gamma_0$ minimizes $M(\gamma;\gamma_0)$ with respect to $\gamma$, that is, with $\Gamma_0 = \{\gamma_* \in \Gamma : \gamma_* = argmin_{\gamma\in\Gamma} M(\gamma;\gamma_0)\}$ denoting the set of minimizers, $\gamma_0 \in \Gamma_0$ holds.*

**Proof**. *See Appendix A.*

If $\Gamma_0$ consists of the singleton $\{\gamma_0\}$, then the statistical model is locally identifiable in the sense that there does not exist a $\gamma \neq \gamma_0$ such that the observations have the same distribution under $\gamma$ and $\gamma_0$. Denote by $P_{\gamma_0}$ the probability under $\gamma_0$, and let $d(\cdot,\cdot)$ be the Euclidean distance on $\Gamma$.

**Theorem 3.2 Consistency**. *Consider the model defined by (5) and (1), and assume that (9) holds. If the functions $\sigma_k(\cdot,\cdot,\cdot)$ are Lipschitz continuous in their first and third argument with Lipschitz constants that are independent of the second argument, and if $\Gamma$ is compact, then the pseudo maximum likelihood estimator $\widehat{\gamma}_N$ defined as the minimizer of (7), is consistent in the sense that, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} P_{\gamma_0}(\inf_{\gamma\in\Gamma_0} d(\widehat{\gamma}_N,\gamma) \geq \varepsilon) = 0.$$

*In particular, if the set of minimizers $\Gamma_0$ contains only $\gamma_0$ , then $\widehat{\gamma}_N$ is consistent under $\gamma_0$ in the classical sense, that is, for all $\epsilon > 0$,*

$$\lim_{N\to\infty} P_{\gamma_0}(d(\widehat{\gamma}_N,\gamma_0) \geq \epsilon) = 0.$$

**Proof**. *See Appendix B.*

**Theorem 3.3 $\sqrt{N}$-consistency**. *Consider the model given by (5) and (1), and assume that (9) holds. Let $\Gamma_0 = \{\gamma_0\}$ hold and let the functions $\sigma_k(\cdot,\cdot,\cdot)$ be twice continuously differentiable in their first and third argument with their first and second derivatives bounded uniformly in the second argument of $\sigma_k(\cdot,\cdot,\cdot)$. Then the matrix of second derivatives of $\sum_{i=1}^{d} p_i m_i(X_{ij},\gamma)$ with respect to $\gamma$ at $\gamma=\gamma_0$ exists a.s. and has finite expectation $-I(\gamma_0)$ under $\gamma_0$. If $\widehat{\gamma}_N$ is consistent under $\gamma_0$ and the matrix $I(\gamma_0)$ is nonsingular, then $\widehat{\gamma}_N$ is $\sqrt{N}$-consistent at $\gamma_0$, that is*

$$\lim_{M\to\infty} \limsup_{N\to\infty} P_{\gamma_0}(\sqrt{N} d(\widehat{\gamma}_N,\gamma_0) > M) = 0,$$

*and $\widehat{\gamma}_N$ is even asymptotically normal at $\gamma_0$.*

**Proof**. *See Appendix C.*

Note that the continuous differentiability conditions and the boundedness of the derivatives of $\sigma_k(\cdot,\cdot,\cdot)$ in Theorem 3.3 imply the Lipschitz continuity of Theorem 3.2.

Under the normality assumption of the errors $\delta_{ij}$ and $\varepsilon_{ijk}$ in (5), the PMLE reduces to the MLE (maximum likelihood estimator) which is asymptotically efficient meaning that its asymptotic distribution is Gaussian with covariance matrix equal to the inverse of the Fisher information matrix $I(\gamma_0)$, which is the asymptotic bound on the performance of the estimators.

## 3.2 Application to a gene regulatory network

The previous results may be applied to the analysis of the behavior of a biochemical system such as gene products interacting in a regulatory network (de

Jong, 2002, Bolouri and Davidson, 2002). The study of gene regulatory networks constitutes a growing field in functional genomics (Kitano, 2003, Tyers and Mann, 2003). The amount of omics data on these networks is increasing at a fast rate (Quackenbush, 2004). The extraction of relevant information from these data requires the development of adequate mathematical tools (Hasty *et al.*, 2001). For example, let us consider the developmental gene regulatory network model, developed by Mjolsness *et al.* (1991), which is formed by $N_g$ genes and accounts for the early *Drosophila melanogaster* segmentation. Mjolsness *et al.* (1991) noted that one can approximate the region of interest for the emergence of early Drosophila segmentation patterns by a line of $K$ nuclei along the embryo anterior-posterior axis. The gene product concentrations $g_{k\ell}(t;\theta)$ for gene product $\ell$, $1 \leq \ell \leq N_g$, in nucleus $k$, $1 \leq k \leq K$, are assumed to satisfy the following set of coupled nonlinear ODE's of reaction-diffusion type, with $1 \leq k \leq K$ and $1 \leq \ell \leq N_g$:

$$\begin{aligned} \frac{dg_{k\ell}(t;\theta)}{dt} &= R_\ell \Phi(\sum_{\ell'=1}^{N_g} W_{\ell\ell'} g_{k\ell'}(t;\theta) + m_\ell g_{k\,bcd}(t) + h_\ell) \\ &\quad + D_\ell[g_{k-1\,\ell}(t;\theta) - 2g_{k\ell}(t;\theta) + g_{k+1\,\ell}(t;\theta)] - \lambda_\ell g_{k\ell}(t;\theta), \end{aligned} \tag{11}$$

with $\theta = ((R_\ell, m_\ell, h_\ell, D_\ell, \lambda_\ell)_{1\leq\ell\leq N_g}, (W_{\ell\ell'})_{1\leq\ell,\ell'\leq N_g})$. Model (11) is a particular case of (1) for which the index $k$ is replaced by the index $(k,\ell)$. The function $\Phi : \mathbb{R} \rightarrow [0,1]$ is assumed to be one-to-one, onto, increasing, and even sigmoidal. Such an S-shaped distribution function accounts for the nonlinearity of the model. An often used example is $\Phi(x) = 0.5[(x/\sqrt{x^2+1}) + 1]$.

The first term at the right-hand side of (11) models the gene regulation and gene product synthesis, the second term models the exchange by diffusion of gene products between neighboring nuclei, and the third term accounts for the decay of the gene product. No diffusion is assumed to occur at both ends of the linear array of nuclei, $k = 1$ and $k = K$, for which equation (11) holds with the second term at the right-hand side ignored. The initial conditions are described in the supplementary information of Jaeger *et al.* (2004).

The interaction between genes $\ell$ and $\ell'$ is represented by a single real number $W_{\ell\ell'}$ accounting for a connection strength or weight. If the product of gene $\ell'$ activates gene $\ell$, then $W_{\ell\ell'}$ is positive; if the product of gene $\ell'$ represses gene $\ell$, then $W_{\ell\ell'}$ is negative; if genes $\ell$ and $\ell'$ do not interact, then $W_{\ell\ell'} = 0$. The bias term $m_\ell g_{k\,bcd}(t)$ arises from the *bicoid* protein (*bcd*), which is treated as an input element in the model that induces the Drosophila segmentation (Reinitz and Sharp, 1995), and which equals the input function $u(t)$ in (1). The parameter $m_\ell$ is the connection strength of *bcd* acting on gene $\ell$, that is $m_\ell = W_{\ell\,bcd}$. The maximum rate of synthesis from gene $\ell$ is denoted by $R_\ell$. The parameter $h_\ell$ is the activation threshold of gene $\ell$, and $D_\ell$ is its diffusion parameter. The quantity $\lambda_\ell$ is the decay rate of the product of gene $\ell$.

In order to infer the true value of the parameter $\theta$ in (11), we may rely on gene expression data collected at the Reinitz Fly Lab[1], Stony Brook University, New York. As concerning these data obtained by confocal laser scanning microscopy, exactly three protein concentrations are measured for each fixed embryo (Kosman *et al.*, 1998, Poustelnikova *et al.*, 2004) whereas the Drosophila segmentation gene network contains $N_g = 16$ interacting genes (Kozlov *et al.*, 2000). Such proteomic data fit into the statistical model (5) of section 2 for which an observation of an "individual" (here a Drosophila embryo) gives insight only into part of the vector of state variables (here the gene product concentrations).

The methodology used by Reinitz and Sharp (1995) to derive an estimator of the true value of the biological parameter $\theta$ from (11) is the minimization of the sum of squared differences between the data and the model, yielding a least squares estimator (LSE).

*Remark 2 Set the averages* $X_{i\,.\,k} = 1/n_i \sum_{j=1}^{n_i} X_{ijk}$*. Noticing that* $\sum_{i,j,k}(X_{ijk} - X_{i\,.\,k})(X_{i\,.\,k} - g_k(t_i;\theta)) = 0$*, one obtains that the LSE also minimizes the quantity* $\sum_{i,k} n_i(X_{i\,.\,k} - g_k(t_i;\theta))^2$*. It is therefore equivalent to derive the LSE from the individual data* $x_{ijk}$ *or from the averaged data* $x_{i\,.\,k}$*.*

The uncertainty about the time at which the Drosophila embryo is observed is not accounted for in the LSE introduced by Reinitz and Sharp (1995). In addition, another possible limitation of the current LSE is that it relies on so-called integrated data which result from the merging of individual, registered data from several embryos belonging to the same temporal class and stained for different combinations of three gene products. The data registration is explained in more detail in Kozlov *et al.* (2000). With the PMLE from section 3.1, we may extract the information contained in the individual unregistered data in a more direct manner, and account for the uncertainty in the time at which the data are collected.

# 4 Conclusion

We have proposed a pseudo maximum likelihood procedure to infer the parameters arising in a system of coupled differential equations (1) based on the statistical model (5). This statistical modelling accounts for two kinds of uncertainties in the observations of the solution of the system: the uncertainty in the measurement of the state variable of the differential equations, and the uncertainty in the time at which the observation of the state variable is made. We have proved the consistency and $\sqrt{N}$-consistency of the PMLE of the parameters arising in

[1]http://flyex.ams.sunysb.edu/flyex/

(1) and (5). We provided results in a quite general setting with a large range of applications. For example, our method may be applied to analyze gene regulatory networks as defined in Mjolsness *et al.* (1991) from proteomic data collected by confocal laser scanning microscopy.

It would be of interest to study the statistical model (5) under the general assumption that the densities of the errors in state variable measurements and in time determination are unknown nuisance functions. Within this semiparametric formalism, a line of research is the investigation of asymptotically efficient estimators of the true value of the finite-dimensional parameter $\theta$. These estimators may be defined as one-step estimators constructed following a Newton-Raphson approach based on the preliminary pseudo maximum likelihood estimators presented in this study. This research will be presented elsewhere.

# Acknowledgements

We are grateful to Richard Gill for stimulating suggestions and to Jaap Kaandorp for introducing us to the problem of parameter identification in (11).

# References

Baake, E., Baake, M., Bock, H. G., Briggs, K. M. (1992) Fitting ordinary differential equations to chaotic data, *Physical Review A.*, **45**, 5524-5529.

Bailey, N. T. J. (1975) The mathematical theory of infectious diseases and its applications, Charles Griffin and Company.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. (1998) Efficient and adaptive estimation for semiparametric models, Springer.

Bellmann, R. E. (2003) Perturbation techniques in mathematics, engineering and physics, Dover Publications.

Bolouri, H., Davidson, E. H. (2002) Modeling transcriptional regulatory networks, *BioEssays*, **24**, 1118-1129.

Cramér, H. (1946) Mathematical methods of Statistics, Princeton University Press.

Cutlip, M. B., Shacham, M. (1999) Problem solving in chemical engineering with numerical methods, Prentice Hall.

Eason, E. A. (1976) A review of least-squares methods for solving partial differential equations, *International Journal for Numerical Methods in Engineering*, **10**, 1021-1046.

Greene, W. H. (2003) Econometric analysis. Prentice Hall.

Hasty, J., McMillen, D., Isaacs, F., Collins, J. J. (2001) Gene regulatory networks: in numero molecular biology, *Nature Reviews Genetics*, **2**, 268-279.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H., Reinitz, J. (2004) Dynamic control of positional information in the early Drosophila embryo, *Nature*, **430**, 368-371.

de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review, *Journal of Computational Biology*, **9**, 67-103.

Kitano, H. (2003) Computational systems biology, *Nature*, **420**, 206-210.

Kosman, D., Small, S., Reinitz, J. (1998) Rapid preparation of a panel of polyclonal antibodies to Drosophila segmentation proteins, *Development Genes and Evolution*, **208**, 290-294.

Kozlov, K., Myasnikova, E., Samsonova, M., Reinitz, J., Kosman, D. (2000) Method for spatial registration of the expression patterns of Drosophila segmentation genes using wavelets, *Computational Technologies*, **5**, 112-119.

Li, Z., Osborne, M. R., Prvan, T. (2005) Parameter estimation of ordinary differential equations, *Journal of Numerical Analysis*, **25**, 264-285.

Mjolsness, E., Sharp, D. H., Reinitz, J. (1991) A connectionist model of development, *Journal of Theoretical Biology*, **152**, 429-453.

Mueller, T. G., Timmer, J. (2004) Parameter identification techniques for partial differential equations, *International Journal of Bifurcation and Chaos*, **14**, 2053-2060.

Osborne, M. R. (1997) Cyclic reduction, dichotomy, and the estimation of differential equations, *Journal of Computational and Applied Mathematics*, **86**, 271-286.

Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M., Reinitz, J. (2004) A database for management of gene expression in situ, *Bioinformatics*, **20**, 2212-2221.

Quackenbush, J. (2004) Data standards for omics science, *Nature Biotechnology*, **22**, 613-614.

Reinitz, J., Sharp, D. H. (1995) Mechanisms of eve stripe formation, *Mechanisms of Development*, **49**, 133-158.

Turchin, P. (2001) Does population ecology have general laws ? *Oikos*, **94**, 17-26.

Tyers, M., Mann, M. (2003) From genomics to proteomics, *Nature*, **422**, 193-197.

van der Vaart, A. W. (1998) Asymptotic Statistics, Cambridge University Press.

Voss, H. U., Timmer, J., Kurths, J. (2004) Nonlinear dynamical system identification from uncertain and indirect measurements, *International Journal of Bifurcation and Chaos*, **14**, 1905-1933.

# Appendix A: Proof of Proposition 3.1

The asymptotic deterministic criterion $M(\cdot;\gamma_0)$ defined in (10) may be rewritten as

$$
\begin{aligned}
M(\gamma;\gamma_0) = \sum_{i=1}^{d} p_i \Bigg\{ & (|\kappa_i|-1)[\frac{\tau_{\varepsilon_0}^2}{\tau_\varepsilon^2} + \log\tau_\varepsilon^2] + \log(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)) \\
& + \frac{1}{\tau_\varepsilon^2}[\tau_{\varepsilon_0}^2 + \tau_{\delta_0}^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta_0) + \sum_{k\in\kappa_i} (g_k(t_i;\theta_0) - g_k(t_i;\theta))^2] \\
& - \frac{\tau_\delta^2}{\tau_\varepsilon^2(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta))} \Bigg[ \tau_{\varepsilon_0}^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta) + \tau_{\delta_0}^2 [\sum_{k\in\kappa_i} \sigma_k(t_i;\theta_0)\sigma_k(t_i;\theta)]^2 \\
& \qquad + [\sum_{k\in\kappa_i} (g_k(t_i;\theta_0) - g_k(t_i;\theta))\sigma_k(t_i;\theta)]^2 \Bigg] \Bigg\}.
\end{aligned}
$$

Applying twice both the Cauchy-Schwarz inequality and the inequality $\alpha/x + \log(x) \geq 1 + \log(\alpha)$, for all $x > 0$, $\alpha > 0$, yields

$$\begin{aligned} M(\gamma;\gamma_0) &\geq \sum_{i=1}^{d} p_i \left\{ (|\kappa_i| - 1)[\frac{\tau_{\varepsilon_0}^2}{\tau_\varepsilon^2} + \log \tau_\varepsilon^2] + \log(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)) \right. \\ &\quad \left. + \frac{\tau_{\varepsilon_0}^2 + \tau_{\delta_0}^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta_0) + \sum_{k\in\kappa_i} (g_k(t_i;\theta_0) - g_k(t_i;\theta))^2}{\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)} \right\} \\ &\geq \sum_{i=1}^{d} p_i \left\{ (|\kappa_i| - 1)[1 + \log(\tau_{\varepsilon_0}^2)] + 1 + \log(\tau_{\varepsilon_0}^2 + \tau_{\delta_0}^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta_0) \right. \\ &\quad \left. + \sum_{k\in\kappa_i} (g_k(t_i;\theta_0) - g_k(t_i;\theta))^2) \right\}. \end{aligned}$$

Denote by $F(\theta;\gamma_0)$ the right-hand side of this string of inequalities. This function $F(\cdot;\gamma_0)$ is minimized at $\theta_0$ with value $F(\theta_0;\gamma_0) = M(\gamma_0;\gamma_0)$. Therefore, the parameter value $\gamma_0$ minimizes $M(\cdot;\gamma_0)$ as well, that is, $\Gamma_0$ contains $\gamma_0$.

# Appendix B: Proof of Theorem 3.2

We apply Wald's consistency result as given in Theorem 5.14 of van der Vaart (1998). Recall that $\gamma = (\theta, \tau_\varepsilon^2, \tau_\delta^2)$ belongs to the compact subset $\Gamma$ of $\Theta \times \mathbb{R}_+^2$. First note that Cauchy-Schwarz implies

$$\begin{aligned} m_i(X_{ij},\gamma) &\geq (|\kappa_i| - 1)\log \tau_\varepsilon^2 + \log(\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)) \\ &\quad + \frac{\sum_{k\in\kappa_i} (X_{ijk} - g_k(t_i;\theta))^2}{\tau_\varepsilon^2 + \tau_\delta^2 \sum_{k\in\kappa_i} \sigma_k^2(t_i;\theta)} \\ &\geq |\kappa_i| \log \tau_\varepsilon^2. \end{aligned}$$

Consequently, we have $E_{\gamma_0}(\inf_{\gamma\in\Gamma} m_i(X_{ij},\gamma)) > -\infty$, where $E_{\gamma_0}$ is the expectation under the true parameter value. Hence it suffices to show that $\gamma \mapsto m_i(X_{ij},\gamma)$ is continuous a.s., $1 \leq i \leq d$. To this end, in view of the expression of $m_i(X_{ij},\gamma)$ in (8), it suffices to prove that for each $t$, the map $\theta \mapsto g(t;\theta)$ is continuous. Note that in an obvious vector notation (1) implies

$$g(t;\theta) - g(s_1;\theta) = \int_{s_1}^{t} \sigma(g(s;\theta), u(s), \theta) ds$$

and hence, for $t_0 \leq s_1 \leq t$,

$$|g(t;\theta_1) - g(t;\theta_2)| \leq |g(s_1;\theta_1) - g(s_1;\theta_2)| + \int_{s_1}^{t} |\sigma(g(s;\theta_1), u(s), \theta_1) - \sigma(g(s;\theta_2), u(s), \theta_2)| ds,$$

by the triangle inequality with $|\cdot|$ the Euclidean norm on $\mathbb{R}^K$. For $t_0 \leq s_1 < s_2$, this yields by the Lipschitz properties of $\sigma(\cdot,\cdot,\cdot)$

$$\begin{aligned} \text{sup}_{s_1\leq t\leq s_2}|g(t;\theta_1) - g(t;\theta_2)| \leq\ & |g(s_1;\theta_1) - g(s_1;\theta_2)| \\ & +C_1(s_2 - s_1)\text{sup}_{s_1\leq s\leq s_2}|g(s;\theta_1) - g(s;\theta_2)| \\ & +C_3(s_2 - s_1)||\theta_1 - \theta_2|| \end{aligned}$$

with $||\cdot||$ the Euclidean norm on $\Theta$. Hence

$$\text{sup}_{s_1\leq s\leq s_2}|g(s;\theta_1) - g(s;\theta_2)| \leq \frac{|g(s_1;\theta_1) - g(s_1;\theta_2)| + C_3(s_2 - s_1)||\theta_1 - \theta_2||}{1 - C_1(s_2 - s_1)}$$

provided $1 - C_1(s_2 - s_1) > 0$. It follows that for $r$ sufficiently small

$$\begin{aligned} \text{sup}_{t_0\leq s\leq t}|g(s;\theta_1) - g(s;\theta_2)| \leq\ & \frac{1}{1-r}\text{sup}_{t_0\leq s\leq t-rC_1^{-1}}|g(s;\theta_1) - g(s;\theta_2)| \\ & +\frac{rC_3C_1^{-1}}{1-r}||\theta_1 - \theta_2|| \end{aligned}$$

holds, and by iteration

$$\begin{aligned} \text{sup}_{t_0\leq s\leq t}|g(s;\theta_1) - g(s;\theta_2)| \leq\ & (1-r)^{-\frac{C_1(t-t_0)}{r}}|g(t_0;\theta_1) - g(t_0;\theta_2)| \\ & +C_3C_1^{-1}||\theta_1 - \theta_2||r\sum_{j=1}^{\lceil C_1(t-t_0)/r\rceil}(\frac{1}{1-r})^j, \end{aligned} \tag{12}$$

where $\lceil x\rceil$ is the smallest integer larger than $x$. Since $g(t_0;\theta)$ does not depend on $\theta$, the right-hand side of (12) equals at most

$$\begin{aligned} & C_3C_1^{-1}||\theta_1 - \theta_2||r(1-r)^{-\frac{C_1(t-t_0)}{r}}\sum_{i=0}^{\infty}(1-r)^i \\ = \ & C_3C_1^{-1}||\theta_1 - \theta_2||(1-r)^{-\frac{C_1(t-t_0)}{r}}, \end{aligned}$$

which as $r \downarrow 0$ converges to $C_3C_1^{-1}||\theta_1 - \theta_2||e^{C_1(t-t_0)}$. This proves the continuity of $\theta \mapsto g(t;\theta)$ (in fact, uniformly in $t$ in a compact) and hence the first part of the theorem, which implies the second part straightforwardly.

# Appendix C: Proof of Theorem 3.3

Denoting differentiation of (the components of) $\sigma(\cdot,\cdot,\cdot)$ in the first argument by $'$ and in the third argument by $\dot{}$, we see that, using matrix notation,

$$\begin{aligned}\dot{g}(t;\theta) &= [\int_{t_0}^{t} \dot{\sigma}(g(s;\theta),u(s),\theta)\exp\{-\int_{t_0}^{s}\sigma'(g(v;\theta),u(v),\theta)dv\}ds] \\ &\quad .\exp\int_{t_0}^{t}\sigma'(g(s;\theta),u(s),\theta)ds \qquad (13)\end{aligned}$$

solves the equation that results from differentiation of (1) with respect to $\theta$, that is

$$\dot{g}'(t;\theta) = \sigma'(g(t;\theta),u(t),\theta)\dot{g}(t;\theta) + \dot{\sigma}(g(t;\theta),u(t),\theta).$$

Note that the boundedness of the derivatives implies the existence of the integrals in (13) and hence the continuous differentiability of $g(t;\theta)$. Repeating this argument, we see that $\theta \mapsto g(t;\theta)$ is twice continuously differentiable and consequently that $\gamma \mapsto \sum_{i=1}^{d} p_i m_i(X_{ij},\gamma)$ is, a.s. Furthermore, for a sufficiently small neighborhood of $\gamma_0$, the supremum over this neighborhood of the absolute values of these second derivatives of $\sum_{i=1}^{d} p_i m_i(X_{ij},\gamma)$ is a function in the $X_{ijk}$ that grows at most quadratically in these $X_{ijk}$. Therefore, it has finite expectation under $\gamma_0$. We have verified now conditions (D1) and (D2) of Corollary 7.2.1, page 303, of Bickel *et al.* (1998), and we note that its condition (D3) means the nonsingularity of $I(\gamma_0)$. This entails the $\sqrt{N}$-consistency of $\widehat{\gamma}_N$. The proof is complete.