In Silico Systems Biology

Parameter Estimation with COPASI

Sven Sahle

BioQuant
MODEL base of LIFE

Introduction

- When starting a modeling project usually many parameters of the model are not known
- How can I find out about the effect of parameter changes in the model?
  - Sensitivities / Metabolic control analysis
  - Parameter scans
- How can I find out about parameter values?

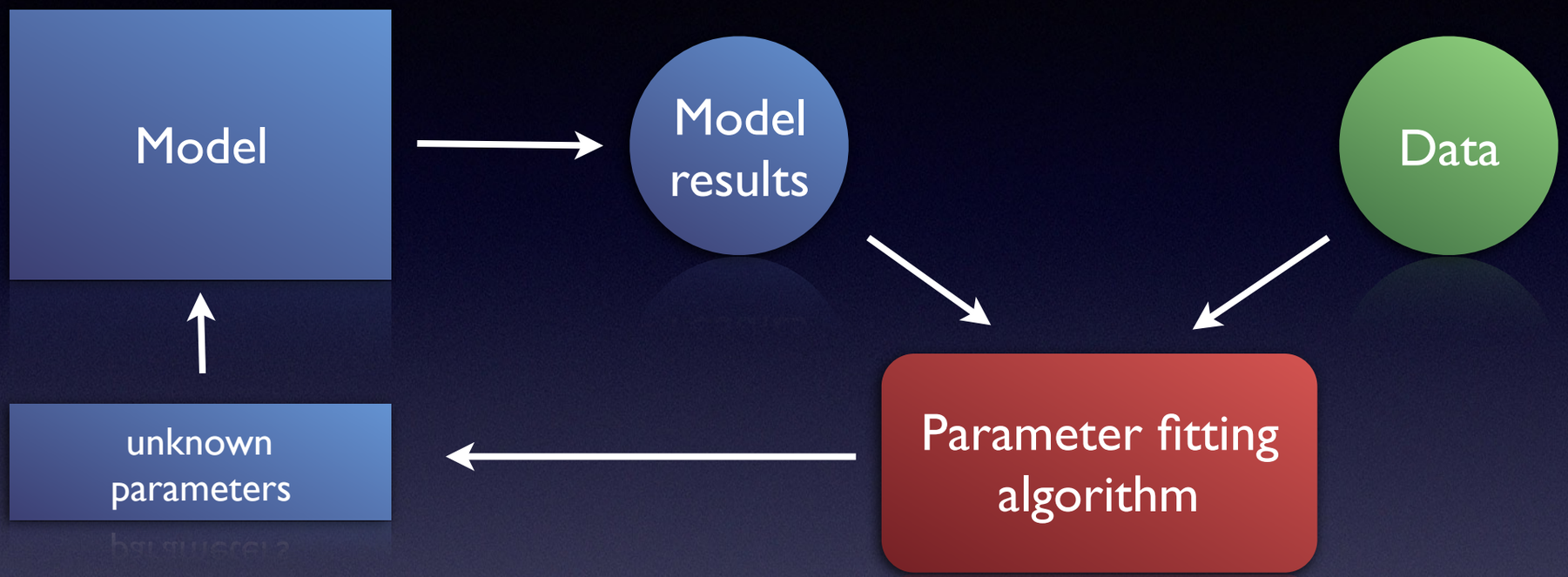
Introduction

- simple approach: try to design an experiment for measuring the specific parameter
  - typically *in vitro* experiment
  - e.g. for rate constants: put different amounts of substrate in a test tube and measure how fast the reaction proceeds
- Problems:
  - often not possible, different from *in vitro* conditions

Introduction

- Systems biology approach: adapt a complete model to experimental data
  - indirect method: use the model to find out about a parameter by measuring something else
  - can also be used to answer more complex questions, such as model identification
  - more difficult

Basic idea of parameter fitting



- Change the parameters of a model so that it „best“ fits the data

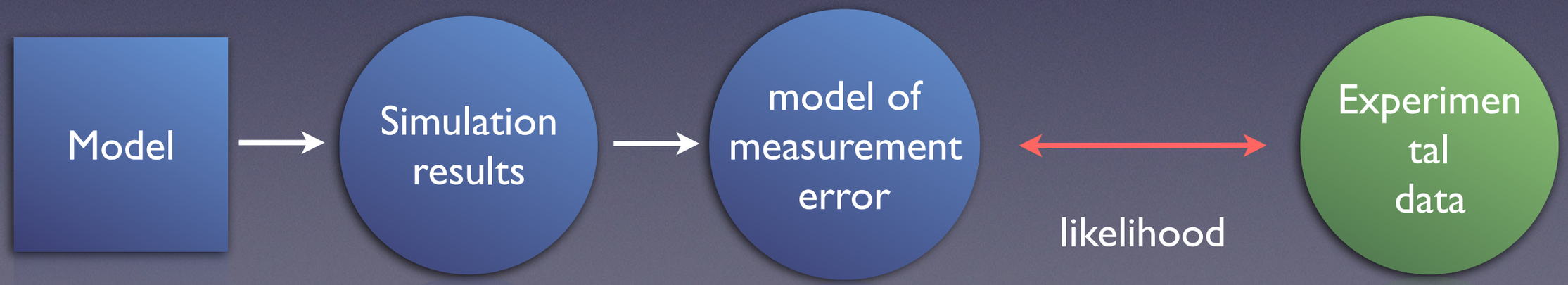
Criteria for parameter fitting

- What is the „best“ fit for a given set of data?
  - This is a mathematically difficult question
  - Fortunately, heuristically there are simple solutions.

- Required information:
  - Model
  - Knowledge about measurement process

Maximum Likelihood

- In principle, if we know the model and the measurement process, we can calculate the probability that the the measurements would be the result of a simulation of the model.
- If this probability is high, the model is good.



Maximum Likelihood

- To do parameter fitting, we need an algorithm that changes the unknown parameters so that the likelihood becomes maximal.
- This is mathematically very nice, but the probality is difficult to calculate in realistic cases.

Least squares method

- If we make some assumptions about the measurement errors, we can find a rather simpler criterion:
- **Assumption:** Error follows a normal distribution, measurement error is uncorrelated
- Leads to **Criterion:** Likelihood is maximal when the difference (as defined on the next slide) between measurements and simulation results is minimal
  - This is easy to calculate but not the most general case.

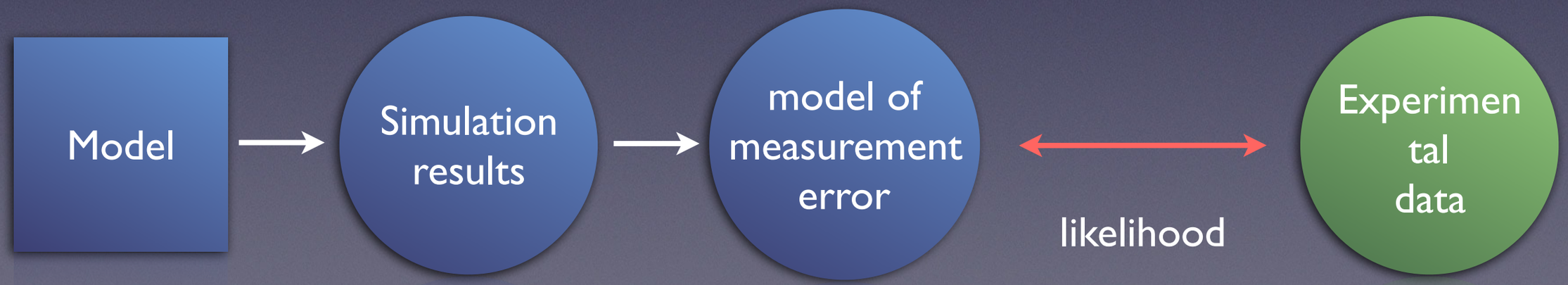
Criteria for parameter fitting

- What is the „best“ fit for a given set of data?
  - This is a mathematically difficult question
  - Fortunately, heuristically there are simple solutions.

- Required information:
  - Model
  - Knowledge about measurement process

Maximum Likelihood

- In principle, if we know the model and the measurement process, we can calculate the probability that the the measurements would be the result of a simulation of the model.
- If this probability is high, the model is good.



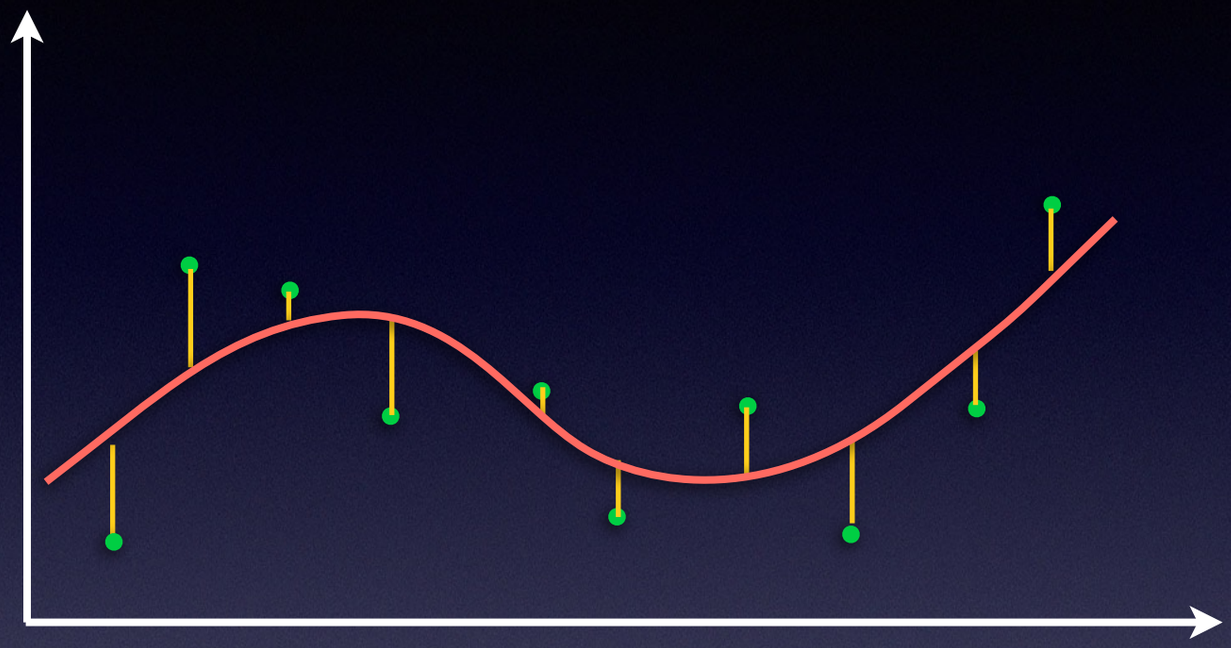
Maximum Likelihood

- To do parameter fitting, we need an algorithm that changes the unknown parameters so that the likelihood becomes maximal.
- This is mathematically very nice, but the probality is difficult to calculate in realistic cases.

Least squares method

- If we make some assumptions about the measurement errors, we can find a rather simpler criterion:
- **Assumption:** Error follows a normal distribution, measurement error is uncorrelated
- Leads to **Criterion:** Likelihood is maximal when the difference (as defined on the next slide) between measurements and simulation results is minimal
  - This is easy to calculate but not the most general case. But it still works.

Least squares distance measure

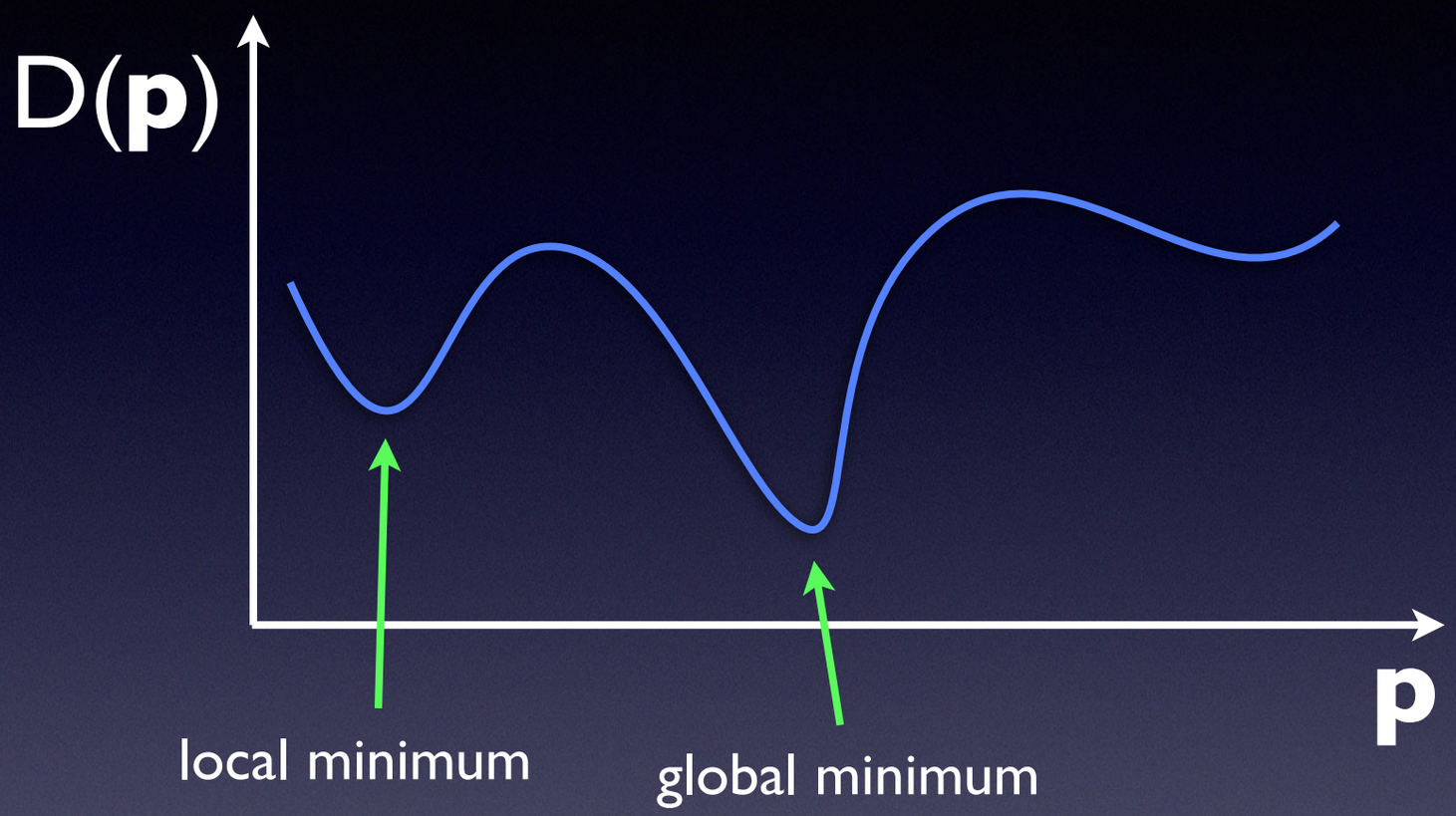


$$D(p) = \sum_{i=1}^{N} (x_i - y_i(p))^2$$

x_i : measured values for time t_i

$y_i(p)$: simulated values for time t_i, parameter p

The target function



The function is usually high dimensional!

- This is the function that needs to be minimized.
- usually has many local minima

Parameter space

- For a complete specification of the parameter fitting problem we still need to specify the unknown parameters:
  - List of parameters with allowed range of values
  - These parameters span an M-dimensional space, the *parameter space*.
  - One specific set of parameters corresponds to a point in parameter space

Optimization problem

- We now need a way to find the set of parameter values (a point in parameter space) for which the distance D is minimal (the best fit).
- A systematic scan of the parameter space is not possible when the dimensionality is large (many unknown parameters)
- Example: 10 parameters with 10 values each: 10^{10} evaluations. Even if we can do 100 simulations/s, it would take 3 years.

Optimization algorithms

- In general it is very difficult to find the parameter values for which D(**p**) is minimal
- It can be shown that there is no optimal optimization algorithm for all cases (this means there is no way to decide which one is best for a given problem)
- This means you should always try several algorithms for difficult parameter estimation problems.

Optimization Algorithms

- Based on derivatives
  - Steepest descent
  - Newton
  - Levenberg-Marquardt
- Using geometry
  - Nelder-Mead (simplex)
  - Hooke-Jeeves
- Based on genetics
  - Genetic algorithm
  - Evolutionary programing
  - Evolution strategy
- Other stochastic
  - random search
  - particle swarm
  - simulated annealing

Specification of a parameter estimation problem

- What kind of information is needed for the computer to do a parameter estimation?
  - The model
  - the experimental data
  - the mapping between experimental data and model simulation results
  - the ranges of possible values for the unknown parameters
  - the optimization algorithm

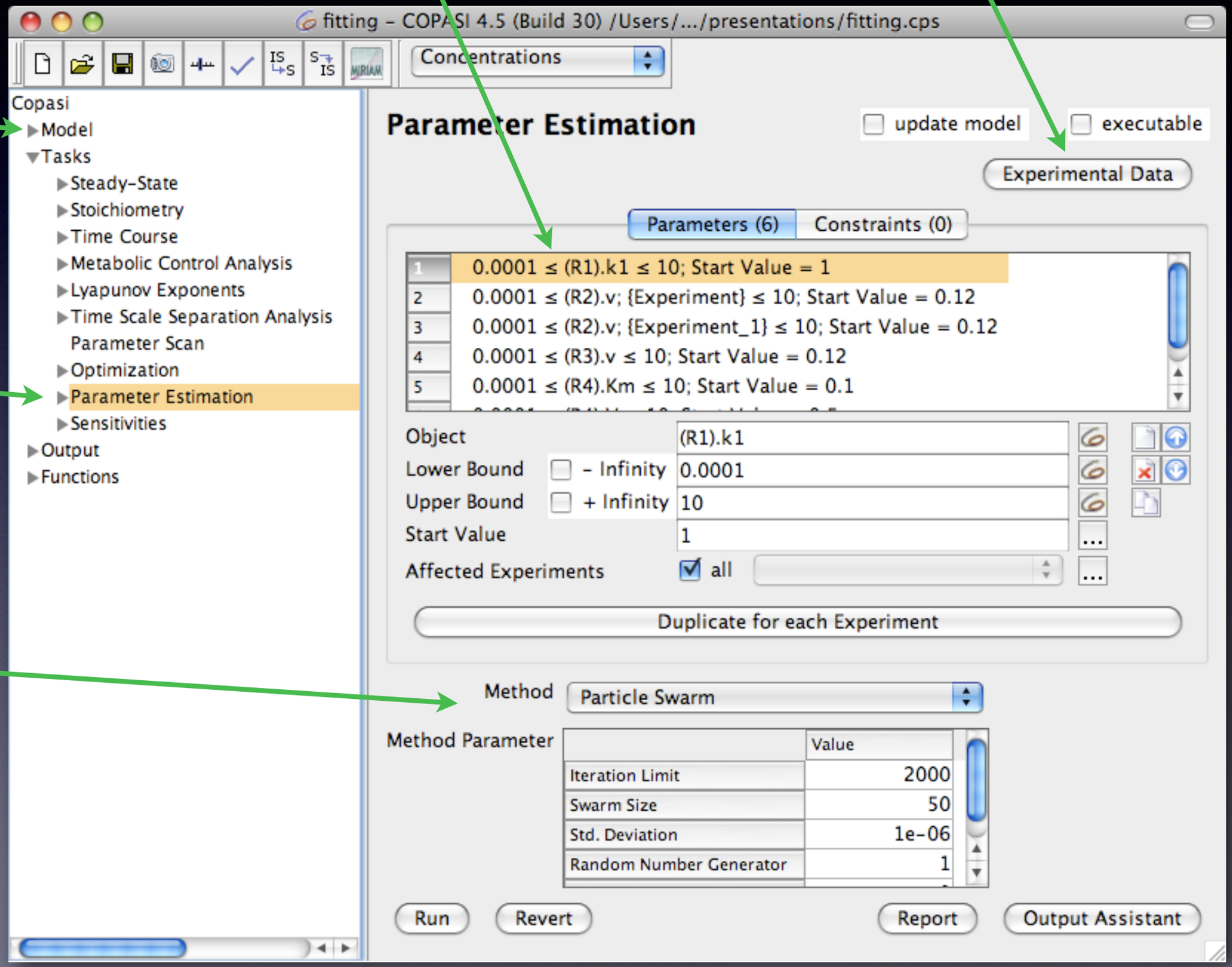
list of unknown parameters (including ranges)

choose experimental data

Model

select parameter estimation

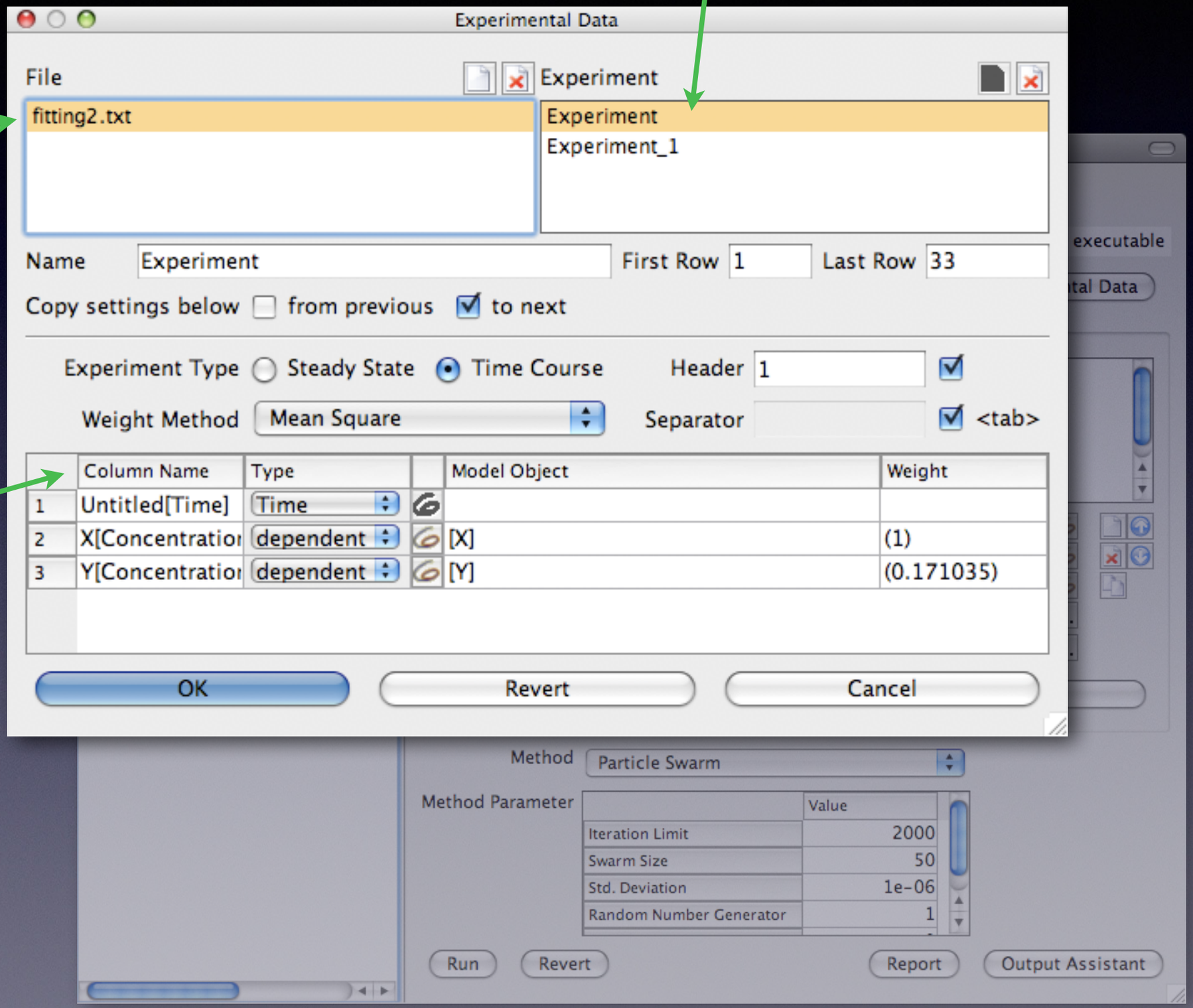
choose optimization algorithm



possibly several
experiments per file

data files

mapping of data to
model elements



Important:

- The result of a parameter fitting always needs to be inspected afterwards!
- Having a good result for a fit does **not** mean that the parameter value is the „true“ one. This depends on the assumptions about the errors and the correctness of the model.
- For the stochastic algorithms the result is not reproducible!

Using several experiments for parameter estimation

- The more data available, the better.
  - So if you have data from several experiments it should be used for parameter estimation simultaneously
- COPASI can deal with an arbitrary number of experiments, also of different kinds (combined time course/steady state, different variables, different time points, etc.)

several experiments...

- Adding data from several experiments is straight forward in COPASI. Several data files can be specified and each can contain several experiments
- Important information: What is the same for all experiments and what is different between experiments? For the things that are different, are they known or unknown?

several experiments...

- Simplest case: Repeated experiments.
  - nothing special needs to be done in COPASI
- Several experiments under different conditions. The conditions are known.
  - Example: Different stimulations in several experiments.
  - In COPASI: The stimulation needs to be a parameter in the model. In the experimental data specification this value is selected as an *independent* parameter. *Independent* data is known data that is provided in the data file. *Dependent* data is data that is used for fitting.

several experiments...

- Experiments where some conditions are different, but not known
  - Example: *In vivo* experiments, even if the experiment is repeated with the same preparation, the initial conditions (inside the organism) are typically different.
  - In COPASI: The user can specify that some parameters are fitted for all experiments, and some are fitted for a specific subset of experiments.