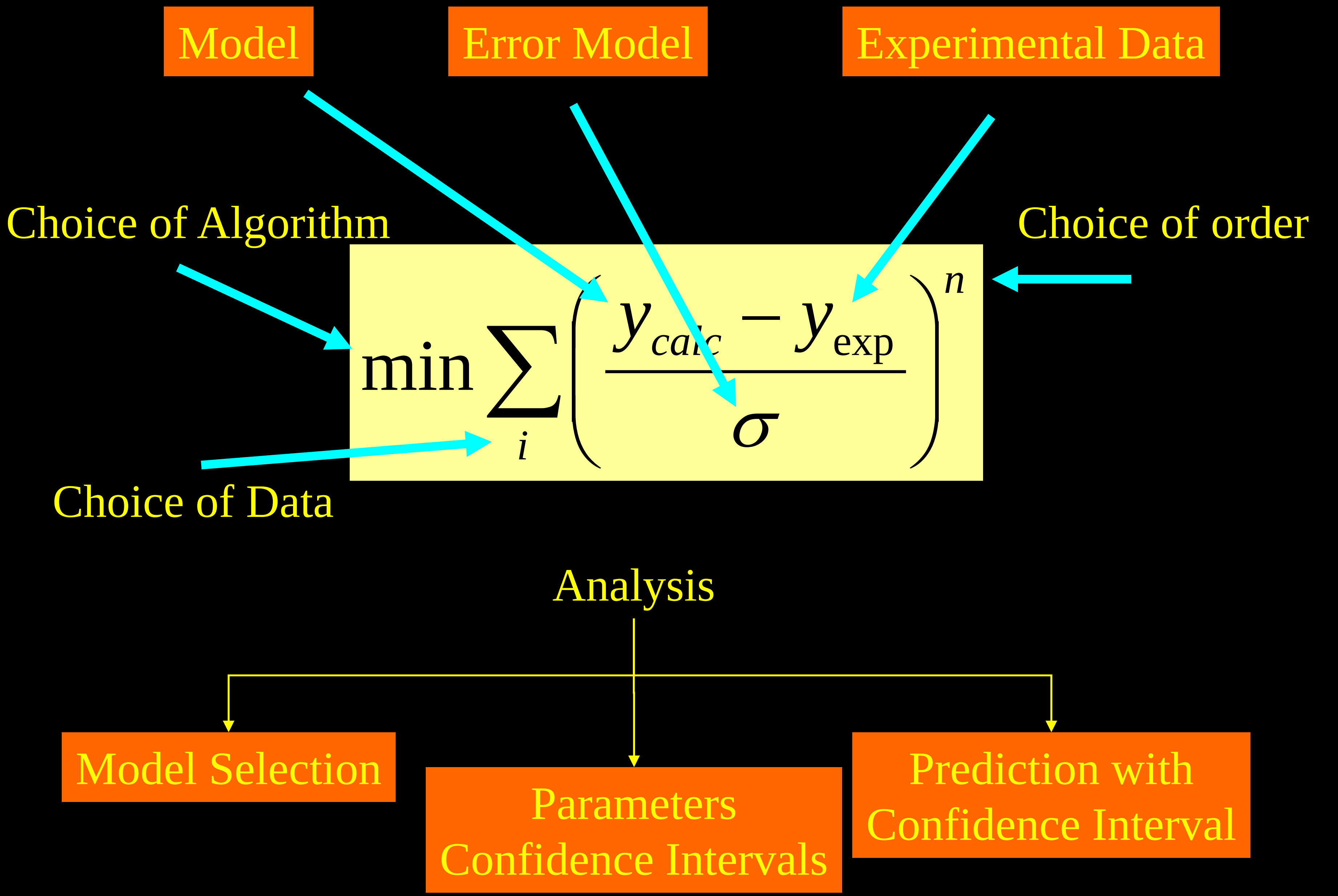
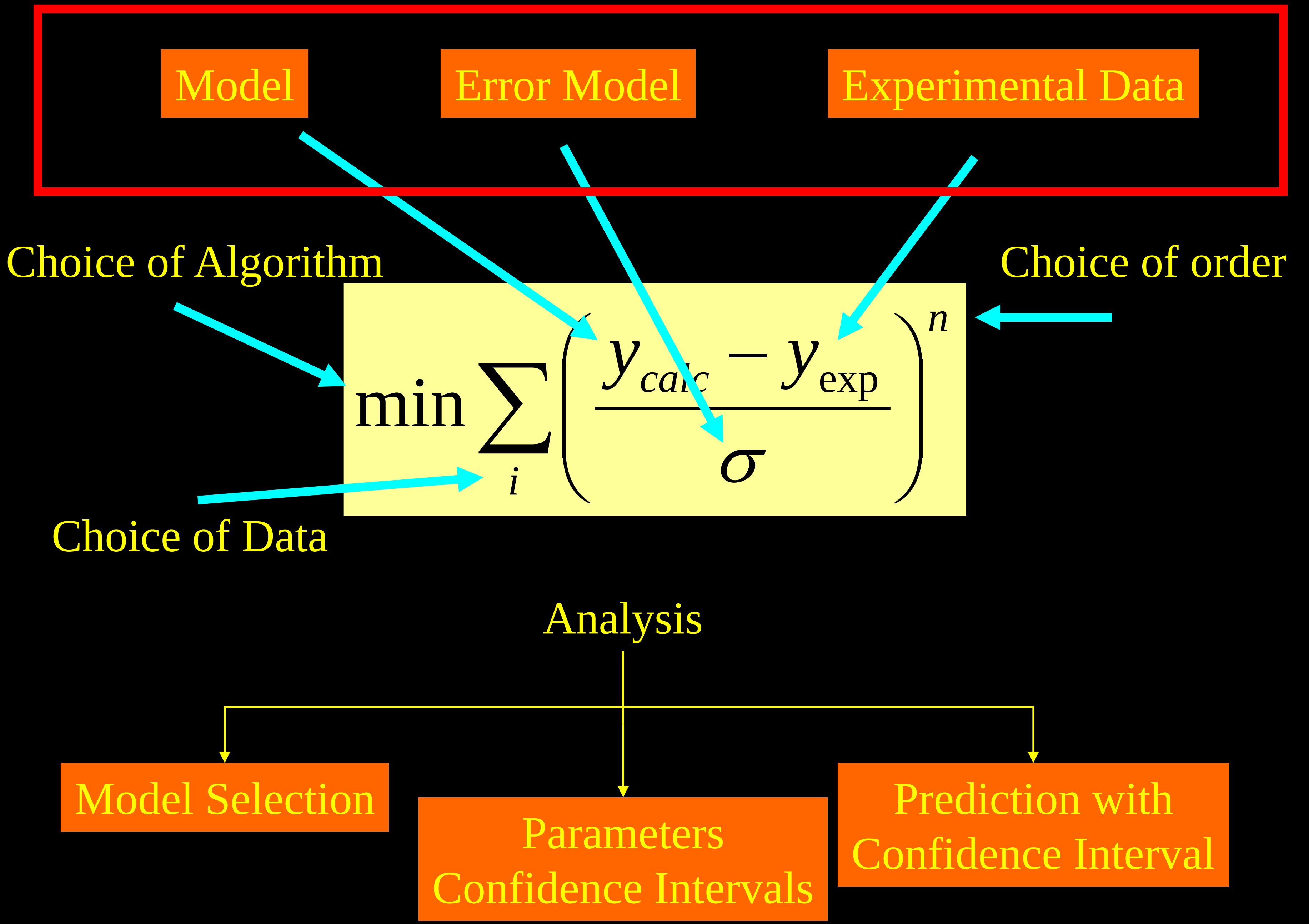
Parameter Estimation
An overview and techniques

Peter J.T. Verheijen

Dept of Chemical Engineering
TU Delft


Model
Error Model
Experimental Data
Choice of Algorithm
Choice of order
$$\min \sum_i \left(\frac{y_{calc} - y_{\exp}}{\sigma} \right)^n$$
Choice of Data
Analysis
Model Selection
Parameters Confidence Intervals
Prediction with Confidence Interval


Model
Error Model
Experimental Data
Choice of Algorithm
Choice of order
$$\min \sum_i \left(\frac{y_{calc} - y_{\exp}}{\sigma} \right)^n$$
Choice of Data
Analysis
Model Selection
Parameters Confidence Intervals
Prediction with Confidence Interval

Model

- Model structure - care: existence of sensititvity
  - Algebraic models => ODE-systems
  - Alg. Models + constraints (>, <, =) => DAE-systems
  - Models with discrete parameters
- Sets of models
  - Hierarchical models (relation between models)
  - Equivalent models (no relation between models)
  - Lumped models (different models, different data)
- Choice of transformation of parameters

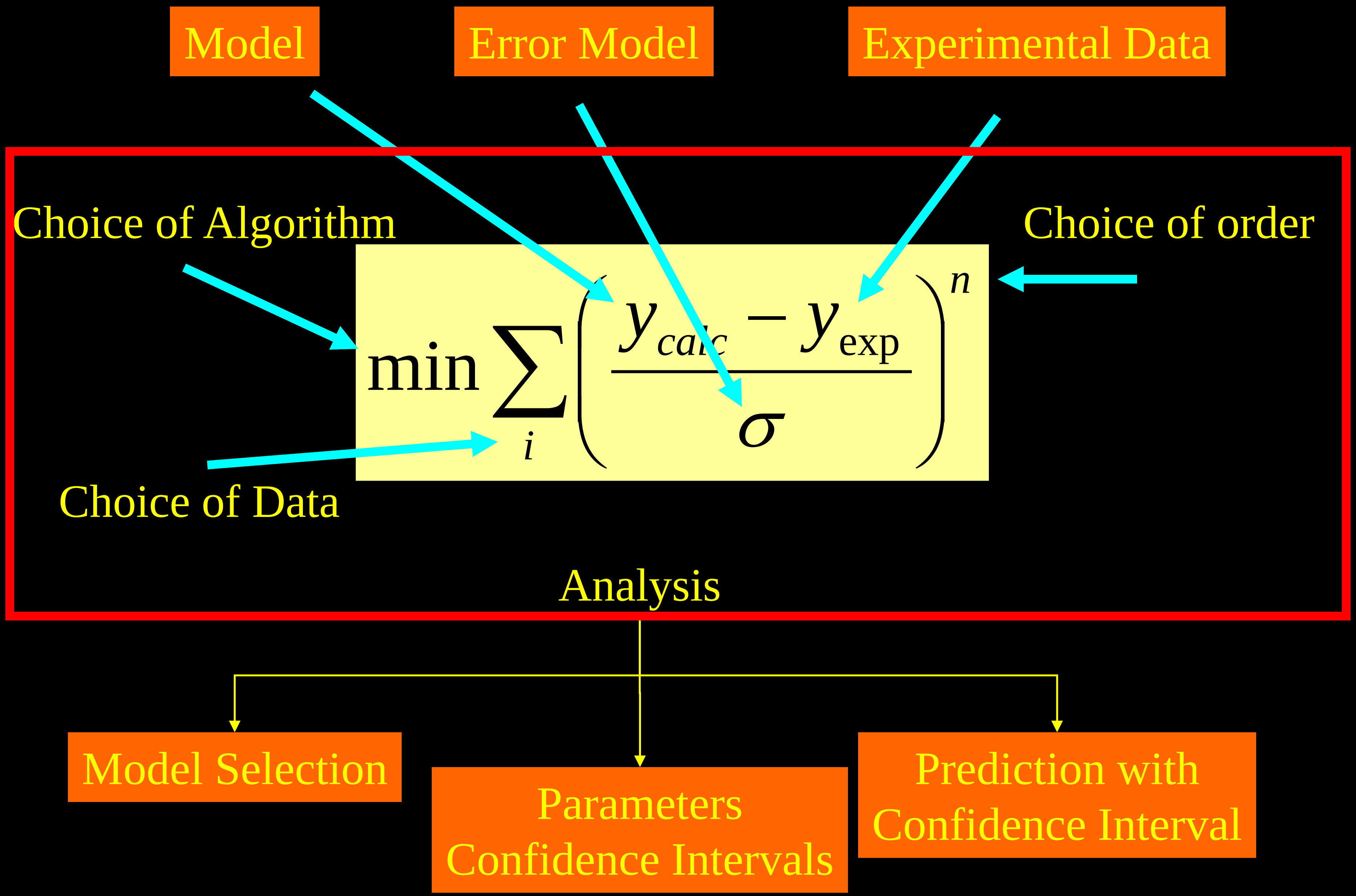
Error model

- Assumed normal distribution => unknown distribution
- Weighted regression is absolute necessary
- Errors - uncorrelated
  - from knowledge of apparatus
  - repeated measurements => Lack-of-fit tests
- Errors - correlated
  - Possible but laborious and relatively little contribution
- Care: data processing/transformations cause transformation of error model

Experimentation and Uncertainty Analysis for Engineers, Coleman and Steele, 1999

Experimental Data

- Pre-processing of data can cause artifacts in modelling (e.g. NMR) => Care
- Datareconciliation - good for presentation of data, but less in the actual fit process itself.
- Outlier detection techniques: very important - point at new physics and chemistry => Example: Least Median of Squares


Model
Error Model
Experimental Data
Choice of Algorithm
Choice of order
$$\min \sum_i \left(\frac{y_{calc} - y_{\exp}}{\sigma} \right)^n$$
Choice of Data
Analysis
Model Selection
Parameters Confidence Intervals
Prediction with Confidence Interval

Choices: Order and Data

- Order:
  - normal is second order, because that relates to normal distribution
  - other orders: choice of weight of extreme residues
  - non-normal distributions, Maximum-Likelihood method
- Data
  - Always do simultaneous fits!
  - On time-domain: fit data on different time-segments in order to isolate the short-term from the long-term effects.

Choice: Algorithm

Parameter Estimation is an Optimization Problem! All optimising routines are applicable

- Direct methods: Nelder-Mead simplex, grid search
- Gradient Methods: Steepest Descent, GRG, etc
- Second Order methods: Gauss-Newton, Levenberg-Marquardt. Note: second order is not precise.
- MINLP: if combined with discrete variables
- Stochastic Methods: Random Search, Simulated Annealing, Genetic Algorithms

Optimization Process: Tips and Tricks

- Good initial estimates are essential: example kinetic systems reduces to linear system in kinetic constants
- Put constraints on search area of parameters: on physical and intuitive grounds
- Use a strategy to increase the set of fittable parameters to the total set of fittable parameters
- Prefer software that use analytical derivatives
- Add a weight function of parameters are preferred to remain ‘close’ to a deisred values; regularization

Analysis

An often forgotten step: VERIFY

- Distribution of residues:
  - over domain of time, place, etc
  - correlation
  - normally distributed
- Lack-of-fit analysis
- Investigation of active constraints: Lagrange Multipliers
- Sensitivity of optimum with respect to constants

Model

Error Model

Experimental Data

Choice of Algorithm

Choice of order

$$\min \sum_{i} \left(\frac{y_{calc} - y_{\exp}}{\sigma} \right)^{n}$$

Choice of Data

Analysis

Model Selection

Parameters
Confidence Intervals

Prediction with
Confidence Interval

Model Selection

Given: a set of models and their fit results

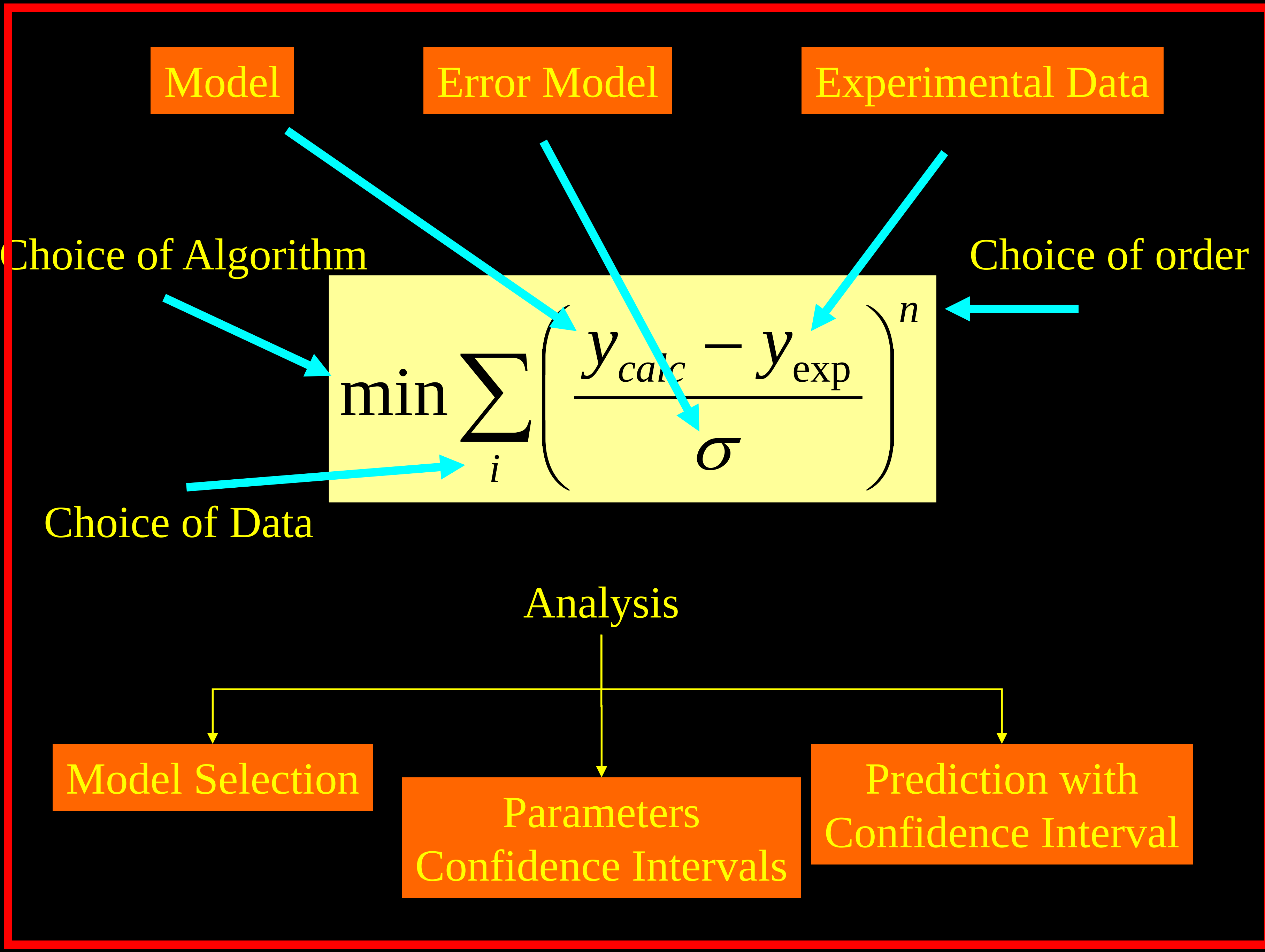
- Model Adequacy
- Hierarchical Models: from Complex to Simple®
  - Using statistical F-test
  - Identification of dominant components: Singular Value Decomposition
  - Superstructure approach: add binary variables - MINLP
- Equivalent models: use Bartlett’s test, Bayes’ rule
- Rate models with criteria to weigh complexity versus accuracy of fit (e.g. AIC, FPE, RSD)

Parameters and Confidence Intervals

- Asymptotic Confidence Intervals
  - follow directly from sensitivity equations
  - Can be determined afterwards (eg after direct method) by perturbation
- Asymmetric Confidence Intervals
  - From Ssres-plot
  - As optimisation problem: find max/min of parameter given a target Ssres
- Monte-Carlo Methods
- Bootstrap methods

Prediction with Confidence Intervals

A specified function of parameters (reactor conversion, selectivity, estimated maximum yield) is a generalization of single parameter. Confidence intervals are determined as previous slide!


Model
Error Model
Experimental Data
Choice of Algorithm
Choice of order
$$\min \sum_i \left(\frac{y_{calc} - y_{\exp}}{\sigma} \right)^n$$
Choice of Data
Analysis
Model Selection
Parameters Confidence Intervals
Prediction with Confidence Interval

Experimental Design

Purpose: choose experimental setup such that confidence interval (or volume) size is minimized

- D-optimal design: volume (determinant)
- single parameter optimal design (single asymptotic error or ‘true’ confidence interval)
- To achieve minimal prediction error eg in design (of a process!)
- Simulate experiment to show the various choices

Optimum Experimental Designs, Atkinson and Donev, 1992 and 1996

Conclusions

- Parameter estimation is foremost an optimization problem
- The use of (confidence) intervals at all levels is a measure of quality and criterion for decision making
- Simulation of complex models before experimentation reduces the experimental effort and, especially, the data processing/parameter estimation post-experiment stage
- Interesting applications: DAE-systems; discrete systems.
- Stochastic optimization techniques are interesting but not efficient. SQP and related methods are reliable.