Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling?

R. Muñoz-Tamayo, L. Puillet, J.B. Daniel, D. Sauvant, O. Martin, M. Taghipoor, and P. Blavy

**Supplementary material S3**

**Calculation of confidence interval from the Fisher Information Matrix and a brief comment on optimal experiment design**

*Assessment of the uncertainty of the parameter estimates*

In this section, we recall the theoretical framework for assessing the uncertainty of the parameter identification by using the Fisher Information Matrix following the classic book of [Walter and Pronzato, 1997](#_ENREF_6).

Let us consider the following model described by ordinary differential equations

$$\frac{dx(t)}{dt}=f\left(x,u,p, t\right), x\left(0\right)=x\_{0} $$

$y\_{m}(t)=g(x,u,p,t)$ (1)

where $t$ is the time, $x$ is the vector of state variables, $y\_{m}$is the vector of model observables, and $u$isthe vector of external stimuli (input vector). The equations contain a set of parameters defined by the vector$ p$, and **f,** **g** are vector functions.

When real experimental data are available, represented here by the vector $y(t)$, we can proceed to the model calibration step by finding the vector $p$that minimizes a cost function of the distance between the real measurements $y(t)$ and the model observables $y\_{m}(t)$.

It is typical to assume that the vector of experimental data collected at time $t\_{i}$ follows

$y\left(t\_{i}\right)=y\_{m}\left(t\_{i},p^{\*}\right)+ε\_{i}, i=1,2, …, n\_{t}$ (2)

where $n\_{t} $is the number of observation times, $y\_{m}\left(t\_{i},p^{\*}\right) $is the predicted observable of the model with $p^{\*}$ the true value of the parameter vector and $ε\_{i}$ is the vector of measurement errors which will be assumed here to follow a normal distribution: $ε\_{i}\~N(0,Σ)$.

The model calibration can be performed by the maximum likelihood (ML) approach. If the covariance matrix $Σ$ is known, the maximum likelihood estimator minimizes the weighted least squares function. Once the parameters estimates ($\hat{p}$) are found by an adequate optimization procedure, we can assess the parameter uncertainty via the computation of the FIM at the estimated value $\hat{p}$as detailed below.

The FIM can be calculated as

 $FIM(\hat{p})=\sum\_{i=1}^{n\_{t}}\left[\frac{∂y\_{m}}{∂p}\right]\_{(t\_{i},\hat{p})}^{T}Σ\left[\frac{∂y\_{m}}{∂p}\right]\_{(t\_{i},\hat{p})}$ (3)

The term $\frac{∂y\_{m}}{∂p}$ contains the sensitivities of the observables with respect to the parameters. The calculation of the sensitivities can be performed by symbolic manipulation of the model equations using dedicated software such as the Matlab® Toolbox IDEAS ([Muñoz-Tamayo *et al.*, 2009](#_ENREF_4)), which is freely available at <http://genome.jouy.inra.fr/logiciels/IDEAS>.

Under a number of technical assumptions that include theoretical identifiability, the covariance matrix $P$of the ML estimator satisfies

$ P \geq FIM^{-1}(p^{\*})$ (4)

This equation is known as the Cramér-Rao inequality. An approximate $\hat{P}$ of the covariance matrix of the parameters can be computed at the Cramér-Rao lower bond evaluated at $\hat{p}$ $ $

$\hat{P}=FIM^{-1}(\hat{p})$ (5)

It must be kept in mind that this approximation is only valid asymptotically, when the number of data points tends to infinity, the statistical hypotheses on the noise are satisfied, and that $\hat{p}$is close to $p^{\*}$**.** Furthermore, this approach is based on a linear approximation of the observables with respect to the parameters, which may be inadequate because of model nonlinearities ([Carson *et al.*, 1983](#_ENREF_2), [Marsili-Libelli *et al.*, 2003](#_ENREF_3), [Raue *et al.*, 2009](#_ENREF_5)). When these idealized conditions are far from being satisfied, this evaluation of the uncertainty on the estimates via the FIM has thus to be considered with caution.

The diagonals of $\hat{P}$are the variances of the parameter estimates. Thus, the square root $σ\_{j}$ of the *j*th diagonal element of $\hat{P} $is an estimate of the standard deviation of the parameter $\hat{p}\_{j}$. On this basis, an approximate 95% confidence interval of the parameter $\hat{p}\_{j}$ can be calculated as $\hat{p}\_{j}\pm 2∙σ\_{j}$. It should be noted that the determination of the covariance matrix $\hat{P}$requires the FIM to be invertible (nonsingular). The condition number of the FIM (*i.e.*, the ratio of the largest eigenvalue of the FIM to the smallest) is a useful indicator of the practical identifiability of the model given the available data. The higher the condition number, the more difficult the optimization is and the lower practical identifiability.

*A brief introduction to optimal experiment design for parameter estimation.*

When designing an experimental configuration with the aim of providing data to be used for model calibration, we expect the resulting data to be highly informative for allowing accurate estimation of the model parameters. The problem of defining such an experimental configuration is the realm of optimal experiment design (OED) for parameter estimation.

Since, the FIM is the core for determining the confidence intervals of the parameter estimates, classical approaches of OED for parameter estimation rely on the optimization of a scalar function of the FIM. The most popular criteria for OED is the D-optimality criterion, which maximizes the determinant of the FIM, and the E-optimality criterion, which maximizes the smallest eigenvalue of the FIM. Maximizing the determinant of the FIM implies minimizing the volume of the confidence ellipsoids for the parameters ([Walter and Pronzato, 1997](#_ENREF_6)), while maximizing the smallest eigenvalue of the FIM implies minimizing the maximum diameter of the confidence ellipsoids for the parameters.

The OED problem can be formulated mathematically as follows

 $\min\_{φ}j\left(FIM\left(p,φ\right)\right)$ (6)

where $j\left(FIM\left(p,φ\right)\right)$ is a scalar cost function of the FIM (*e.g.*, $det\left(FIM\left(p,φ\right)\right)$) and $φ$is the design vector that defines the experimental configuration (*e.g.* sampling times, initial conditions, stimuli). Since the true values of the model parameters are unknown, the OED problem is defined with a nominal parameter set $p^{0}$**,** whose values are obtained from literature or experimental data. This value can be further refined in an iterative process. It should be noted that the OED problem is constrained by experimental limitations and is only possible when there are some degrees of freedom in the procedure for data collection. Finally, the solution of the OED problem requires efficient optimization algorithms as discussed in the dedicated literature ([Walter and Pronzato, 1997](#_ENREF_6); [Balsa-Canto *et al.*, 2008](#_ENREF_1)).

References

Balsa-Canto E, Alonso AA and Banga JR 2008. Computational procedures for optimal experimental design in biological systems. IET Syst Biol 2, 163-172.

Carson ER, Cobelli C and Finkelstein L 1983. The Mathematical Modeling of Metabolic and Endocrine Systems: Model Formulation, Identification, and Validation. John Wiley & Sons, New York.

Marsili-Libelli S, Guerrizio S and Checchi N 2003. Confidence regions of estimated parameters for ecological systems. Ecological Modelling 165, 127-146.

Muñoz-Tamayo R, Laroche B, Leclerc M and Walter E 2009. IDEAS: a Parameter Identification Toolbox with Symbolic Analysis of Uncertainty and its Application to Biological Modelling. In Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France, pp. 1271-1276.

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U and Timmer J 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923-1929.

Walter E and Pronzato L 1997. Identification of Parametric Models from Experimental Data. Springer, London.