

Advanced Model Identification using Global Optimization

Theoretical introduction.

Eva Balsa-Canto and Julio R. Banga



Process Engineering Group
IIm-CSIC
Vigo-Spain
e-mails: {ebalsa,julio}@iim.csic.es

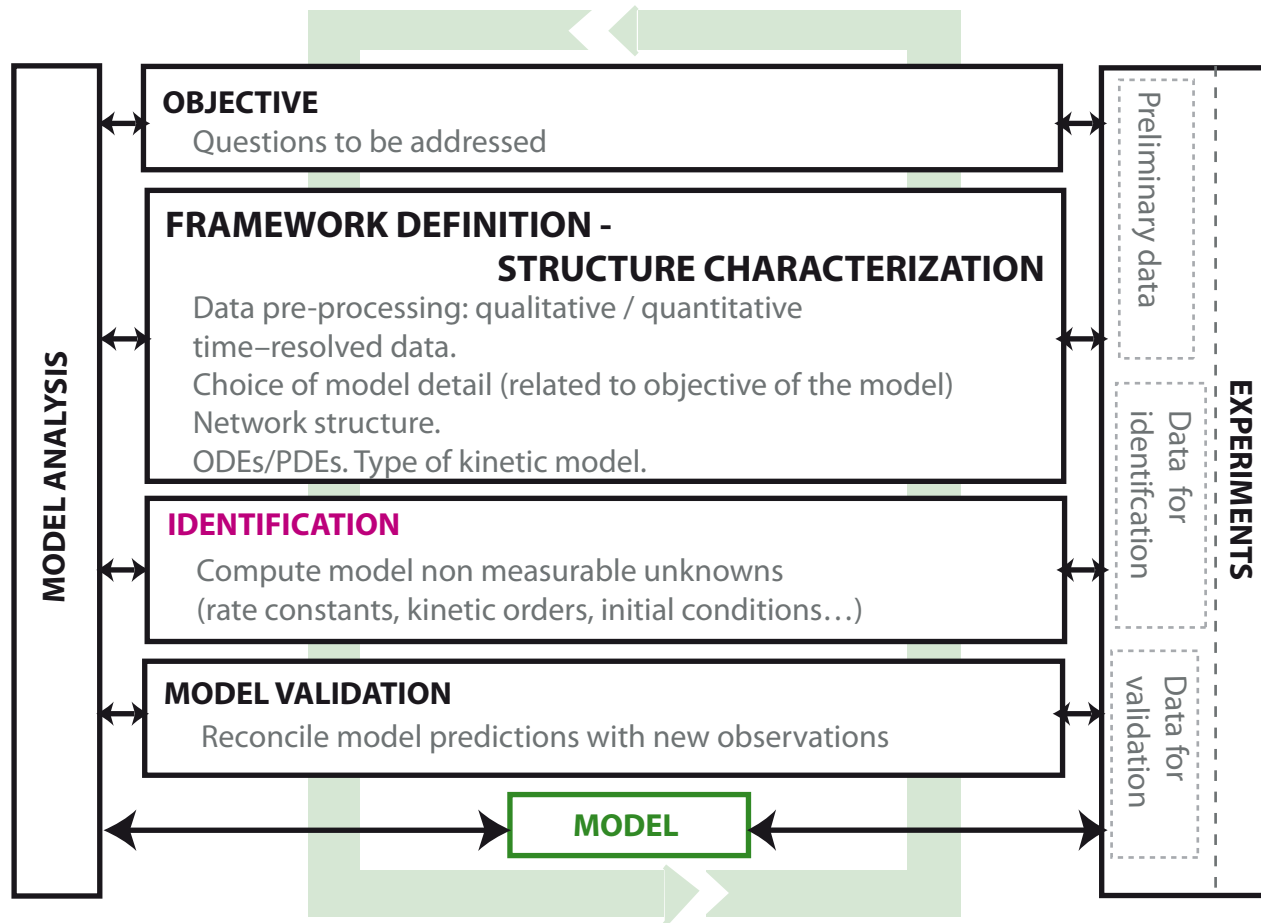


Outline

- ❑ Model (parametric) identification as part of the model building loop
 - > Structural properties: *structural identifiability*
 - > Sensitivity analysis and ranking of parameters
 - > *Model calibration*: problem formulation and numerical solution
 - > Practical difficulties: *multimodality* and lack of *practical identifiability*
 - > Model calibration with *global optimization methods*
 - > *Optimal experimental design*

- ❑ AMIGO: A matlab toolbox for Advanced Model Identification using Global Optimization

Model building loop



Experimental data ...

- i) qualitative data may help to define which components are involved, to define the 'cartoon' of interactions, to decide on the type of modelling
- ii) quantitative data is to be used for parametric identification and validation

A mathematical model ...

- i) helps to better understand the biological phenomena
- ii) enables the possibility of performing experiments 'in silico'
- iii) collects the available knowledge in a format that can be easily communicated

Devising a model involves several steps from the definition of the purpose of the model and finishing with a (first) working model.

Mathematical models in biology

General Objectives

- > To explain biological processes that result in an observed phenomena.
- > To fill the gap between the molecular level and observed patterns.
- > To predict unobserved phenomena.
- > To identify key mechanisms.
- > To identify missing components or concepts.
- > To guide experiments.
- >

Some popular examples:

- > Bacterial exponential growth models
- > Lotka-Volterra model (predator-prey)
- > Hodgkin–Huxley model (action potentials in neurons)
- > Goodwin oscillator
- > Michaelis-Menten kinetics
- > Noble (heart model)
- > ...

Framework definition- Structure characterization

- > Depend on the available data: qualitative/quantitative, time/spatial resolved data, diffusion/convection/reaction mechanisms
- > Different types of modelling frameworks: computational models such as boolean networks, mathematical models such as those in differential equations, etc.
- > Most of models under consideration: deterministic or stochastic ODEs or PDEs.

We will focus on models in terms of deterministic differential equations. 

Note that PDEs are to be transformed in ODEs by typical methods (FEM, FD, ROMs,...)

(Parametric) Identification (I)

Mathematical model $M(\mathbf{p})$

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)$$

$$\mathbf{y}^\varepsilon(t_s, \mathbf{u}, \mathbf{p}) = \mathbf{g}(\mathbf{x}(\mathbf{u}, \mathbf{p}, t_s), \mathbf{p}, t)$$

$$s = 1, \dots, n_s^{o,\varepsilon}$$

The mathematical model will consist on two essential elements:

i) The set of **ordinary differential equations** describing the system behavior

$\mathbf{x} \in X \subset \mathbb{R}^{n_x}$, vector of state variables

$\mathbf{u} \in U \subset \mathbb{R}^{n_u}$, vector of control variables, external factors or stimuli

$\mathbf{p} \in \mathcal{P} \subset \mathbb{R}^{n_p}$, vector of parameters in the model

t , time

ii) The **observation function** describing the relationship between the states in the model and the available measured quantities

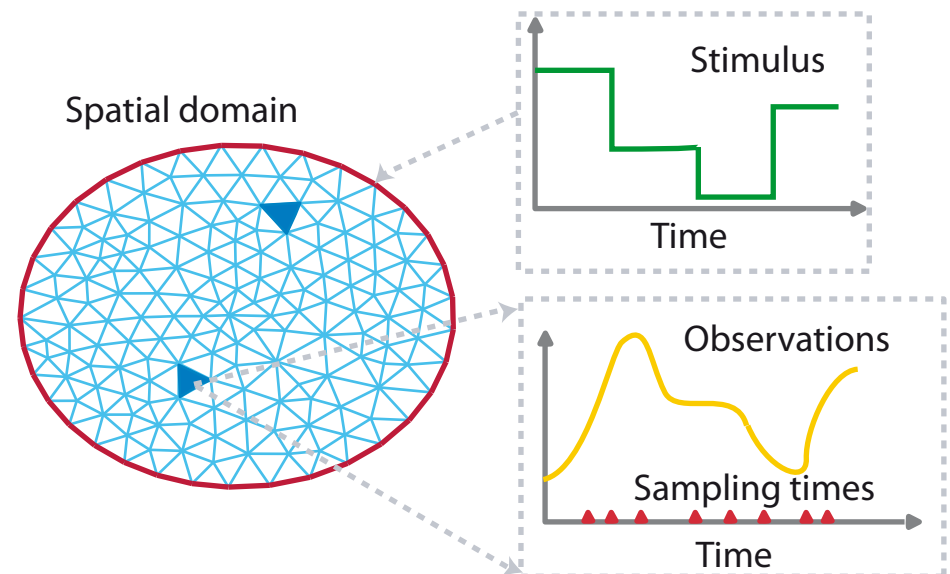
$\mathbf{y}^\varepsilon \in Y \subset \mathbb{R}^{n_o}$, n_o vectors of discrete time measurements per experiment

$n_s^{o,\varepsilon}$, number of sampling times per observable per experiment

Experimental scheme ε

The experimental scheme collects all information related to the experiment ε :

- i) Observed / measured quantities
- ii) Stimuli conditions
- iii) Spatial domain of interest (if relevant)
- iv) Locations (if spatial domain is relevant)
- v) Experiment duration
- vi) Sampling times
- vii) Experimental noise: type/ quantity (if available)



(Parametric) Identification (II)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) \quad \mathbf{M}(\mathbf{p})$$

$$\mathbf{y}^\varepsilon(t_s, \dot{\mathbf{x}} = \bar{\mathbf{f}}(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{p}}, t) \quad \bar{\mathbf{M}}(\bar{\mathbf{p}})$$

$$\bar{\mathbf{y}}^\varepsilon(t_s, \bar{\mathbf{u}}, \bar{\mathbf{p}}) = \bar{\mathbf{g}}(\bar{\mathbf{x}}(\bar{\mathbf{u}}, \bar{\mathbf{p}}, t_s), \bar{\mathbf{p}}, t)$$

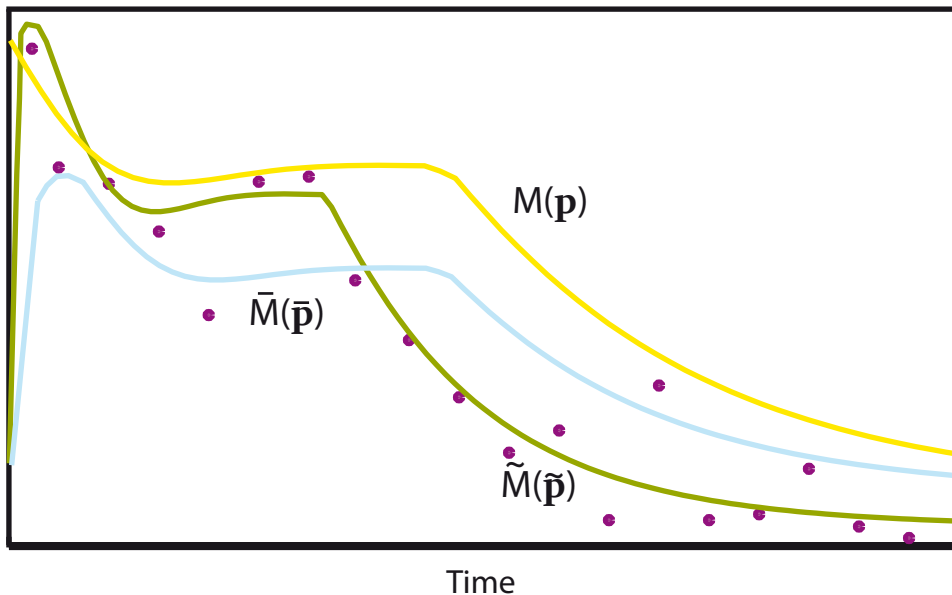


$$\dot{\tilde{\mathbf{x}}} = \tilde{\mathbf{f}}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}, \tilde{\mathbf{p}}, t) \quad \tilde{\mathbf{M}}(\tilde{\mathbf{p}})$$

$$\tilde{\mathbf{y}}^\varepsilon(t_s, \tilde{\mathbf{u}}, \tilde{\mathbf{p}}) = \tilde{\mathbf{g}}(\tilde{\mathbf{x}}(\tilde{\mathbf{u}}, \tilde{\mathbf{p}}, t_s), \tilde{\mathbf{p}}, t)$$

Given a *battery of models* which are potentially able to represent the system behavior, the objective is to find *which is the model* and the corresponding *model* unknowns that make the *mathematical equations to reproduce the experimental data*.

Model unknowns are assumed to be “constants” that may correspond to a (sub-) set of $\mathbf{p}$, regarded here as $\boldsymbol{\theta}$, together with unknown initial conditions, $\boldsymbol{\theta}_y^\varepsilon$



Model unknowns are collected in the following vector:

$$[\boldsymbol{\theta} \quad \boldsymbol{\theta}_y^\varepsilon]$$

Note that initial conditions may depend on the experiment, thus it may be necessary to estimate their values for every different experiment.

Structural properties: distinguishability and identifiability

Structural properties:

Properties of the model(s) which are independent of the parameter values

Assume that a finite number of model structures compete for the description of a given set of experimental data. It is then natural to ask whether the experimental set up makes it possible to select the best structure and estimate its parameters. This questions find answers in the idealized framework of ***distinguishability*** and ***identifiability*** studies.

Distinguishability

Given a battery of model structures, will the experimental data allow to distinguish among candidate models?

$$M(\mathbf{p}) = \tilde{M}(\tilde{\mathbf{p}}) ?$$

Methods to test distinguishability properties

- > Taylor series approach
- > Differential algebra
- > ...

When competing model structures have been proved to be distinguishable it is then necessary to device mechanisms to effectively discriminate among the candidates.

Optimal experimental design for model discrimination is intended to plan the experiments that may help to decide on the quality of the different models in the battery.

Structural properties: identifiability

Structural Identifiability

Will it be possible to uniquely estimate the model unknowns under given “ideal” (noise free and continuous) experimental conditions (observables + type of stimuli) ?

$$M(\mathbf{p})=M(\mathbf{p}^*)$$

$$\mathbf{p}=\mathbf{p}^* ?$$

A simple illustrative example

$\dot{x}_1 = p_1 x_1 x_2$ Will it be possible to uniquely estimate
 $\dot{x}_2 = p_2 u$ p_1 and p_2 by observing $y=x_1$ with
 u arbitrary but $\neq 0$?

The solution to the system is:

$$x_1 = x_{10} \exp(0.5 p_1 p_2 u t^2 + p_1 x_{20} t) = y$$

$$x_2 = (p_2/p_1) (\exp(p_1 t) - 1)$$

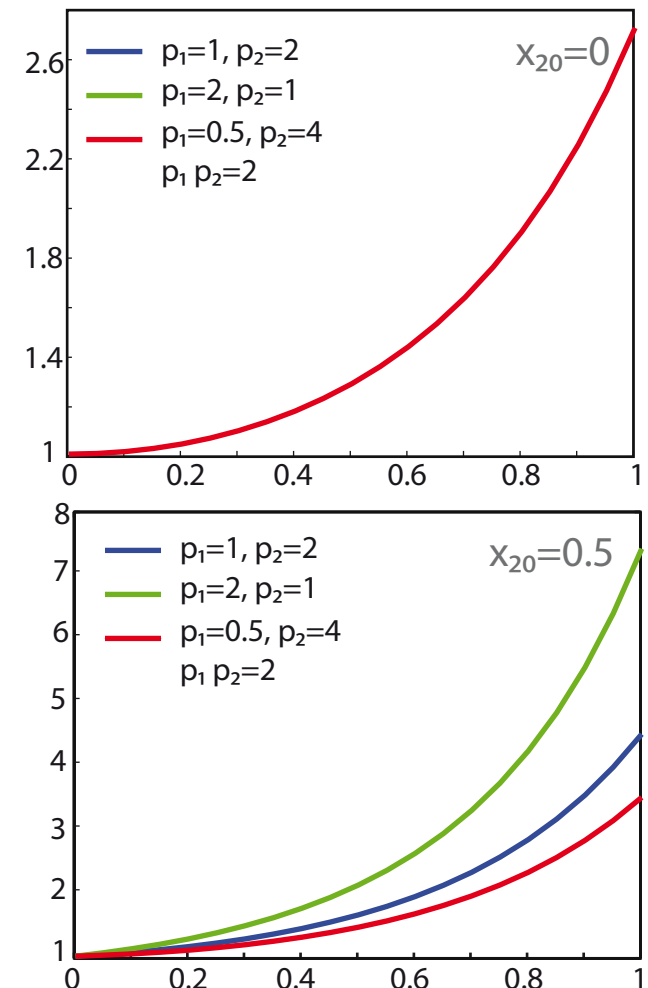
Therefore it is clear that if:

i) $x_{10} \neq 0$ & $x_{20}=0$, then the parameters are **not identifiable**!

$$y = x_{10} \exp(0.5 p_1 p_2 u t^2)$$

ii) $x_{10} \neq 0$ & $x_{20} \neq 0$, then the parameters become **identifiable**!

$$y = x_{10} \exp(0.5 p_1 p_2 u t^2 + p_1 x_{20} t)$$



Structural properties: identifiability

Test structural identifiability properties

Conditions

- The data is continuous (in time and space) and noise-free
- Inputs and observables may or not be flexible

Methods for non linear models

- > Taylor series approach
- > Similarity transformation approach
- > Differential algebra

All methods have advantages and disadvantages, probably the most popular are the Taylor method since it may be applied to any type of non-linear terms and the differential algebra approach since the analysis may, in some cases, be independent of the initial conditions.

Taylor series approach

The observation functions and their successive time derivatives are evaluated at some known time point (usually an initial condition). These derivatives are expressed in terms of the model parameters. Since the Taylor coefficients are unique and, in principle, measurable, the identifiability problem reduces to determining the number of solutions for the parameters in a set of algebraic equations which are, in general, nonlinear in the parameters.

Differential algebra based approach

The basic idea is to derive the differential polynomial formulation of the model. Compute the characteristic set with respect to the ranking $u \ll y < p \ll x < \dot{x}$. Consider the polynomials with leaders y and reduce them to monic form, i.e. with leading coefficient one. To finish with, consider the set of coefficients of the polynomials and try to rewrite the parameters as functions of those coefficients.

Recommended reading: [2] Pohjanpalo H. System identifiability based on the power series expansion of the solution. *Math. Biosci.* 41:21-33, (1978)

[3] Margaria G et al. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Math. Biosci.* 174: 1-26, (2001)

Structural properties: identifiability

REMARKS

- > Complexity of both methods increases very rapidly with the size of the model and the number of parameters. This is particularly the case when the number of observables is low as compared to the number of parameters.
- > It may be necessary to impose conditions on the stimuli (continuous, class infinite, (non) zero derivatives, etc.)
- > The use of symbolic manipulation software is suggested.

A simple illustrative example

$$\begin{aligned}\dot{x}_1 &= p_1 x_1 x_2 && \text{Will it be possible to uniquely estimate } p_1 \text{ and } p_2 \\ \dot{x}_2 &= p_2 x_1 + u && \text{by observing } y=x_1 \text{ with non-zero initial conditions?}\end{aligned}$$

Taylor series approach

Taylor coefficients

$$t_1 = \frac{dy}{dt}(t=0^+) = p_1 x_{10} x_{20}$$

$$t_2 = \frac{d^2y}{dt^2}(t=0^+) = p_1 x_{10}(u_0 + p_2 x_{10} + p_1 x_{20}^2)$$

Solution of the system of equations

$$p_1 = t_1 / (x_{10} x_{20})$$

$$p_2 = -t_1 u_0 x_{01} - t_1^2 x_{02} + t_2 x_{01} x_{02} / (t_1 x_{01}^2)$$

The system is (globally) identifiable

Differential algebra approach

Ranking

$$u < y < \dot{y} < \ddot{y} < x_1 < x_2 < \dot{x}_1 < \dot{x}_2$$

Differential polynomial formulation

$$A_1: y - x_1$$

$$A_2: \dot{x}_1 - p_1 x_1 x_2$$

$$A_3: \dot{x}_2 - p_2 x_1 - u$$

Characteristic set

$$A_1: y - x_1$$

$$A_2: y \ddot{y} - \dot{y}^2 - p_1 p_2 y^3 + p_1 y^2 u$$

$$A_3: \dot{x}_2 - p_2 y + u$$

The system is (globally) identifiable

Sensitivity analysis (I)

Local sensitivity analysis

To measure how sensitive is the model (usually the observables) to “small” changes on the parameters (around a given value of the parameters).

Absolute (local) sensitivities

$$S_p^{\varepsilon,o}(t_s^{\varepsilon,o}) = \frac{\partial y^{\varepsilon,o}}{\partial \theta_p}(t_s^{\varepsilon,o}); \quad p = 1 \dots n_\theta$$

Relative (local) sensitivities

$$S_p^{\varepsilon,o}(t_s^{\varepsilon,o}) = \frac{\Delta \theta_p}{\Delta y^{\varepsilon,o}} \frac{\partial y^{\varepsilon,o}}{\partial \theta_p}(t_s^{\varepsilon,o}); \quad p = 1 \dots n_\theta$$

Global sensitivity analysis [4]

To measure how sensitive is the model (usually the observables) to changes on the parameters over the full range of plausible values of the parameters.

- > Sampling methods: Monte Carlo, Latin Hypercube Sampling,...
- > FAST, Fourier Amplitude Sensitivity Test
- > Method of Sobol'
- > Bayesian sensitivity analysis
- > ...

What can sensitivity analysis offer for parametric identification? [5]

- > Allows to detect relevant steps in a given system
- > Allows to rank the parameters in order of importance for the observables
- > The rank may assist in fixing parameters to guarantee structural identifiability or improve practical identifiability
- > Provides information for optimal experimental design

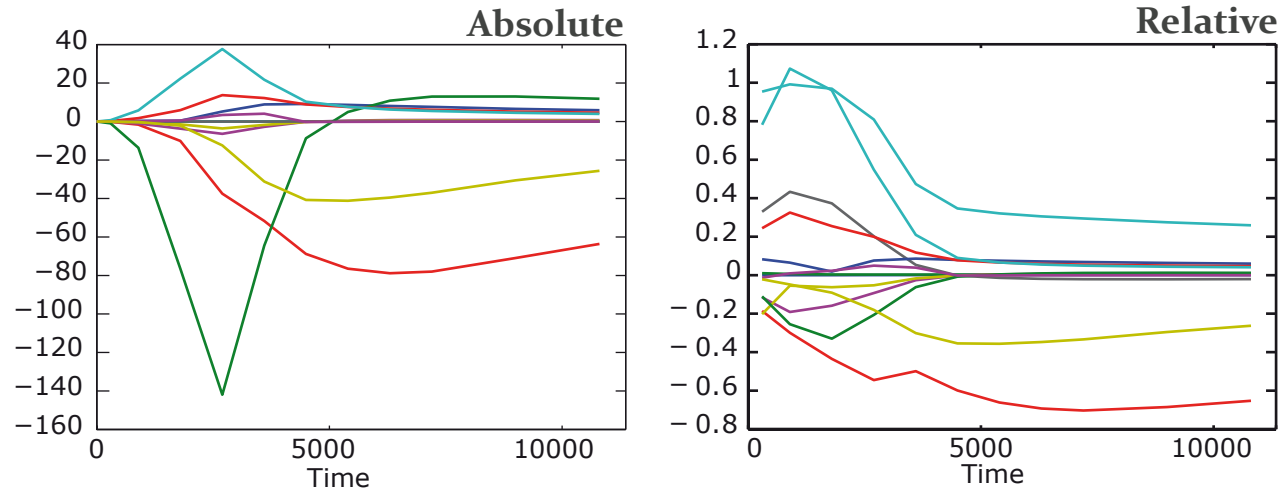
Recommended reading: [4] Saltelli A. et al. *Sensitivity analysis as an ingredient of modelling*. Statistical Science, 15(4): 377-395, (2000).

[5] Balsa-Canto et al. *An optimal identification procedure for model identification in systems biology: Applications in cell signalling*. In: Algöwer & Reuss (Eds). *Foundations of Systems Biology in Engineering*. 51-56, (2007)

Sensitivity analysis (II)

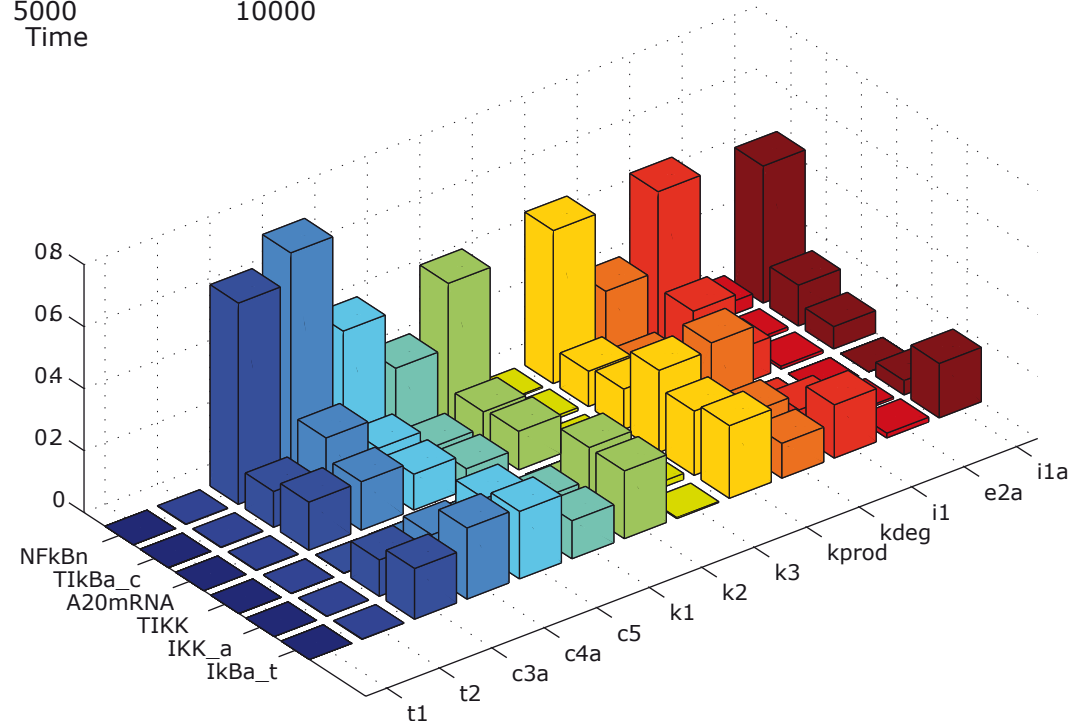
Typical sensitivity plots

Evolution with time



Temporal accumulative effect

Ex: MSQR Relative sensitivities



Ranking of parameters (I)

- > To assess individual parameters importance
- > Originally suggested several criteria to local rank of parameters [6]
- > Generalization to global rank. Relative local parametric sensitivities are computed for a number of n_{lhs} samples using the **Latin Hypercube Sampling** approach within parameter bounds [7].

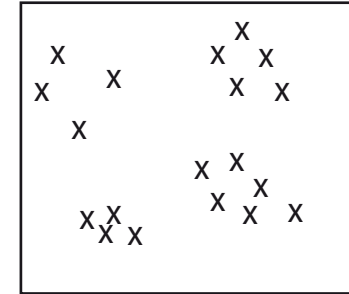
$$\delta_p^{msqr} = \frac{1}{n_{lhs} n_{\epsilon} n_o n_s} \sqrt{\sum_{lhs=1}^{n_{lhs}} \sum_{\epsilon=1}^{n_{\epsilon}} \sum_{o=1}^{n_o} \sum_{s=1}^{n_s} (s_p^{\epsilon,o}(t_s^{\epsilon,o}))^2}$$

$$\delta_p^{mabs} = \frac{1}{n_{lhs} n_{\epsilon} n_o n_s} \sum_{lhs=1}^{n_{lhs}} \sum_{\epsilon=1}^{n_{\epsilon}} \sum_{o=1}^{n_o} \sum_{s=1}^{n_s} |s_p^{\epsilon,o}(t_s^{\epsilon,o})|$$

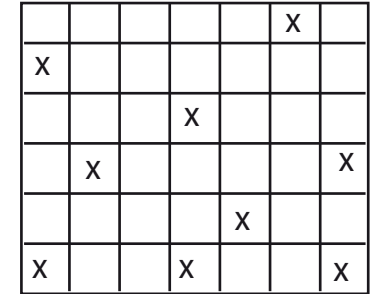
$$\delta_p^{mean} = \frac{1}{n_{lhs} n_{\epsilon} n_o n_s} \sum_{lhs=1}^{n_{lhs}} \sum_{\epsilon=1}^{n_{\epsilon}} \sum_{o=1}^{n_o} \sum_{s=1}^{n_s} s_p^{\epsilon,o}(t_s^{\epsilon,o})$$

$$\delta_p^{max} = \frac{1}{n_{lhs}} \sum_{lhs=1}^{n_{lhs}} \max_{k,i,\epsilon} s_p^{\epsilon,o}(t_s^{\epsilon,o})$$

$$\delta_p^{min} = \frac{1}{n_{lhs}} \sum_{lhs=1}^{n_{lhs}} \min_{k,i,\epsilon} s_p^{\epsilon,o}(t_s^{\epsilon,o})$$



**Conventional
Monte Carlo sampling**



**Latin Hypercube
sampling**

Ordering the parameters in f.ex. decreasing order of f.ex. δ_p^{msqr} provides the rank of parameters.

δ_p^{msqr} , δ_p^{mabs} : quantify how sensitive is the model to a given parameter considering in δ_p^{mabs} interactions among parameters.

δ_p^{max} , δ_p^{min} : indicate the presence of outliers and provide information about the sign

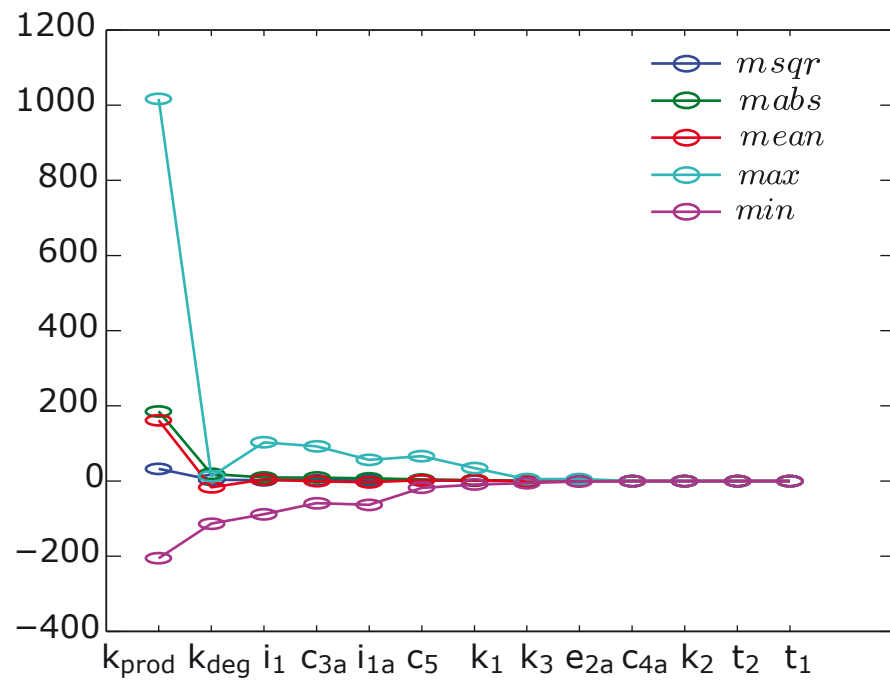
δ_p^{mean} : averaged effect

Remark that the summations will in general hide the different contributions from the different observables unless they are in the same order of magnitude. Similar analyses may be performed by observables.

Ranking of parameters (II)

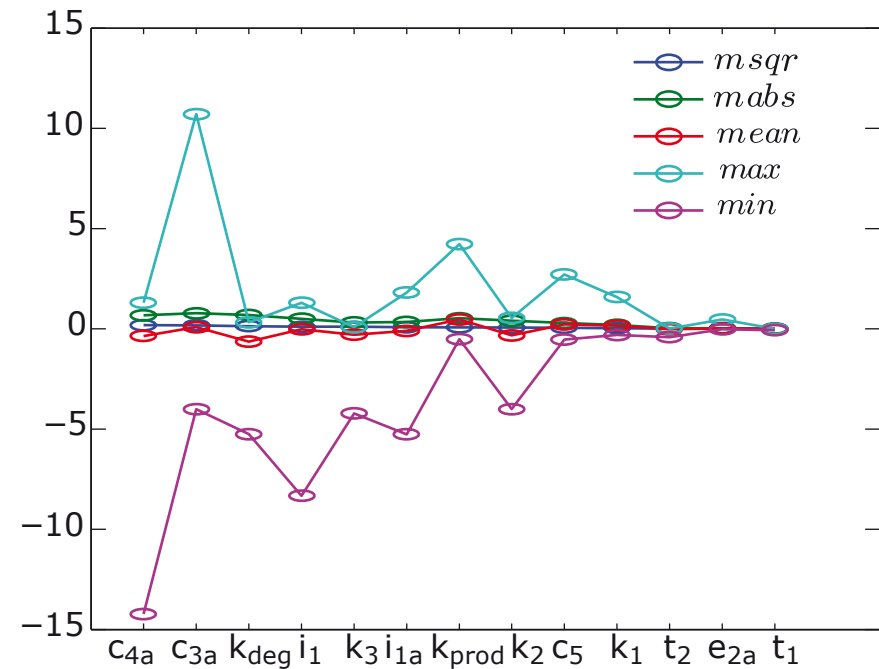
Typical ranking plots

Full local ranking of parameters



Parameters ordered by decreasing $msqr$

Relative local ranking of parameters



Model calibration. Problem formulation (I)

The model calibration problem consists on finding the model unknowns to *minimize the distance among the model predictions and the experimental data*.

Elements in the problem formulation

The model

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)$$

$$\mathbf{y}^\varepsilon(t_s, \mathbf{u}, \mathbf{p}) = \mathbf{g}(\mathbf{x}(\mathbf{u}, \mathbf{p}, t_s), \mathbf{p}, t)$$

$$s = 1, \dots, n_s^{\varepsilon, o}$$

Model unknowns

$$[\boldsymbol{\theta} \ \boldsymbol{\theta}_y^\varepsilon]$$

Experimental data

$$\mathbf{y}_m^\varepsilon(t_s) \quad s = 1, \dots, n_s^{\varepsilon, o}$$

“Distance measure”

Its definition depends on the prior information available:

- distribution in the parameters
- experimental noise

?

Scalar function to “measure” the distance among model predictions and experimental data

- > **(Generalized) Least squares and modulus estimation:** almost no prior information is considered.
- > **Maximum likelihood:** a probabilistic distribution in the noise is considered but without considering any uncertainty in the parameters.
- > **Bayesian estimation:** introduces information about a probabilistic distribution in the parameters and noise. (Not to be addressed here).

Recommended reading: [1] Walter & Pronzato. *Identification of Parametric Models from Experimental Data*. Springer, 1997.

[8] Seber & Wild. *Nonlinear regression*. Wiley series in Probability and Mathematical Statistics. Wiley. (1989).

Model calibration. Problem formulation (II)

Generalized least squares

$$J(\boldsymbol{\theta}) = \sum_{\varepsilon=1}^{n_{\varepsilon}} \sum_{o=1}^{n_o^{\varepsilon}} (\mathbf{y}^{\varepsilon,o}(\boldsymbol{\theta}) - \mathbf{y}m^{\varepsilon,o})^T \mathbf{Q}^{\varepsilon,o} (\mathbf{y}^{\varepsilon,o}(\boldsymbol{\theta}) - \mathbf{y}m^{\varepsilon,o})$$

For simplicity in the notation vector $\boldsymbol{\theta}$ will collect all unknowns to be estimated, i.e. $[\boldsymbol{\theta} \ \boldsymbol{\theta}_y^{\varepsilon}]$!

- > Quadratic cost functions are the most commonly used
- > For linear models in the parameters (LP) the best estimate can be analytically obtained
- > $\mathbf{Q}^{\varepsilon,o}$ is a nonnegative definite symmetric weighting matrix. The weighting coefficients $(\omega_s^{\varepsilon,o}, s = 1, \dots, n_s^{\varepsilon,o})$ located in the diagonal of the matrix are positive or zero and fixed *a priori*. The choice of the weights will express the relative confidence in the various experimental data and the consequent importance attached to the model performance with regard to each observable and each experimental scheme.
 - $\omega_s = 1$, assigns the same level of importance to all data
 - $\omega_s = 0$, eliminates a datum deemed not relevant
 - $\omega_s = \max(\mathbf{y}m^{\varepsilon,o})^2$, the square of the maximum experimental data for the observable o and the experiment ε , reduces the effect of having observations of different orders of magnitude
- > The solution $\boldsymbol{\theta}^* = \arg \min J(\boldsymbol{\theta})$ is regarded as the (weighted) least squares estimator or L_2 estimator

Generalized least modulus

$$J(\boldsymbol{\theta}) = \frac{1}{n_m} \sum_{\varepsilon=1}^{n_{\varepsilon}} \sum_{o=1}^{n_o^{\varepsilon}} \sum_{s=1}^{n_s^{\varepsilon,o}} \omega_s^{\varepsilon,o} \left| \mathbf{y}_s^{\varepsilon,o}(\boldsymbol{\theta}) - \mathbf{y}m_s^{\varepsilon,o} \right|$$

n_m : total number of experimental data

- > Penalize large errors less than quadratic costs
- > The choice of weighting factors is similar to the least squares case
- > The solution is regarded as the (weighted) least modulus estimator or L_1 estimator

Model calibration. Problem formulation (III)

Maximum log-likelihood

$$J(\theta) = \ln(\pi_y(y^m|\theta))$$

- > The maximum (log-)likelihood method looks for the value of the model unknowns that give the highest likelihood to the observed data.
- > This approach allows to introduce in the design of the cost function the available information on the nature of the experimental noise.

To simplify the notation the case of one experiment with one observable is now considered.

Typical situation: output-additive independent random variables

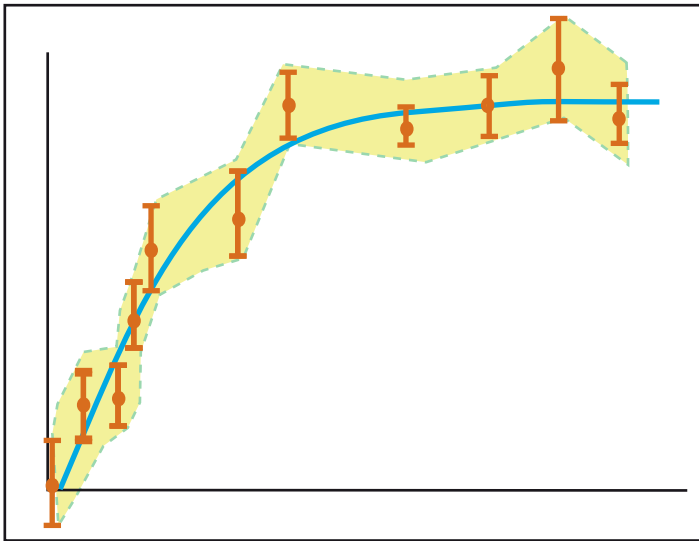
- > Assume that the observables satisfy: $y^m(t_s) = y(t_s, \theta^*) + e_s$
- > For the true value of the parameters θ^* the error satisfies: $y^m(t_s) - y(t_s, \theta^*) = e_s$
- > Since the model is deterministic and e_s belongs to a sequence of independent random variables with probability density function $\pi_{e_s}(e_s)$, the log-likelihood is then expressed as a sum of terms associated to the error for a given sampling time:

$$\ln(\pi_y(y^m|\theta)) = \sum_{s=1}^{n_s} \ln \pi_{e_s}(y^m(t_s) - y(t_s, \theta))$$

- > The type of probability density function will condition the type of function to be maximized:
a **Gaussian distribution** is usually assumed in practice.
 - **Gaussian noise with known or constant (homocedastic) variance**
 - **Gaussian noise with unknown varying variance (heterocedastic)**

Model calibration. Problem formulation (IV)

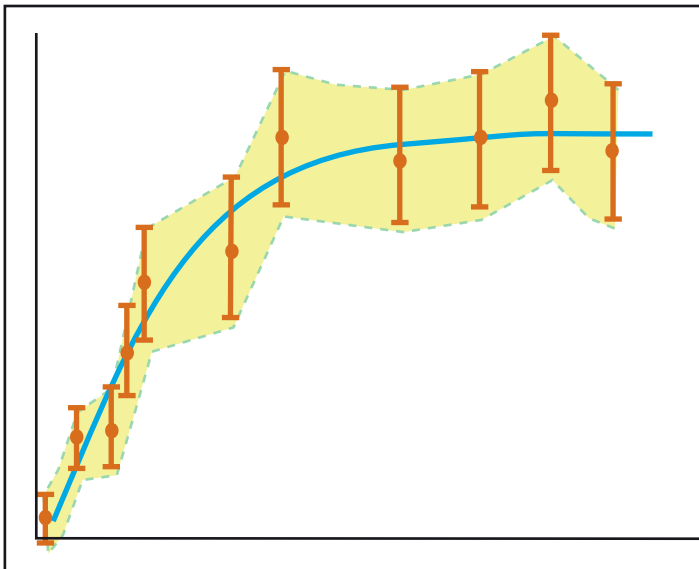
Type of experimental noise



Gaussian noise with known or constant (homocedastic) variance

$$\ln(\pi_{\mathbf{y}}(\mathbf{y}^m|\boldsymbol{\theta})) = -\frac{1}{2} \sum_{s=1}^{n_s} \left(\frac{\mathbf{y}_s(\boldsymbol{\theta}) - \mathbf{y}^m_s}{\sigma_s} \right)^2$$

- > Remark that in this case a L_2 estimator is obtained by fixing the weights to the inverse of the variance of the associated noise.
- > This estimator can be used even if the noise does not follow a Gaussian distribution, provided the values of the variances are known. It is called the **Gauss-Markow estimator**.



Gaussian noise with unknown varying variance (heterocedastic)

$$\sigma_s^2 = a |\mathbf{y}_s(\boldsymbol{\theta})|^b$$
$$J(\boldsymbol{\theta}, b) = n_s \ln \left(\frac{1}{n_s} \sum_{s=1}^{n_s} \frac{(\mathbf{y}_s(\boldsymbol{\theta}) - \mathbf{y}^m_s)^2}{|\mathbf{y}_s(\boldsymbol{\theta})|^b} \right) + b \sum_{s=1}^{n_s} \ln |\mathbf{y}_s(\boldsymbol{\theta})|^b$$

Model calibration. Numerical techniques (I)

The non-linear optimization problem

Find $[\theta \ \theta_y^\varepsilon]$ so as to minimize:

$$J(\theta)$$

subject to the system dynamics:

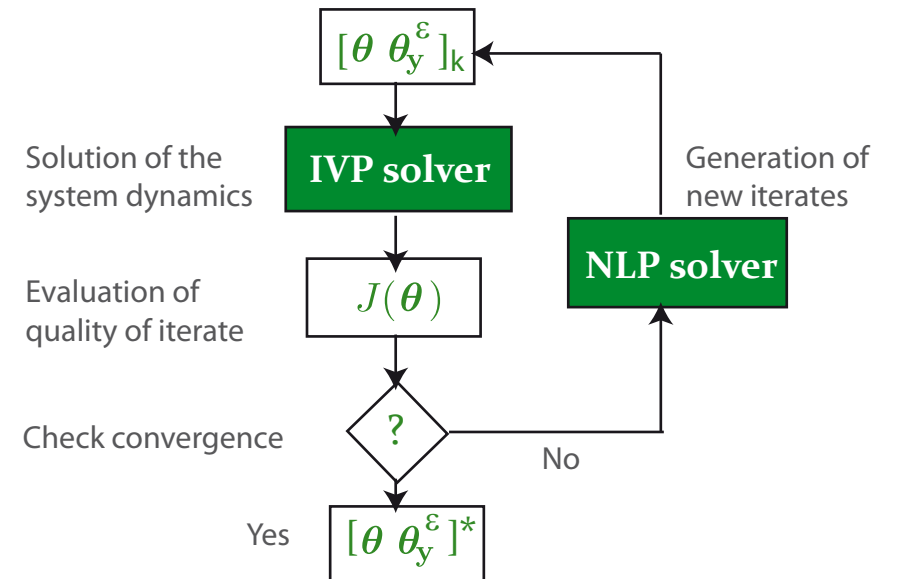
$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)$$

$$\mathbf{y}^\varepsilon(t_s, \mathbf{u}, \mathbf{p}) = \mathbf{g}(\mathbf{x}(\mathbf{u}, \mathbf{p}, t_s), \mathbf{p}, t) \quad s = 1, \dots, n_s^{\varepsilon, o}$$

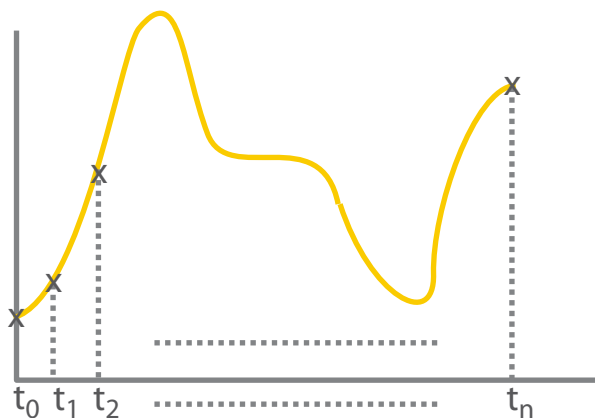
and bounds on the unknowns:

$$[\theta \ \theta_y^\varepsilon]_{\min} \leq [\theta \ \theta_y^\varepsilon] \leq [\theta \ \theta_y^\varepsilon]_{\max}$$

Numerical solution



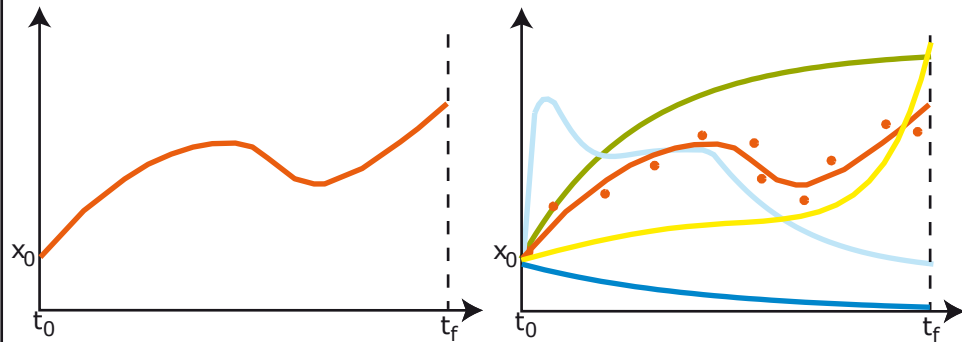
Initial value problem solver



- > The goal is to determine (approximately) the unknown function $\mathbf{x}(t)$
- > Numerical methods are based on some type of discretization of the interval $[t_0, t_f]$ and the approximation of the states by using local interpolation.
- > How the mesh is selected, the number and position of elements in the mesh, the possibilities for mesh adaptation, ect... result in a huge variety of methods suitable for stiff and non-stiff problems.
- > Runge-Kutta and BDF (backward differentiation formulae) methods are among the most popular.
- > BDF methods allow for the direct computation of parametric sensitivities

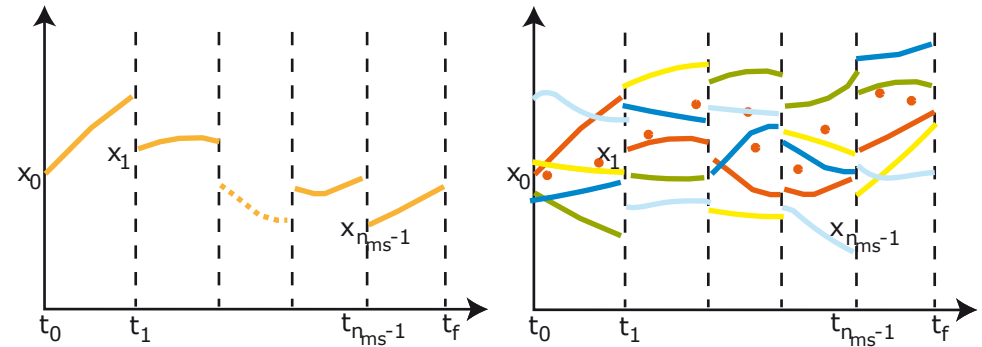
Model calibration. Numerical techniques (IX)

Single shooting



- > Solves the system dynamics for every new iterate
- > Intermediate solutions may be very far from the optimum
- > Simulation may become instable for some particular systems for some particular iterates
- > The most widely used approach in combination with local and global NLP solvers
- > Implementations with global methods overcome multimodality

Multiple shooting



- > The dynamics is partitioned in a number of shooting elements (less or equal that number of sampling times)
- > The trajectory is discontinuous and the initial conditions in the intermediate steps become decision variables
- > Constraints are necessary to force continuity in the optimum
- > Intermediate solutions are usually closer to the experimental data
- > Reduces multimodality but the cost to pay is a higher number of decision variables
- > Current implementations combined with local constrained optimization methods

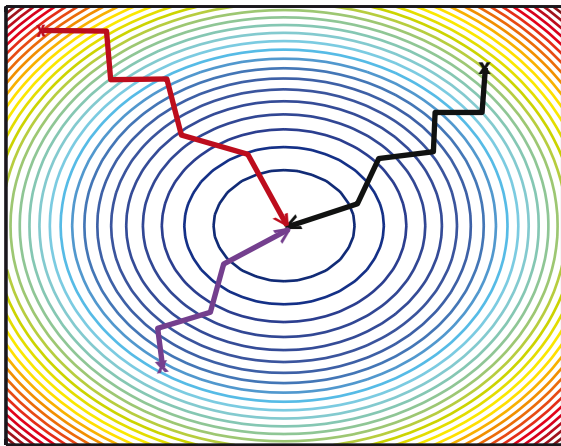
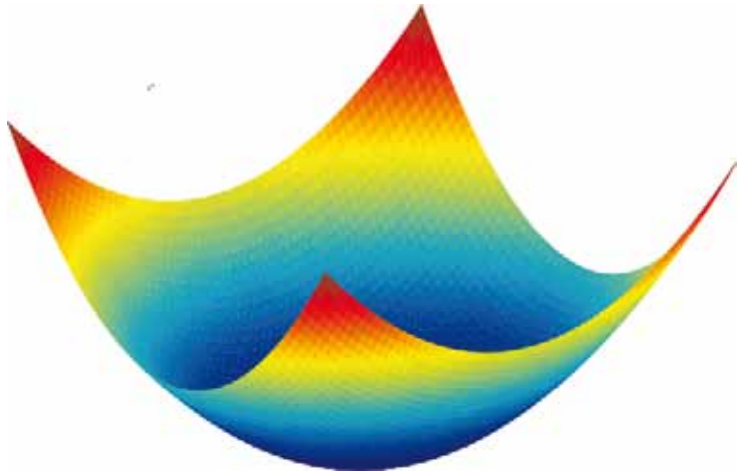
Recommended reading: [10] Peifer M, Timmer J. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *IET Systems Biology*, 1(2):78-88, (2007)

[11] Balsa-Canto et al. Hybrid optimization method with general switching strategy for parameter estimation. *BMC Systems Biology*, 2:26, (2008) doi:10.1186/1752-0509-2-26

Model calibration. Numerical techniques (III)

Non-linear optimization solvers

An illustrative 2D example.



- > Optimization methods are designed to generate a sequence of solutions that eventually converges to the minimum of the cost function
- > **Local solvers** proceed with an iterate at a time and evolve using some kind of search direction:

$$\theta_{k+1} = \theta_k + \alpha f(\theta_k)$$

- > Depending on the way the search direction is computed local solvers may be classified in direct and indirect :

Direct methods make use of the objective function values to generate new iterates. Ex: simplex, Hooke & Jeeves, Nelder & Mead, etc...

Indirect methods exploit information about the gradient (and the Hessian) so as to increase efficiency. Ex: quasi-Newton, Truncated Newton, L-BFGS, Levenberg-Marquardt, SQP, etc...

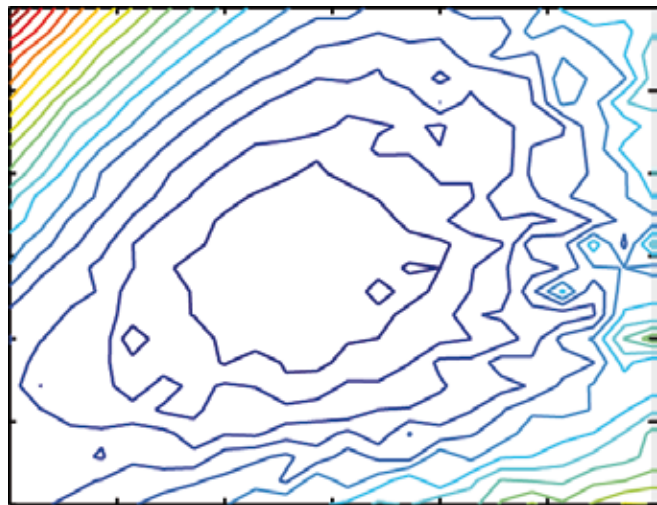
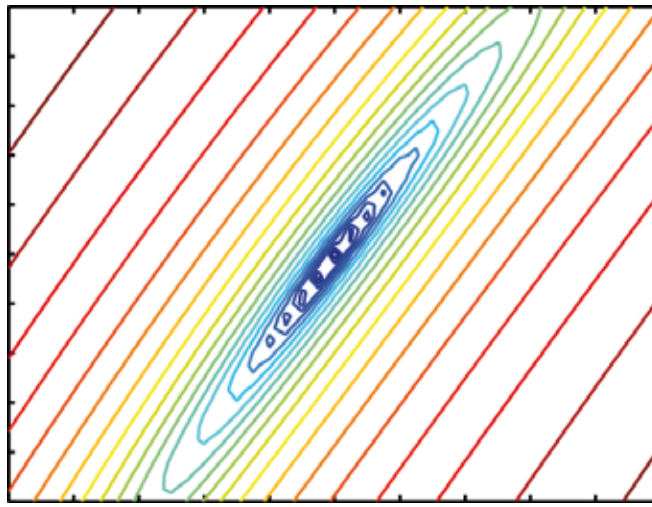
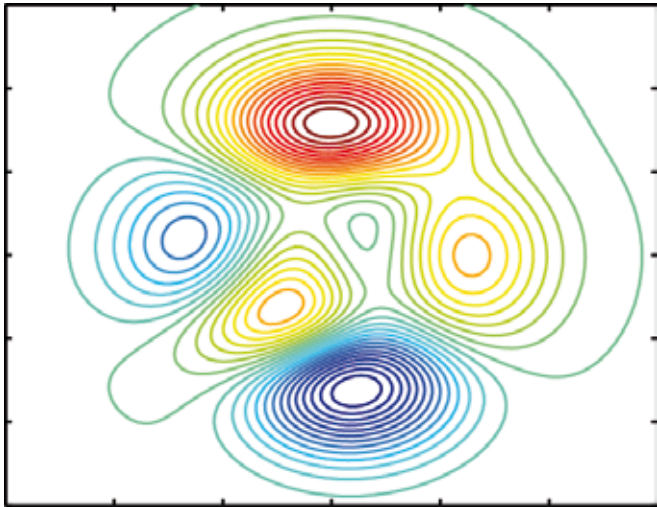
Recommended reading: [12] Nash & Sofer. *Linear and Nonlinear programming*. Mc-Graw-Hill, (1996)
[13] Fletcher. *Practical methods of Optimization*. John Wiley & Sons. 2nd Edition, (1987)

Practical difficulties (I)

- Nonlinear dynamic models (GMA models, power-law models, highly non-linear kinetics....)
- Large scale dynamics (increasing level of detail in the models, larger systems, “crosstalk” among systems...)
- Large number of model unknowns
- The order of magnitude of the unknowns may be unknown (large search space, flat areas,...)
- Presence of several suboptimal solutions (multimodality) in the optimization problem
- Quantity, Quality and Variety of data (limited number of observables and sampling times, constraints in the experimental schemes, experimental noise,...)
- Practical identifiability problems
- If the fit is bad, is that a modelling error or a “bad” solution in model calibration?

Practical difficulties (II)

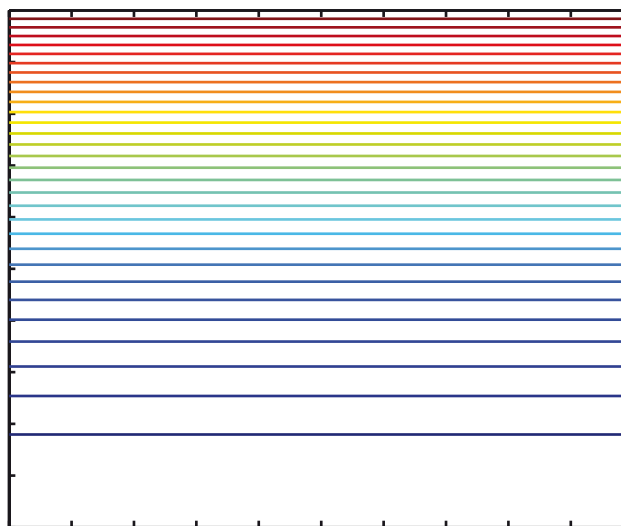
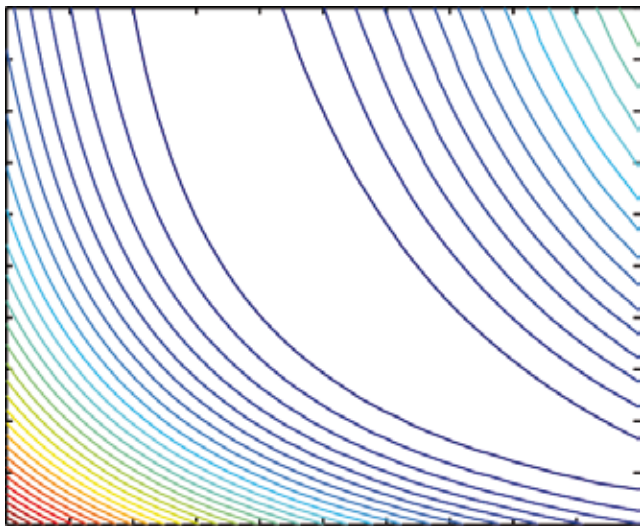
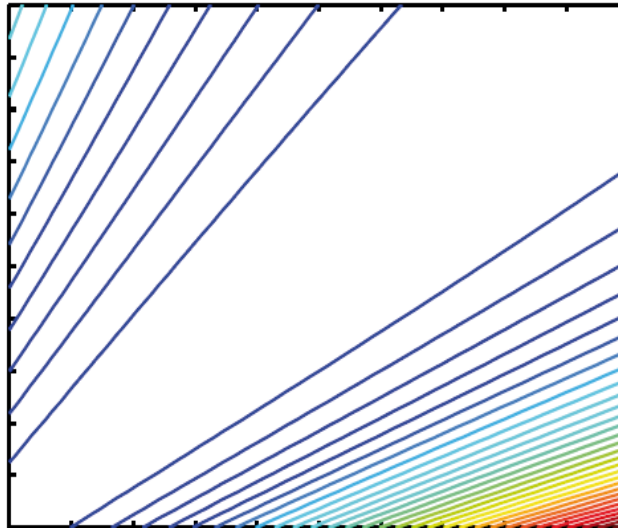
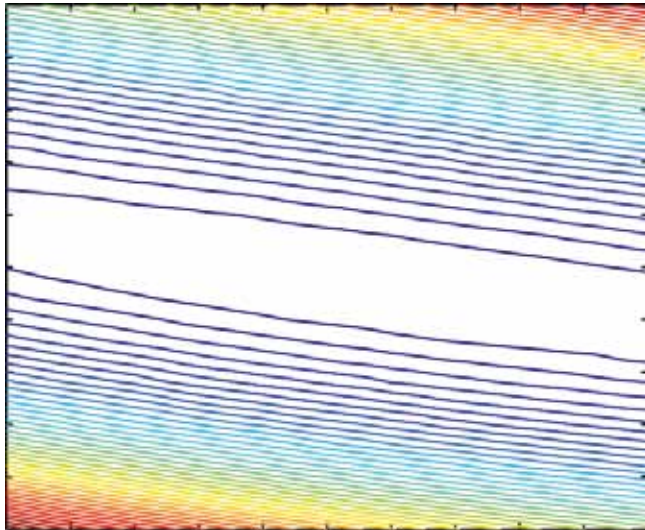
Simple 2D illustrative examples



- > Local optimization solvers tend to converge to sub-optimal solutions or to “get loss” when dealing with multimodal or noisy functions and large flat areas
- > Running a local solver from a sufficiently large number of initial guesses (multi-start) may help to detect multimodality
- > *Global optimization methods* have emerged as successful alternatives to deal with this type of difficulties

Practical difficulties (III)

Simple 2D illustrative examples



- > It is impossible to give unique solutions for the unknowns no matter the optimization solver is used.
- > The ***lack of practical identifiability*** is often related with the type of experimental scheme (type of stimuli, number of experiments, number of data, ...) and the experimental noise.
- > Contour plots and robust Monte Carlo based approaches may be used to detect lack of or poor practical identifiability.
- > ***Optimal experimental design*** has emerged as a successful alternative to improve practical identifiability

Global optimization methods (I)

Why is important not to converge to local solutions? An illustrative numerical example...

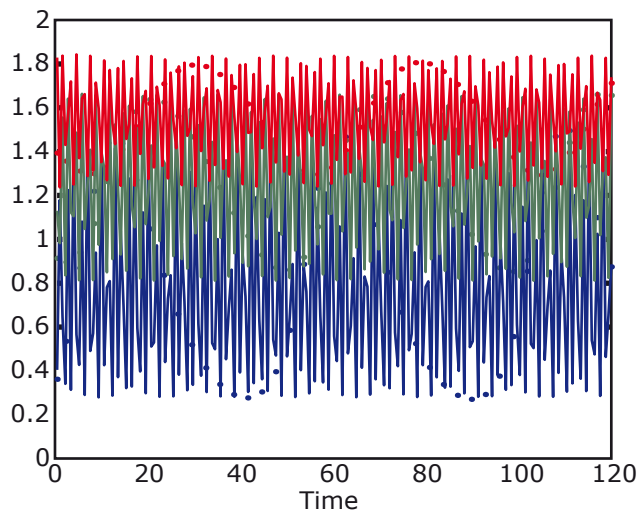
The Goodwin oscillator

$$\begin{aligned}\dot{x} &= \frac{a}{A + z^\sigma} - bx & x(0) &= 0.3617, \\ \dot{y} &= \alpha x - \beta y & y(0) &= 0.9137, \\ \dot{z} &= \gamma y - \delta z & z(0) &= 1.3934,\end{aligned}$$

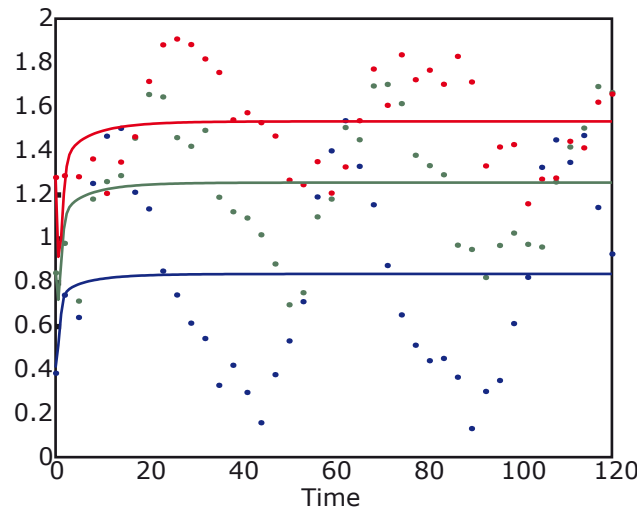
Nominal (optimal) value for the parameters

$a = 3.4884$, $A = 2.1500$, $b = 0.0969$, $\alpha = 0.0969$,
 $\beta = 0.0581$, $\gamma = 0.0969$, $\sigma = 10$, and $\delta = 0.0775$

- > Simulated experimental data has been generated by using the nominal value of the parameters and under the assumption of (1) non-experimental noise and (2) Gaussian 10% variance experimental noise
- > The objective is to estimate all model parameters by measuring all states.
- > Using local methods we arrive to solutions which are reasonable in terms of cost function, but
- > Is the model wrong or are the parameters wrong?



Non experimental noise, $J^* \sim 1 \times 10^{-2}$

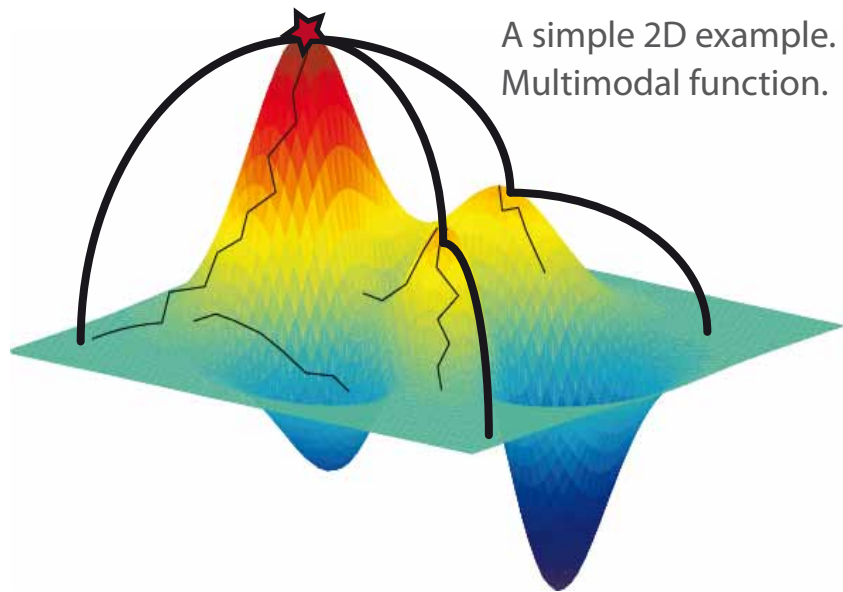


With experimental noise, $J^* \sim 10$

Global optimization methods
are required so as to convergence
to the best possible solution.

Global optimization methods (II)

Non-linear optimization solvers



Global solvers

- > Intended for the solution of multimodal problems
- > The successful methodologies combine effective mechanisms of exploration of the search space and exploitation of the previous knowledge obtained by the search.
- > Depending on how the search is performed and the information are they exploiting the alternatives may be classified in three major groups:
 - **Deterministic**
 - **Stochastic**
 - **Hybrid**

Deterministic global methods

- > Take advantage of the problem's structure, use gradient or Hessian information and may even guarantee convergence within a preselected level of accuracy.
- > Although very promising and powerful, there are still limitations to their application, mainly due to rapid increase of computational cost with the size of the considered system and the number of its parameters.

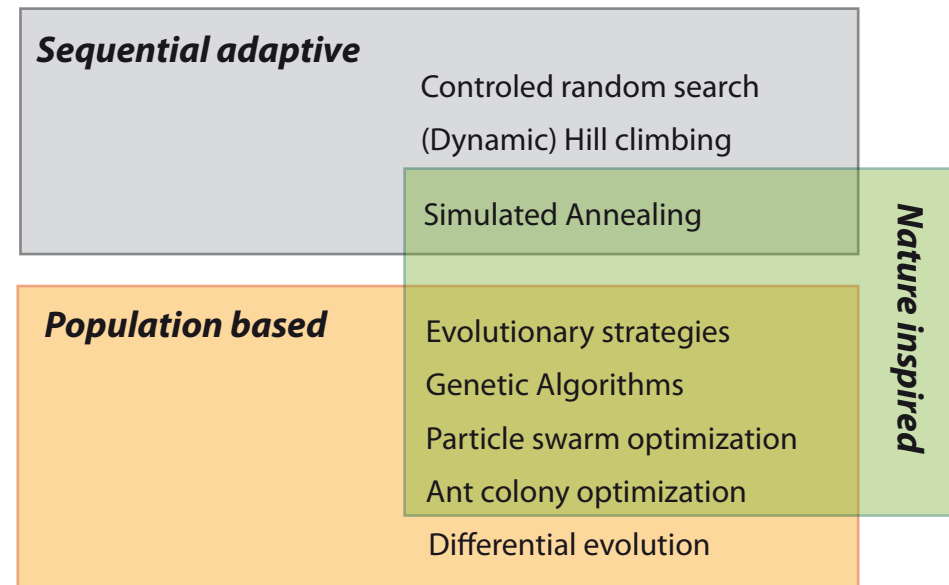
Recommended Reading:

- [14] Esposito WR, Floudas C. Global optimization for the parameter estimation of differential-algebraic systems. *Ind & Eng Chem Res*, 39:1291-1310, (2000)
- [15] Gau CY, Stadtherr MA. Reliable Nonlinear Parameter Estimation Using Interval Analysis: Error in Variable Approach. *Comp & Chem Eng*, 24:631-637, (2000)
- [16] Lin Y, Stadtherr MA. Deterministic global optimization for parameter estimation of dynamic systems. *Ind & Eng Chem Res*, 45:8438-8448, (2006)
- [17] Polisetty P et al. Identification of metabolic system parameters using global optimization methods. *Theor Biol & Med Mod*, 3:4, (2006)

Global optimization methods (III)

Stochastic global methods

- > Do not require any assumptions about the problem's structure.
- > Make use of pseudo-random sequences to determine search directions towards the global optimum.
- > The main advantage of these methods is that they rapidly arrive to the proximity of the solution although further refinements may be too costly.
- > Convergence to the global optimum can not be, in general guaranteed.
- > The number of methods has rapidly increased in last decades. The most successful approaches lie in one (or more) of the following groups:
 - Pure random search and **adaptive sequential methods**
 - Clustering methods
 - **Population based methods**
 - **Nature inspired methods**



Recommended Reading: [18] Michalewicz & Fogel. *How to solve it: Modern heuristics*. Springer (2000)
[19] Dréo et al. *Metaheuristics for hard optimization. Methods and case studies*. Springer (2006)

Global optimization methods (IV)

Some well-known examples

> **Simulated Annealing:** it is inspired on the process of slow and controlled cooling of a melted metals to obtain crystalline structures. Each step the SA algorithm replaces the current solution by a random "close" solution, chosen with a given probability that depends on the difference between the objective function and the so called temperature, that is gradually decreased during the process. It allows for "uphill" moves thus reducing the possibilities of becoming stuck at local minima. The major reported disadvantage of the method is the slow convergence as compared to other stochastic methods.

Recommended reading: [20] Kirkpatrick et al. *Optimization by simulated annealing*. *Science*, 220: 671-680, (1983).

[21] Mendes P and Kell DB. *Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation*. *Bioinformatics* 14: 869-883, (1998)

> **Evolutionary Search:** use techniques inspired by evolution such as *mutation, selection, crossover*. The algorithms start by generating a random population of candidate solutions over the search space. The best sub-set of candidates (parents) in terms of cost function (fitness) is selected and used to generate offsprings through recombination and/or mutation. The new candidates compete with old ones for their place in the next generation (survival of the fittest). Probably the most representative is **SRES (Stochastic Ranking Evolutionary Search)**[18], which has been successfully used for model calibration [19].

Recommended Reading:

[22] Runarsson T, Yao X. *Stochastic ranking for constrained evolutionary optimization*. *IEEE Trans Evol Comp*, 564:284-294, (2000)

[23] Moles et al. *Parameter estimation in biochemical pathways: a comparison of global optimization methods*. *Genome Res*, 13:2467-2474, (2003)

> **Differential Evolution [22]:** is a population based evolutionary method. It is based on the use of the so called **differential mutation**, i.e. generates a new candidate by adding to an old candidate the weighted difference between two other candidates; and a crossover between the new and old candidates.

This method has become very popular in last years, lots of different applications may be found in the developers web page: <http://www.icsi.berkeley.edu/~storn/code.html>

Recommended Reading:

[24] Storn R, Price K. *Differential Evolution – a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces*. *J Global Optim*, 11:341-359, (1997)

Global optimization methods (V)

Hybrid methods

> Combine global stochastic methods with local methods so as to enhance their strengths:

- *Global: fast approach to the vicinity of the solution*

- *Local: Fast convergence to the closest solution*

while compensating for their weaknesses:

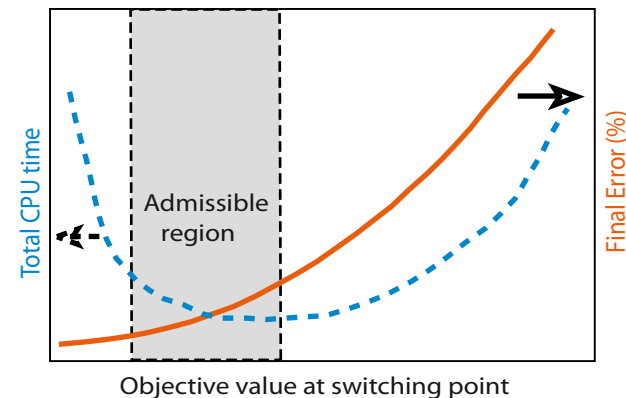
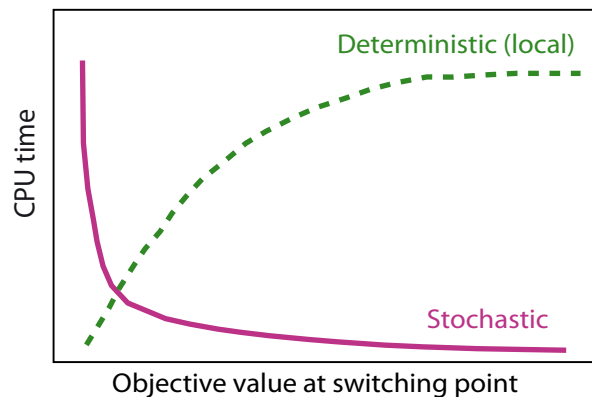
- *Global: Slow refining of the solution*

- *Local: convergence to local solutions when started far from the global*

> There are to crucial aspects to take into account to develop a hybrid [23]:

- **Design**, i.e. the selection of the methods to be combined

- **Tuning**, i.e. the selection of the switching point



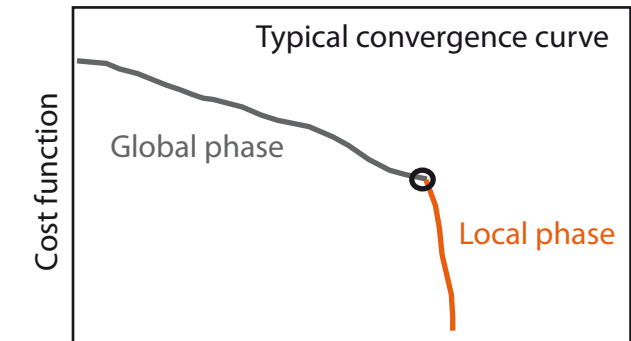
Recommended Reading: [25] Balsa-Canto E, Vassiliadis V, Banga J: Dynamic Optimization of Single- and Multi-Stage Systems Using a Hybrid Stochastic-Deterministic Method. Ind Eng Chem Res, 44(5):1514-1523, (2005)

Global optimization methods (VI)

> Types of hybrids:

- **Sequential in two phases**

- Different combinations of SRES and DE with local (first order) methods [24,25]
- “Manual” switching [24]
- Automatic switching [25]



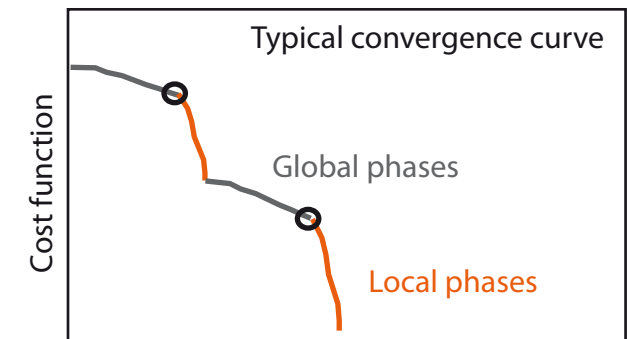
Recommended Reading:

[26] Rodriguez-Fernandez M et al. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83:248-265, (2006)

[11] Balsa-Canto et al. Hybrid optimization method with general switching strategy for parameter estimation. *BMC Systems Biology*, 2:26, (2008)
doi:10.1186/1752-0509-2-26

- **Sequential in several phases**

- Use several switching points from different regions of the search space
- Those switching points are determined following a specific heuristic
- The recently developed **Scatter Search** method [26, 27, 28]



<http://www.iim.csic.es/~gingproc/software.html>

Recommended Reading:

[27] Egea JA et al. Scatter Search for Chemical and Bio-Process Optimization. *J Glob Opt*, 37(3):481-503, (2007)

[28] Egea JA. New heuristics for global optimization of complex bioprocesses. PhD thesis. Univ. of Vigo, Spain. (2008)

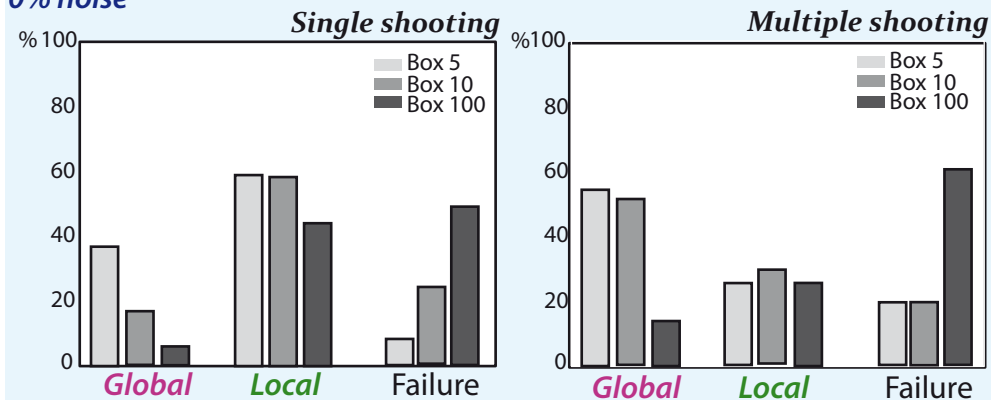
[29] Rodriguez-Fernandez M et al. Novel Metaheuristic for Parameter Estimation in Nonlinear Dynamic Biological Systems. *BMC Bioinformatics*, 7:483, (2006)

Global optimization methods (VII)

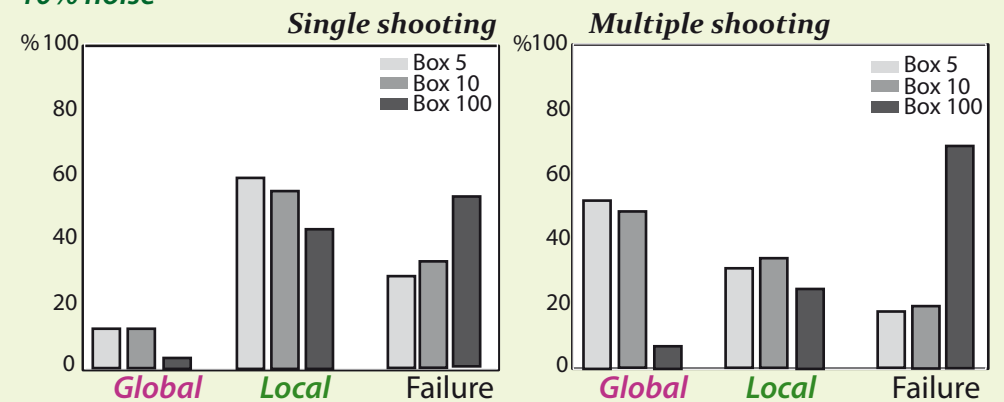
Back to the Goodwin oscillator problem

Multi-start of a Local solver

0% noise

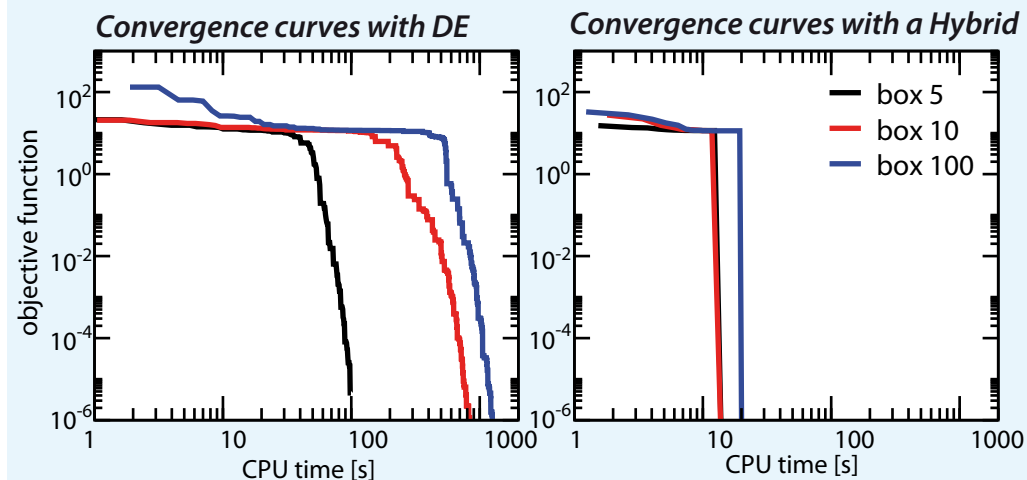


10% noise

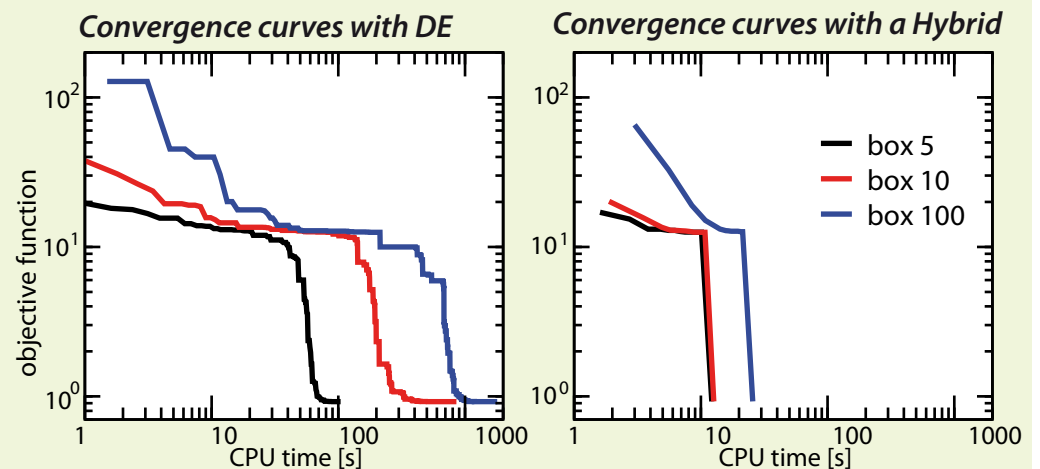


Solution with Global solvers

0% noise

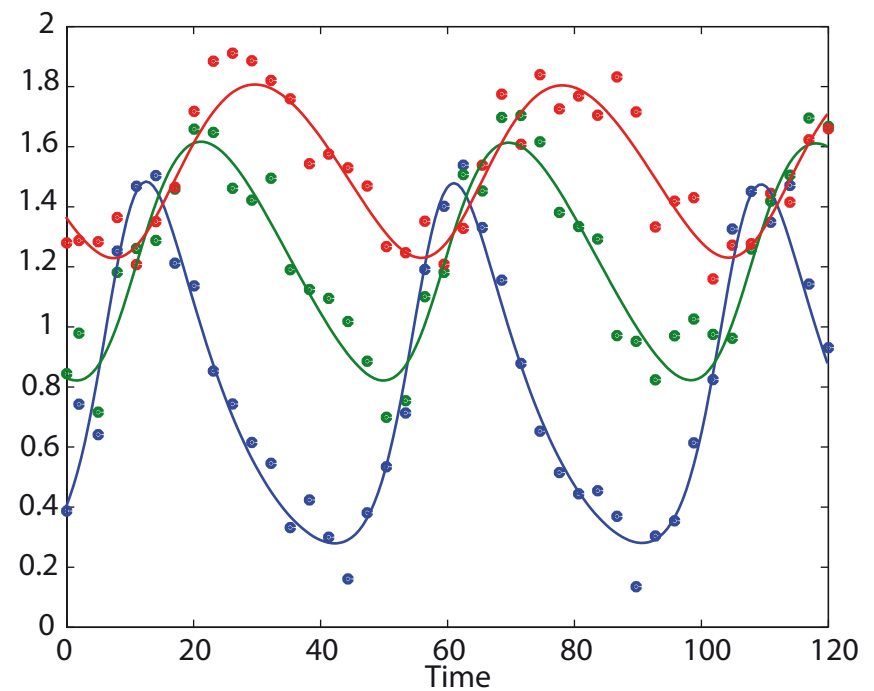
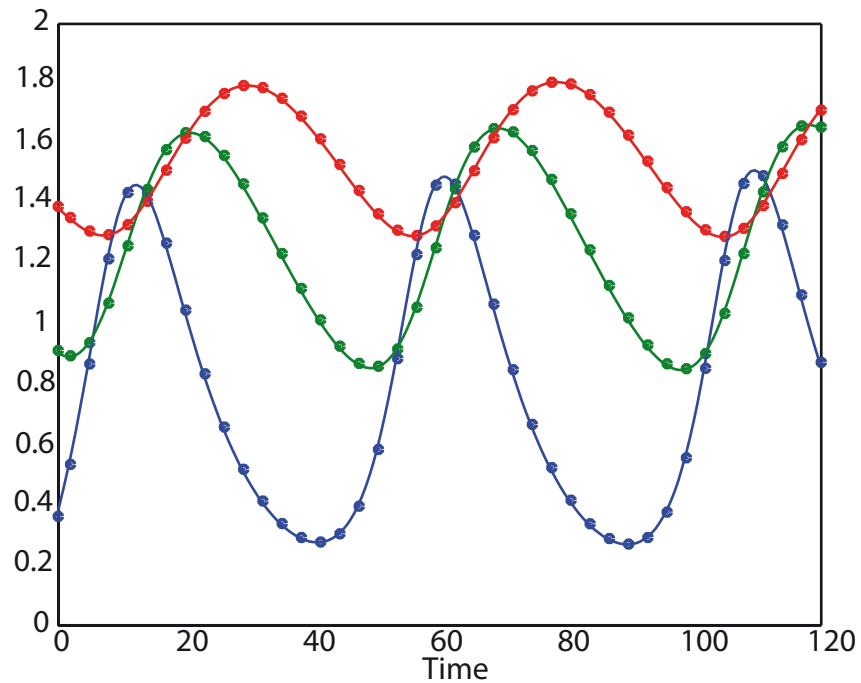


10% noise



Global optimization methods (VIII)

Back to the **Goodwin oscillator problem**, the global solution



Practical identifiability analysis (I)

Practical Identifiability

Will it be possible to uniquely estimate the model unknowns under given experimental conditions (observables + type of stimuli + experimental noise + limited number of data) ?

$$M(\theta) = M(\theta^*) \quad \theta = \theta^* ?$$

Structural vs Practical identifiability

> The questions are pretty similar, however there are some differences which make the analysis and the consequences of the lack of identifiability completely different:

- ❑ Structural analysis is performed under ideal noise free experimental conditions whereas for the practical identifiability experimental noise is crucial.
- ❑ Lack of structural identifiability will be in most of the cases impossible to solve. Only in some cases it will be possible to compute/obtain an experimental scheme that may eliminate the structural identifiability problems. In most examples one should consider reformulating the model or aggregating terms or parameters.
- ❑ Lack of practical identifiability will be in general terms solvable, provided the experimental constraints allow for designing sufficiently rich experiments and adequate global optimization solvers are employed to deal with the presence of suboptimal solutions.
- ❑ Structural identifiability analysis is by far more complicated, complex symbolic manipulations are required and this might make a full analysis impossible for highly nonlinear large scale models. Practical identifiability analysis may be computationally intensive but it is extremely helpful to assess parameter estimates reliability and to compare possible experimental designs.

Practical identifiability analysis (II)

- > To evaluate the practical identifiability analysis one may use *cost contours* and/or *confidence regions*.

Cost contours in the unknowns space

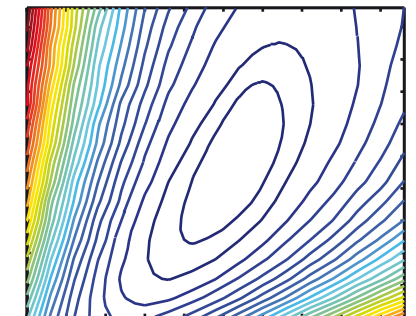
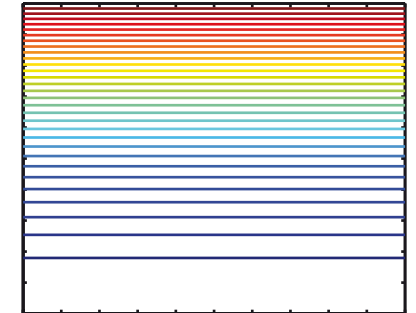
The cost contour at a certain level j^* of cost function $J(\boldsymbol{\theta})$ corresponds to a hypersurface satisfying $J(\boldsymbol{\theta}) = j^*$.

If the Hessian of the cost is positive-definite in the vicinity of the minimum, the cost contours may then be approximated by n_θ -dimensional hyper-ellipsoids (with n_θ the number of unknowns).

Drawing the cost contours consists of evaluating the cost function in a grid covering the “vicinity” of the optimum and then use a routine to determine the iso-cost surfaces.

Visual inspection of 2D projections of the cost contours may provide information about the practical identifiability: for example, contours extend to infinity when some unknowns are not identifiable; contours may also reveal the presence of several suboptimal solutions, thus the parameters being “difficult” to identify.

Cost contours may also be used to quantify confidence regions (see [1,8]). This approach will not be considered here.



2D examples, several contours

Practical identifiability analysis (III)

Crammèr-Rao inequality. Confidence intervals.

> The Crammèr-Rao inequality provides a lower bound on the covariance matrix of the estimators:

$$Cv \geq F^{-1}(\theta^*)$$

where F is the **Fisher Information Matrix** given by:

$$F(\theta) = E_{y^m|\theta} \left\{ \left[\frac{\partial}{\partial \theta} \ln \pi_y(y^m|\theta) \right] \left[\frac{\partial}{\partial \theta} \ln \pi_y(y^m|\theta) \right]^T \right\}$$

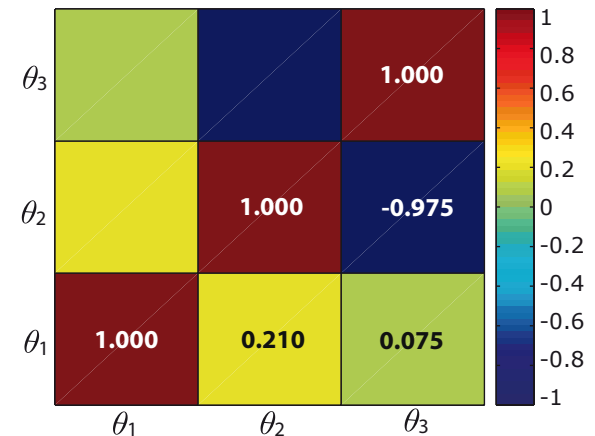
> The **confidence intervals** for the parameters can be then computed as: $\theta_i^* \pm 2Cv_{ii}$

> The **correlation** between pairs of unknowns $Cr_{ij} = \frac{Cv_{ij}}{\sqrt{Cv_{ii}Cv_{jj}}}$, in such a way that two unknowns

(θ_i, θ_j) are highly correlated if $Cr_{ij}=+1$ or -1 and fully uncorrelated if $Cr_{ij}=0$.

> Remark that the Crammèr-Rao inequality is valid under the following assumptions:

- the set of all data vectors with probability density $\pi_y(y^m|\theta)$ does not depend on θ
- the gradient of $\ln \pi_y(y^m|\theta)$ with respect to the parameters is integrable
- the Fisher Information Matrix exists and is invertible



Illustrative example of correlation matrix.

Recommended reading: [1] Walter & Pronzato. Identification of Parametric Models from Experimental Data. Springer, 1997.

[30] Ljung. System identification: Theory for the user. Prentice Hall, New Jersey, 1999.

Practical identifiability analysis (IV)

Monte-Carlo based methods

Taking into account that the repetition of identical experiments will in general not yield the same results due to the experimental noise and possible perturbations, Monte-Carlo based methods aim to determine statistical characteristics of the population of estimates yielded by the set of all possible realizations of the experimental data.

Since it is impossible to reproduce the experiments a sufficiently large number of times the idea is to generate pseudo-measurements for a given experimental scheme under given experimental noise conditions and to estimate the corresponding model unknowns.

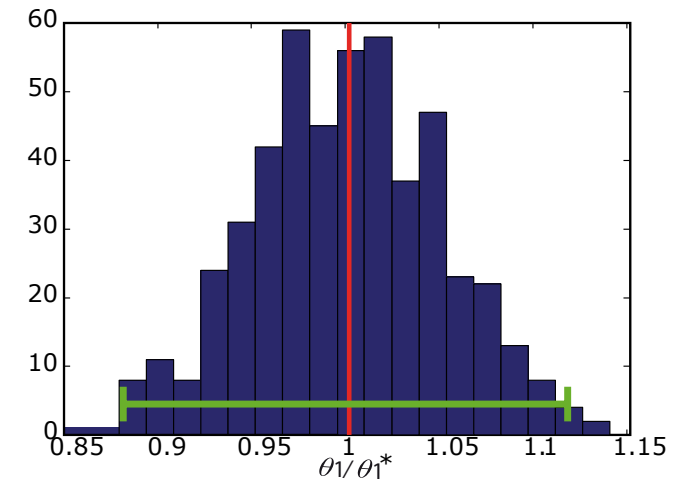
The numerical information about the confidence region is obtained through the manipulation of the resulting matrix of solutions.

Quantitative results on confidence pseudo-volume and eccentricity and visual inspection by pairs of parameters helps to detect correlation among parameters.

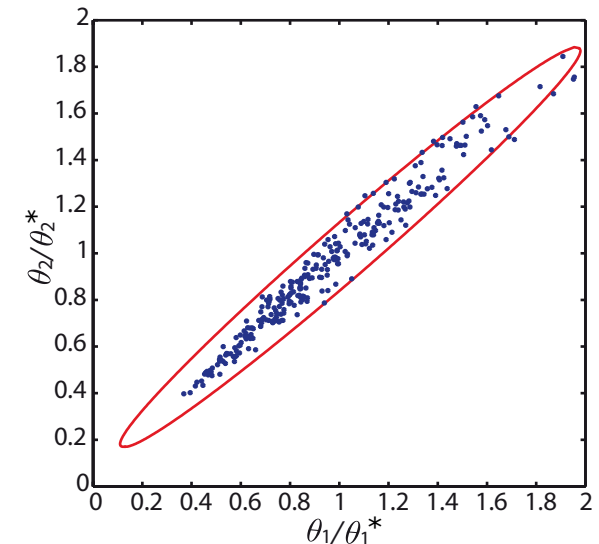
Recommended reading:

[31] Joshi M, et al. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Eng.*,8:447-455, (2006).

[32] Balsa-Canto E, et al. Computational design of optimal dynamic experiments in systems biology: a case study in cell signaling. In M. Canovas, J.L. Iborra, & A. Manjon, Eds, *Understanding and Exploiting Systems Biology in Bioprocesses and Biomedicine*, pages 103-117. Fundacion CajaMurcia, (2006).



Illustrative example of Monte Carlo based confidence interval.



"Cloud" of solutions by pairs of parameters

Optimal experimental design (I)

Planning experiments for modelling in systems biology is quite complex and several experimental phases should be distinguished. Let's consider, for example, the case of modelling biological networks (metabolic pathways, signalling cascades,...):

1st phase, the experiments are devoted to uncover biological details within networks, which elements (proteins, metabolites, genes) are involved, the interactions among such components and probably new complexes. This may help to elucidate the network structure.

To develop mathematical models of such networks the level of detail should be constrained by the question under consideration and the availability of quantitative data. Note that models with too many degrees of freedom may pose serious risks for model calibration and ulterior model predictive capabilities. Therefore once the model and the observation function have been defined, the possibility of uniquely estimating the parameters, structural identifiability, should be assessed.

2nd phase, the objective is to gather the necessary quantitative data for model calibration.

3rd phase, it to gather new data for model validation

Optimal Experimental Design (OED) : the determination of the (dynamic) scheme of measurements that generates the richest information for the estimation of parameters with the greatest precision and/or uncorrelation.

It is therefore devoted to the second experimental phase and it is model dependent.

Recommended reading:

[33] Balsa-Canto et al. *Computational Procedures for Optimal Experimental Design in Biological Systems*. IET Systems Biology, 2008. In press.

Optimal experimental design (II)

> There are many design issues prior to performing experiments:

- what to measure?
- which stimuli should be manipulated? and how?
- initial state of the system under consideration?
- when to measure (sampling times)?
- the number of experiments?

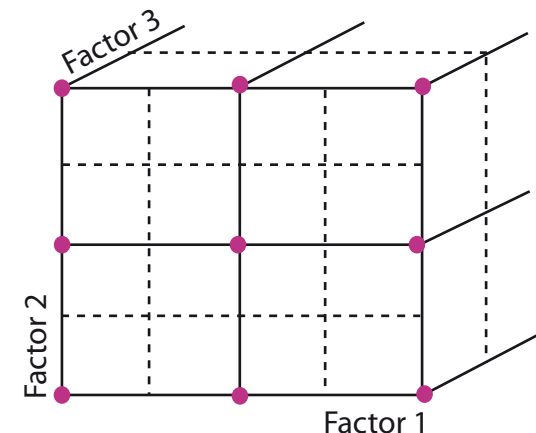
> Because of the human and economic costs associated to performing these experiments, sample sizes are usually restricted, thus efficient use of available resources is critical.

The objective of *optimal experimental design* for model calibration is to calculate the *best scheme of measurements* in order to *maximize the richness (quantity and quality) of the information* provided by the experiments while *minimizing the experimental burden*.

The simplest experimental design: Multilevel Factorial plan

Consist of testing the system responses to different constant values (levels) for the factors influencing the system so as to select or screen out the few important main effects.

- > Number of experiments increases very rapidly with the number of factors and levels
- > May provide limited information about the system dynamics



Optimal experimental design (III)

> The design of optimal experiments for model calibration requires:

- the model of the system including the observation functions
- the definition of a measure of the quality of the experiment for model calibration: *scalar cost function*
- the definition of the degrees of freedom (stimuli, sampling times...): *decision variables/functions*
- the definition of the experimental limitations: *algebraic constraints*
- an optimization procedure to solve the problem: *dynamic optimization method*

Scalar cost function (I)

For the purpose of model calibration the cost function should be related to the uncertainty on the parameters. In general, as stated before, it is assumed that the inverse of the Fisher Information Matrix ($\mathbf{F}$) provides a measure of this uncertainty. Therefore the cost function may be defined as a scalar measure of the $\mathbf{F}$ as follows:

$$J_{OED} = \phi(\mathbf{F})$$

The general formulation of the $\mathbf{F}$ (in page 35) allows to consider different types of probability density functions. Note however that the most widely used in the literature corresponds to a Gaussian function with known and fixed variance (homocedastic case):

$$\mathbf{F} = \sum_{\varepsilon=1}^{n_{\varepsilon}} \sum_{o=1}^{n_o} \sum_{s=1}^{n_s^{\varepsilon,o}} \nabla_{\theta} y^{\varepsilon,o}(t_s | \theta^*) \mathbf{Q}^{\varepsilon,o}(t_s) \nabla_{\theta} y^{\varepsilon,o}(t_s | \theta^*)^T$$

The generalization to the heterocedastic case may be found in [Appendix B, 32].

Remark that the design will be local in the sense that the $\mathbf{F}$ is computed for a given value of the parameters.

Recommended reading:

[34] García MR. Identification and Real Time Optimisation in the Food Processing and Biotechnology Industries. PhD Thesis. Univ. of Vigo. Spain. 2008.
(Appendix

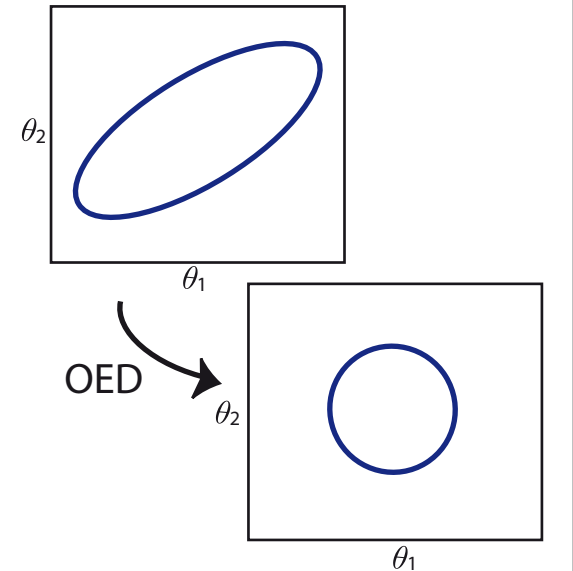
Optimal experimental design (IV)

Scalar cost function (II)

The selection of the scalar cost ϕ is often related to shape and volume of the quadric associated to the $\mathbf{F}$ or to its inverse. Note that for the non-singular cases the quadric associated to the inverse of the $\mathbf{F}$ will correspond to a hyper-ellipsoid (as shown in the figures for the 2D case), the asymptotic confidence hyper-ellipsoid.

The objective will be to make this hyper-ellipsoid as small and spherical as possible. Several criteria have been defined with this purpose, possibly the most popular are:

- **D optimality**: Maximizes the determinant of the $\mathbf{F}$
- **E optimality**: Maximizes the minimum eigenvalue of the $\mathbf{F}$



Decision variables/functions & associated constraints

Associated to the definition of the experimental scheme (page 5):

- Observed / measured quantities: often the maximum the number of allowed observables the better (structural identifiability should be guaranteed).
- Stimuli conditions: may be the stimuli changed with time and how? pulse-wise, step-wise, the maximum amount of stimuli, etc...?
- Experiment(s) duration: maximum and minimum duration for the experiments?
- Sampling times: how many, minimum distance between sampling times?
- Experimental noise: type/ quantity (if available)
- If spatial domain plays a role, number and allowed locations of sensors?

All this information will be kept in the following constraints:

$$\mathbf{u}^L < \mathbf{u}(t) < \mathbf{u}^U \quad \mathbf{v}^L < \mathbf{v} < \mathbf{v}^U$$

Optimal experimental design (V)

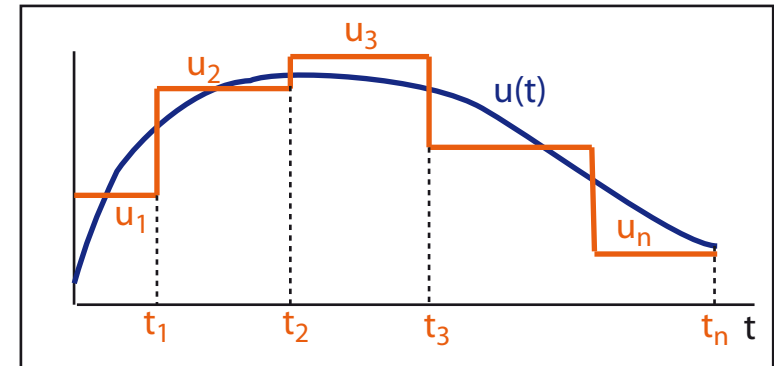
(Dynamic) optimization problem and its solution

- > The OED problem may be formulated as a general dynamic optimization problem as follows:
Calculate the experimental conditions: $\mathbf{u}(t)$ and $\mathbf{v}$, so as to optimize a scalar measure of the Fisher Information Matrix:

$$J_{OED} = \phi(\mathbf{F})$$

subject to:

- The system dynamics and the observation functions
 - The algebraic constraints related to experimental limitations
- > The **control vector parameterization method** may be used to transform the original dynamic optimization problem into a non-linear programming problem.
 - > The resultant NLP may be solved by using suitable optimization methods.
 - > Note that most of the OED problems result to be multimodal thus the use of **global optimization methods** is recommended.



Recommended reading:

- [31] Balsa-Canto et al. Computational Procedures for Optimal Experimental Design in Biological Systems. IET Systems Biology, 2008. In press.
- [33] Banga JR et al. Computation of optimal identification experiments for nonlinear dynamic process models: an stochastic global optimization approach. Ind & Eng Chem Res. 41:2425-2430, (2002).

Optimal experimental design (VI)

An illustrative example

STAT5 signalling module

Model

$$\dot{x}_1 = -k_1 x_1 \text{Epo}R_A(t) + k_2 x_3(t - \tau)$$

$$\dot{x}_2 = -x_2^2 + k_1 x_1 \text{Epo}R_A(t)$$

$$\dot{x}_3 = -k_2 x_3 + x_2^2$$

$$\dot{x}_4 = -k_2 x_3(t - \tau) + k_2 x_3$$

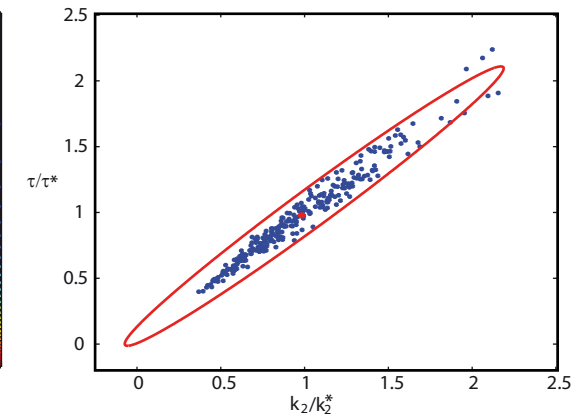
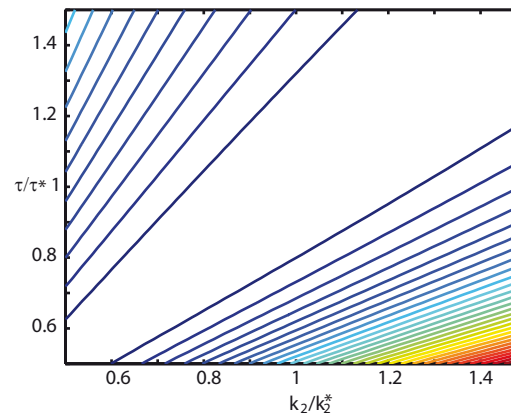
$$\mathbf{x}_0 = [3.71 \ 0 \ 0 \ 0]^T$$

$$y_1 = \sigma_1 (x_2 + x_3)$$

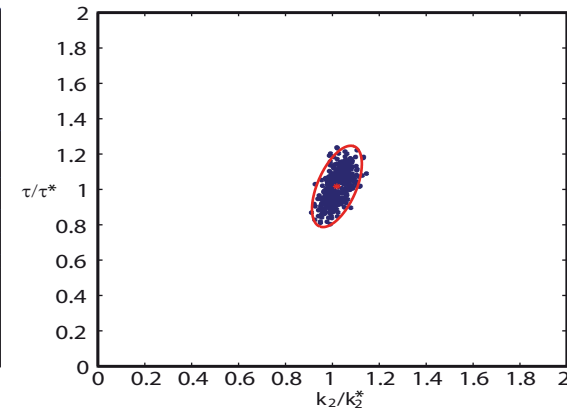
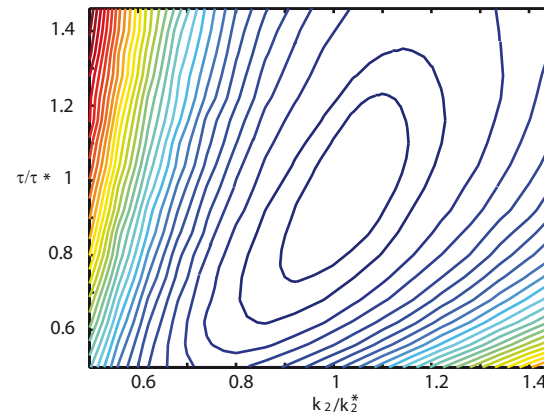
$$y_2 = \sigma_2 (x_1 + x_2 + x_3)$$

Experimental constraints

1. No more than 4 stimulus steps within the bounds $0 < \text{Epo}R_A(t) < 1$ per experiment.
2. No more than 20 sampling times per experiment.
3. The overall experimental scheme duration is bounded between 40 min and 60 min per experiment.



Practical identifiability analysis of $[k_2, \tau]$ for a typical experiment with sustained stimulus and 20 equidistant sampling times. a) Least squares contour plot; b) Monte Carlo scatter plot. Figures reveal not only the high correlation between the parameters but the expected maximum uncertainty in the parameters which corresponds to more than the 80%.



Practical identifiability analysis of $[k_2, \tau]$ for an optimal dynamic pulse-wise experiment. a) Least squares contours which are almost elliptical; b) Monte Carlo scatter plot. The optimal dynamic experiment tends to both uncorrelate the parameters and reduce the size of the confidence regions as compared to sustained stimulus experiments. A maximum expected uncertainty of the 24.2% corresponds to the estimation of τ .

Recommended reading:

[33] Balsa-Canto E et al. Computational Procedures for Optimal Experimental Design in Biological Systems. IET Systems Biology, 2008. In press.

References

- [1] Walter & Pronzato. *Identification of Parametric Models from Experimental Data*. Springer, 1997.
- [2] Pohjanpalo H. System identifiability based on the power series expansion of the solution. *Math. Biosci.* 41:21-33, (1978)
- [3] Margaria G, Riccomagno E, Chappel MJ, Wynn HP. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Math. Biosci.* 174: 1-26, (2001).
- [4] Saltelli A, Tarantola S, Campolongo F. Sensitivity analysis as an ingredient of modelling. *Statistical Sci.*, 15(4): 377-395, (2000).
- [5] Balsa-Canto E, Alonso AA, Banga JR. An optimal identification procedure for model identification in systems biology: Applications in cell signalling. In: Algöwer & Reuss (Eds). *Foundations of Systems Biology in Engineering*. (2007)
- [6] Brun R, Reichert P. Practical identifiability of large environmental simulation models. *Water Resources Res.* 37:1015-1030, (2001)
- [7] Balsa-Canto E, Alonso AA, Banga JR. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Systems Biology.* 4:11,(2010)
- [8] Seber & Wild. *Nonlinear regression*. Wiley series in Prob. & Math.Stat. (1989).
- [9] Hairer E, Wanner G. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. Springer Series in Comp. Math. 14, (1999)
- [10] Peifer M, Timmer J. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *IET Systems Biol*, 1(2):78-88, (2007)
- [11] Balsa-Canto E, Peifer M, Banga JR, Timmer J, Fleck C. Hybrid optimization method with general switching strategy for parameter estimation. *BMC Systems Biol*, 2:26, (2008) doi:10.1186/1752-0509-2-26
- [12] Nash & Sofer. *Linear and Nonlinear programming*. Mc-Graw-Hill, 1996
- [13] Fletcher. *Practical methods of Optimization*. John Wiley & Sons. 1987)
- [14] Esposito WR, Floudas C. Global optimization for the parameter estimation of differential-algebraic systems. *Ind & Eng Chem Res*, 39:1291-1310, (2000)
- [15] Gau CY, Stadtherr MA. Reliable Nonlinear Parameter Estimation Using Interval Analysis: Error in Variable Approach. *Comp & Chem Eng*, 24:631-637, (2000)
- [16] Lin Y, Stadtherr MA. Deterministic global optimization for parameter estimation of dynamic systems. *Ind & Eng Chem Res*, 45:8438-8448, (2006)
- [17] Polisetty P, Voit EO, Gatzke EP. Identification of metabolic system parameters using global optimization methods. *Theor Biol & Med Mod*, 3:4, (2006)
- [18] Michalewicz & Fogel. *How to solve it: Modern heuristics*. Springer, 2000
- [19] Dréo J, Petrowski A, Siarry P, Taillard E. *Metaheuristics for hard optimization. Methods and case studies*. Springer, 2006
- [20] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science*, 220: 671-680,(1983).
- [21] Mendes P, Kell DB. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* 14: 869-883, (1998)
- [22] Runarsson T, Yao X. Stochastic ranking for constrained evolutionary optimization. *IEEE Trans Evol Comp*, 564:284-294, (2000)
- [23] Moles CG, Mendes P, Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13:2467-2474, (2003)
- [24] Storn R, Price K. Differential Evolution – a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J Global Optim*, 11:341-359, (1997)
- [25] Balsa-Canto E, Vassiliadis V, Banga J: Dynamic Optimization of Single- and Multi-Stage Systems Using a Hybrid Stochastic-Deterministic Method. *Ind Eng Chem Res*, 44(5):1514-1523, (2005)
- [26] Rodriguez-Fernandez M, Mendes P, Banga JR. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosyst*, 83:248-265, (2006)
- [27] Egea JA, Rodriguez-Fernandez M, Banga JR, Martí R. Scatter Search for Chemical and Bio-Process Optimization. *J Glob Opt*, 37(3):481-503, (2007)
- [28] Egea JA. New heuristics for global optimization of complex bioprocesses. PhD thesis. Univ. of Vigo, Spain. (2008)
- [29] Rodriguez-Fernandez M, Egea JA, Banga JR. Novel Metaheuristic for Parameter Estimation in Nonlinear Dynamic Biological Systems. *BMC Bioinf.*, 7:483, (2006)
- [30] Ljung. *System identification: Theory for the user*. Prentice Hall, New Jersey, 1999.
- [31] Joshi M, Seidel-Morgenstern A, Kremling A. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Eng.*, 8:447-455, (2006).
- [32] Balsa-Canto E, et al. Computational design of optimal dynamic experiments in systems biology: a case study in cell signaling. In M. Canovas, J.L. Iborra, & A. Manjon, Eds, *Understanding and Exploiting Systems Biology in Bioprocesses and Biomedicine*, pages 103-117. Fundacion CajaMurcia, (2006).

References

- [33] Balsa-Canto E, Alonso AA, Banga JR. *Computational Procedures for Optimal Experimental Design in Biological Systems*. IET Systems Biology, 2008. In press.
- [34] García MR. *Identification and Real Time Optimisation in the Food Processing and Biotechnology Industries*. PhD Thesis. Univ. of Vigo. Spain. 2008. (Appendix B).
- [35] Banga JR, Versyck KJ, van Impe JF. *Computation of optimal identification experiments for nonlinear dynamic process models: an stochastic global optimization approach*. Ind & Eng Chem Res. 41:2425-2430, (2002).

Further reading

- [36] Schittkowski K. *Numerical Data Fitting in Dynamical Systems – A Practical Introduction with Applications and Software* Kluwer Academic; 2002.
- [37] Jaqaman K, Danuser G. *Linking data to models: data regression*. Nat. Rev. Mol. Cell Bio., 7(11):813{819, (2006)
- [38] van Riel NAW. *Dynamic modelling and analysis of biochemical networks: Mechanism-based models and model-based experiments*. Brief. Bioinform., 7(4):364{374, (2006)
- [39] Kremling A, Saez-Rodriguez. *Systems biology - an engineering perspective*. J. Biotechnol., 129:32-351, (2007)

.....