The Systems Biology Cycle: Computational Workflow & Uncertainty Quantification

Maria Navarro
maria.navarro@cwi.nl

Centrum Wiskunde & Informatica
Amsterdam

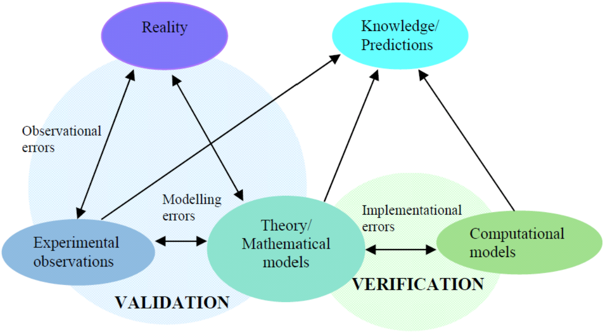
Scientific Meeting 06/06/2014

INTRODUCCTION

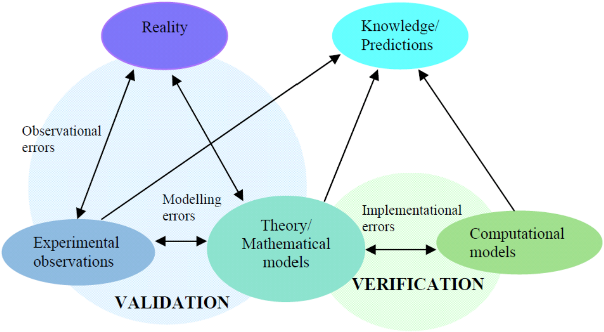
- Postdoc in the Scientific Computing for Systems Biology / Life Science (LS)
- **BioPreDyn:** From Data to Models (European project, FP7,#289434) with J.G. Blom and A.T. Valderrama

Outline

Errors


Reality
Knowledge/ Predictions
Observational errors
Modelling errors
Implementational errors
Experimental observations
Theory/ Mathematical models
Computational models
VALIDATION
VERIFICATION

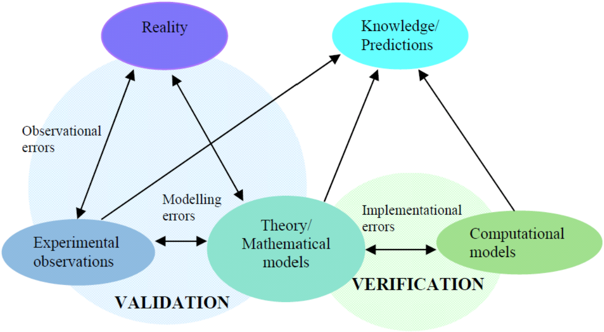
Errors



Experimental models: controlled version of reality

model organisms, synthetic biology

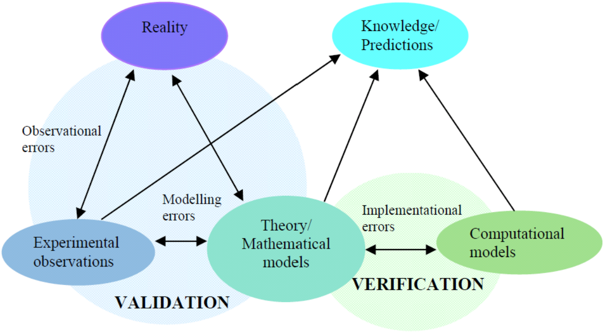
Errors



Mathematical models: *precise* description of an *approximation* of reality

assumptions, equations

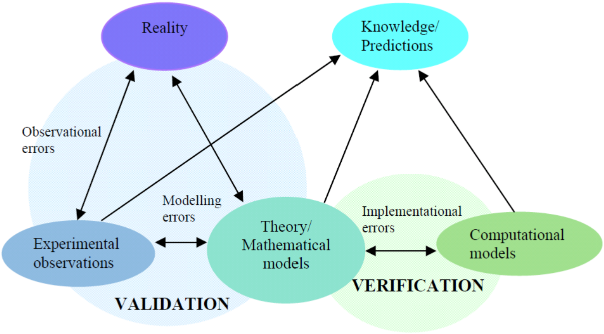
Errors



Computational models: implementation of mathematical model

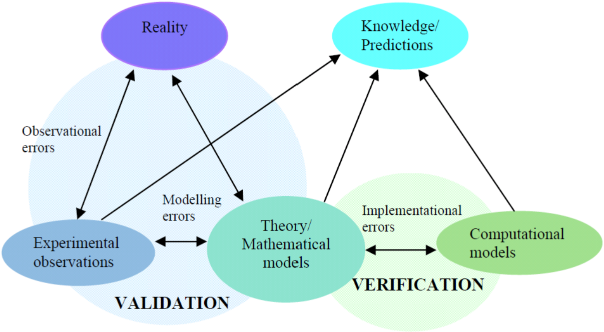
equation-based, model identification, verification, validation

Errors

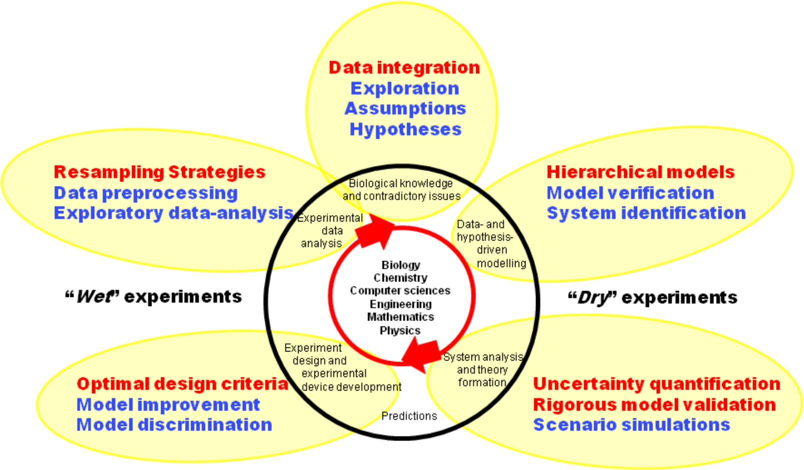


(Controllable) errors everywhere

Errors

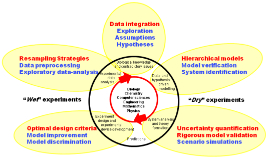


(Controllable) errors everywhere
THINK before you do


Data integration
Exploration
Assumptions
Hypotheses
Resampling Strategies
Data preprocessing
Exploratory data-analysis
Hierarchical models
Model verification
System identification
Biological knowledge and contradictory issues
Experimental data analysis
Data- and hypothesis-driven modelling
Biology
Chemistry
Computer sciences
Engineering
Mathematics
Physics
"Wet" experiments
"Dry" experiments
Experiment design and experimental device development
System analysis and theory formation
Predictions
Optimal design criteria
Model improvement
Model discrimination
Uncertainty quantification
Rigorous model validation
Scenario simulations

"Dry": Mathematical / Computational Model

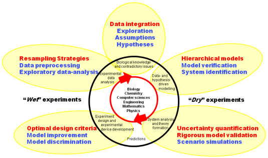
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)

“Dry”: Mathematical / Computational Model

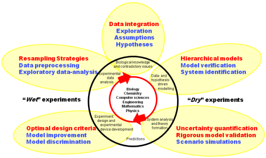
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)

"Dry": Mathematical / Computational Model

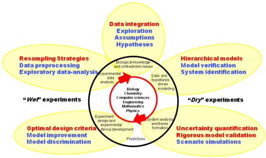
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)

"Dry": Mathematical / Computational Model

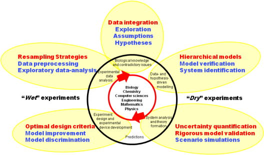
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)
4. Fit parameters with (part of) data (residual analysis, model validation)

"Dry": Mathematical / Computational Model

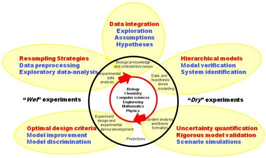
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)
4. Fit parameters with (part of) data (residual analysis, model validation)
5. Analyse propagation of exp. errors on parameters (practical ident., FIM)

“Dry”: Mathematical / Computational Model

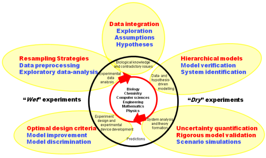
Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)
4. Fit parameters with (part of) data (residual analysis, model validation)
5. Analyse propagation of exp. errors on parameters (practical ident., FIM)
6. (Model discrimination (distinguish between various hypotheses/assumptions))

"Dry": Mathematical / Computational Model

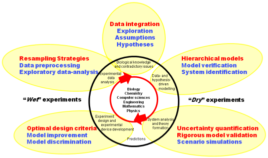
Experimental data for some observables (data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)
4. Fit parameters with (part of) data (residual analysis, model validation)
5. Analyse propagation of exp. errors on parameters (practical ident., FIM)
6. (Model discrimination (distinguish between various hypotheses/assumptions))
7. Optimal Experimental Design (experimental options, aim)

"Dry": Mathematical / Computational Model

Experimental data for some observables
(data analysis: exp. error, exploratory)



1. Mathematical Model (Hypotheses+assumptions+literature+data analysis)
2. Analyze model and parameter identifiability (math. analysis + structural ident.)
3. Computational model (verification: comp. model $\stackrel{?}{=}$ math. model)
4. Fit parameters with (part of) data (residual analysis, model validation)
5. Analyse propagation of exp. errors on parameters (practical ident., FIM)
6. (Model discrimination (distinguish between various hypotheses/assumptions))
7. Optimal Experimental Design (experimental options, aim)
8. Model predictions

Uncertainty Quantification

1 Why Uncertainty Quantification?

2 The Systems Biology Cycle

3 Polynomial Chaos Expansion

4 PCE for correlated inputs

Deterministic mathematical model

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)), \qquad \mathbf{y}(0) = \mathbf{y}_0$$

Uncertainty in *initial values* or *parameters* (e.g. from error in data)

ξ (multi-variate) random variable, $\mathbf{y}(t, \xi)$ random process:

Stochastic mathematical model

$$\mathbf{y}'(t, \xi) = \mathbf{f}(t, \mathbf{y}(t, \xi), \xi), \qquad \mathbf{y}(0, \xi) = \mathbf{y}_0(\xi)$$

How to compute $\mathbf{y}(t, \xi)$?

Deterministic mathematical model

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)), \qquad \mathbf{y}(0) = \mathbf{y}_0$$

Uncertainty in *initial values* or *parameters* (e.g. from error in data)

ξ (multi-variate) random variable, $\mathbf{y}(t, \xi)$ random process:

Stochastic mathematical model

$$\mathbf{y}'(t, \xi) = \mathbf{f}(t, \mathbf{y}(t, \xi), \xi), \qquad \mathbf{y}(0, \xi) = \mathbf{y}_0(\xi)$$

How to compute $\mathbf{y}(t, \xi)$? **Polynomial Chaos Expansion**

PCE: decomposition of random process y (cf Fourier, Taylor)

Wiener (1938) separation of variables

$$y(t,\xi) = \sum_{p=0}^{\infty} y_p(t)\Psi_p(\xi), \qquad \xi \text{ RV}$$

Ψ polynomials; orthogonal wrt **pdf** $\rho(\xi)$

$$< \Psi_i, \Psi_j >= \int_S \Psi_i(\xi)\Psi_j(\xi)\rho(\xi)d\xi = 0, \quad i \neq j$$

PCE: decomposition of random process y (cf Fourier, Taylor)

Wiener (1938) separation of variables

$$y(t,\xi) = \sum_{p=0}^{\infty} y_p(t)\Psi_p(\xi), \qquad \xi \text{ RV}$$

$\hat{p}$-th order approximation: $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$

PCE: **decomposition** of random process y (cf Fourier, Taylor)

Wiener (1938) separation of variables

$$y(t,\xi) = \sum_{p=0}^{\infty} y_p(t)\Psi_p(\xi), \qquad \xi \text{ RV}$$

$\hat{p}$-th order approximation: $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$

Project series or equation onto space $\{\Psi_0, ..., \Psi_{\hat{p}}\}$

PCE: decomposition of random process y (cf Fourier, Taylor)

Wiener (1938) separation of variables

$$y(t,\xi) = \sum_{p=0}^{\infty} y_p(t)\Psi_p(\xi), \qquad \xi \text{ RV}$$

$\hat{p}$-th order approximation: $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$

Project series or equation onto space $\{\Psi_0, ..., \Psi_{\hat{p}}\}$

Solution: response surface, pdf $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$

Basics

Response surface, pdf $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$

Response surface, pdf $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$
Explicit mean, variance, Sobol indices; higher moments: cheap integrals

Mean $\mu_y(t) = \int_S y(t,\xi)\rho(\xi)d\xi = y_0(t)$

Response surface, pdf $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$
Explicit mean, variance, Sobol indices; higher moments: cheap integrals

Mean $\mu_y(t) = \int_S y(t,\xi)\rho(\xi)d\xi = y_0(t)$

Variance $\sigma_y^2(t) = \int_S (y(t,\xi))^2\rho(\xi)d\xi - \mu_y^2(t) = \sum_{p=1} y_p(t)^2||\Psi_p||^2$

Response surface, pdf $y(t,\xi) \approx \sum_{p=0}^{\hat{p}} y_p(t)\Psi_p(\xi)$
Explicit mean, variance, Sobol indices; higher moments: cheap integrals

Mean $\mu_y(t) = \int_S y(t,\xi)\rho(\xi)d\xi = y_0(t)$

Variance $\sigma_y^2(t) = \int_S (y(t,\xi))^2\rho(\xi)d\xi - \mu_y^2(t) = \sum_{p=1} y_p(t)^2||\Psi_p||^2$

Sobol indices $S_l(t) = \frac{y_l(t)^2||\Psi_l||^2}{\sigma_y^2(t)}$ first order, independent random variables

- $n > 1$ RVs: multidimensional pdf's
  # expansion terms: $N = \frac{(n+\hat{p})!}{n!\,\hat{p}!} - 1$

- $n > 1$ RVs: multidimensional pdf's
  # expansion terms: $N = \frac{(n+\hat{p})!}{n!\,\hat{p}!} - 1$
- arbitrarily distributed independent RVs
  generalized Polynomial Chaos

- $n > 1$ RVs: multidimensional pdf's
  # expansion terms: $N = \frac{(n+\hat{p})!}{n!\,\hat{p}!} - 1$
- arbitrarily distributed independent RVs
  generalized Polynomial Chaos
- **arbitrarily multivariate distribution (2014)**
  incl. correlations between RVs
  In collaboration with J. Witteveen - Scientific Computing

Stochastic mathematical model

$$\mathbf{y}'(t,\xi) = \mathbf{f}(t,\mathbf{y}(t,\xi),\xi), \qquad \mathbf{y}(0,\xi) = \mathbf{y}_0(\xi)$$

ξ CORRELATED multi-variate random variable and $\rho(\xi)$ its pfd

How to compute $\mathbf{y}(t,\xi)$ taking correlations into account?

Stochastic mathematical model

$$\mathbf{y}'(t,\xi) = \mathbf{f}(t,\mathbf{y}(t,\xi),\xi), \qquad \mathbf{y}(0,\xi) = \mathbf{y}_0(\xi)$$

ξ CORRELATED multi-variate random variable and $\rho(\xi)$ its pfd

How to compute $\mathbf{y}(t,\xi)$ taking correlations into account?
multivariate arbitrarily distributed PCE

Orthogonal polynomials for correlated inputs

The polynomials have to be orthogonal with respect to the inner product $<,>$ with $\rho(\boldsymbol{\xi})$

$$\langle \Phi_i(\boldsymbol{\xi}), \Phi_j(\boldsymbol{\xi}) \rangle = \int_{\Xi} \Phi_i(\boldsymbol{\xi}) \Phi_j(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} = ||\Phi_i||^2 \delta_{ij},$$

A) For **independent** random variables: Φ_i's are tensor products of one-dimensional polynomials, $\Phi_i(\boldsymbol{\xi}) = \prod_{j=1}^{n} \tilde{\Phi}_i(\xi_j)$

B) For **correlated** random variables: By applying GRAM-SCHMIDT orthogonalization to a set of of linearly independent polynomials $\{e_j(\boldsymbol{\xi})\}_{j=0}^{N}$ given by

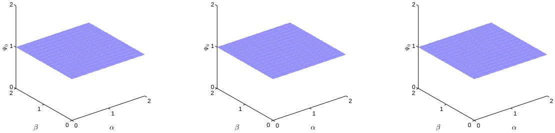
$$e_j(\boldsymbol{\xi}) = \prod_{l=1}^{n} \xi_l^{j_l}, \quad j = 0, \ldots, N, \; j_l \in \{0, \ldots, p\}, \text{ and } \sum_{l=1}^{n} j_l \leq p,$$

E.g., if $n = p = 2$ then $N + 1 = 5$ and the set of linearly independent polynomials $\{e_j(\xi_1, \xi_2)\}_{j=0}^{5}$ equals $\{1, \xi_1, \xi_2, \xi_1^2, \xi_2^2, \xi_1\xi_2\}$

Orthogonal polynomials for correlated inputs

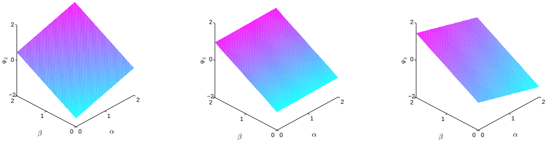
Order 0:

$\varrho = -0.5$ $\varrho = 0$ $\varrho = 0.5$



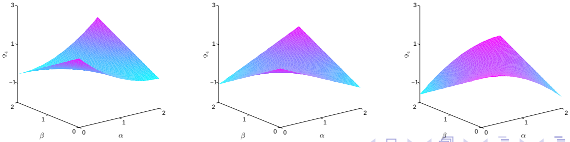
Order 1:

$\varrho = -0.5$ $\varrho = 0$ $\varrho = 0.5$



Order 2:

$\varrho = -0.5$ $\varrho = 0$ $\varrho = 0.5$



Explicit mean, variance, Sobol indices

Explicit mean, variance, Sobol indices

Mean $\mu_{\boldsymbol{y}}(t) = \int_S \boldsymbol{y}(t, \boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi} = \boldsymbol{y}_0(t)$

Explicit mean, variance, Sobol indices

Mean $\mu_{\boldsymbol{y}}(t) = \int_S \boldsymbol{y}(t, \boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi} = \boldsymbol{y}_0(t)$

Variance $\sigma^2_{\boldsymbol{y}}(t) = \int_S (\boldsymbol{y}(t, \boldsymbol{\xi}))^2 \rho(\boldsymbol{\xi})d\boldsymbol{\xi} - \boldsymbol{\mu}^2_{\boldsymbol{y}}(t) = \sum_{p=1} \boldsymbol{y}_p(t)^2 ||\Psi_p||^2$

Orthogonal polynomials for correlated inputs

Explicit mean, variance, Sobol indices

Mean $\mu_{\boldsymbol{y}}(t) = \int_S \boldsymbol{y}(t,\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi} = \boldsymbol{y}_0(t)$

Variance $\sigma^2_{\boldsymbol{y}}(t) = \int_S (\boldsymbol{y}(t,\boldsymbol{\xi}))^2\rho(\boldsymbol{\xi})d\boldsymbol{\xi} - \boldsymbol{\mu}^2_{\boldsymbol{y}}(t) = \sum_{p=1} \boldsymbol{y}_p(t)^2||\Psi_p||^2$

Sobol indices $\mathrm{S}_{\boldsymbol{I}}(t) = \frac{\mathrm{Cov}[M_{\boldsymbol{I}}(t,\boldsymbol{\xi}_{\boldsymbol{I}}),\boldsymbol{y}]}{\sigma^2_{\boldsymbol{y}}(t)} = S^u_I(t) + S^c_I(t)$, with

$$S^u_I(t) = \frac{\mathrm{Var}[\mathrm{E}_{\boldsymbol{\xi}_{\neg I}}[\boldsymbol{y}|\boldsymbol{\xi}_I]]}{\sigma^2_{\boldsymbol{y}}(t)},\ \text{and}$$

$$S^c_I(t) = \frac{\mathrm{Cov}[\mathrm{E}_{\boldsymbol{\xi}_{\neg I}}[\boldsymbol{y}|\boldsymbol{\xi}_I],\boldsymbol{y}-\mathrm{E}_{\boldsymbol{\xi}_{\neg I}}[\boldsymbol{y}|\boldsymbol{\xi}_I]]}{\sigma^2_{\boldsymbol{y}}(t)}$$

CWI

Example

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{C},$$

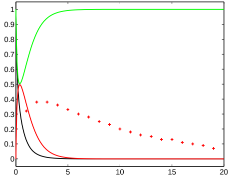
$$\mathrm{C} \overset{k_3}{\rightarrow} \mathrm{E} + \mathrm{P}$$

Example

$$E + S \underset{k_2}{\overset{k_1}{\rightleftarrows}} C,$$

$$C \xrightarrow{k_3} E + P$$

Parameter estimation

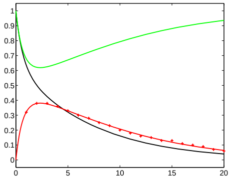


Example

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{C},$$

$$\mathrm{C} \overset{k_3}{\rightarrow} \mathrm{E} + \mathrm{P}$$

Parameter estimation

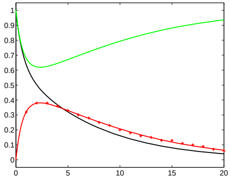


Example

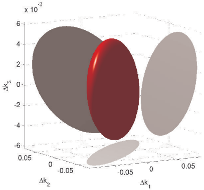
$$E + S \underset{k_2}{\overset{k_1}{\rightleftarrows}} C,$$

$$C \xrightarrow{k_3} E + P$$

Parameter estimation



Uncertainty in parameters

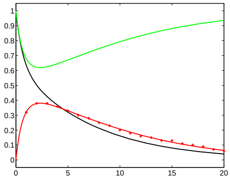


Example

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{C},$$

$$\mathrm{C} \overset{k_3}{\rightarrow} \mathrm{E} + \mathrm{P}$$

Parameter estimation



Uncertainty in parameters

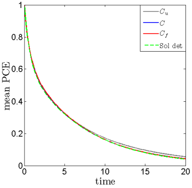
$\Delta^I(k_1)$	$\Delta^I(k_2)$	$\Delta^I(k_3)$
0.295	0.169	0.008

$$R_{20} = \begin{pmatrix} 1 & 0.9 & -0.37 \\ 0.9 & 1 & -0.45 \\ -0.37 & -0.45 & 1 \end{pmatrix}$$

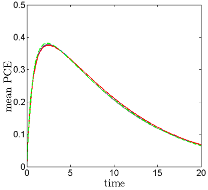
Correlations effect?

Example

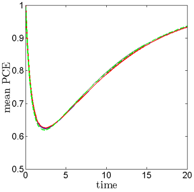
FIGURE : Means



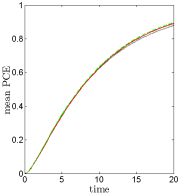
(A) Substrate S



(B) Complex C

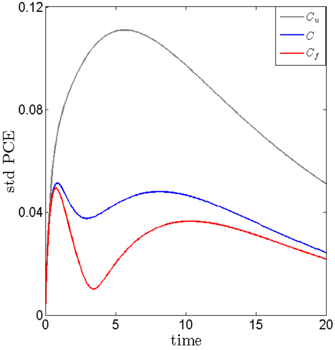


(C) Enzyme E

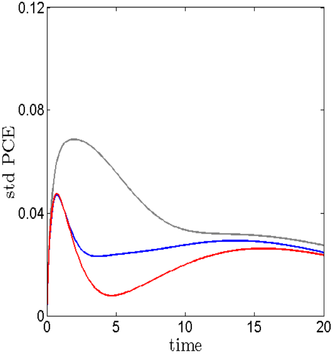


(D) Product P

FIGURE : Significant qualitative effect of correlation on standard deviation



(A) Substrate S



(B) Complex C

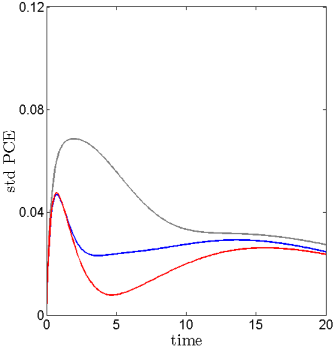


FIGURE : Significant qualitative effect of correlation on standard deviation

(A) Enzyme E

(B) Product P

Conclusions and Future works

- Study the UQ of the QoIs helps us to know the quality of the model predictions
- Sobol indices: ranking influence of RVs (uncertain parameters) $\rightarrow$ reduction
- Take correlations into account
- Develop Gauss quadrature based on new polynomials

Thank you