Identifying overparametrized biological wastewater treatment models with prior information regarding a subset of parameters

Chemical Engineering
Departmental Seminar

Queen's University
Kingston, ON

27 OCT 2011

Peter A. Vanrolleghem

Identifying overparametrized biological wastewater treatment models with prior information regarding a subset of parameters

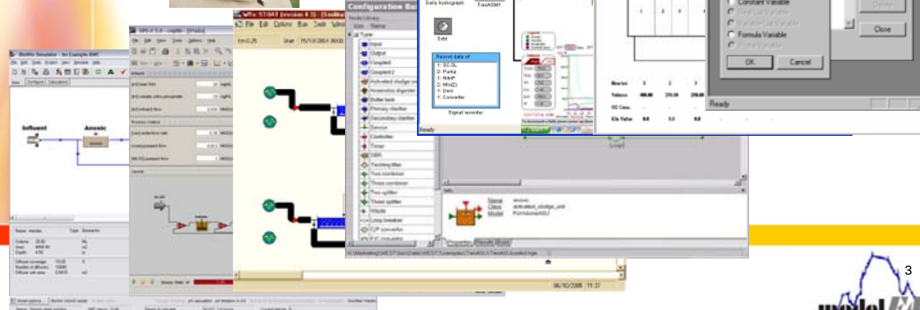
Chemical Engineering
Departmental Seminar

Queen's University
Kingston, ON

27 OCT 2011

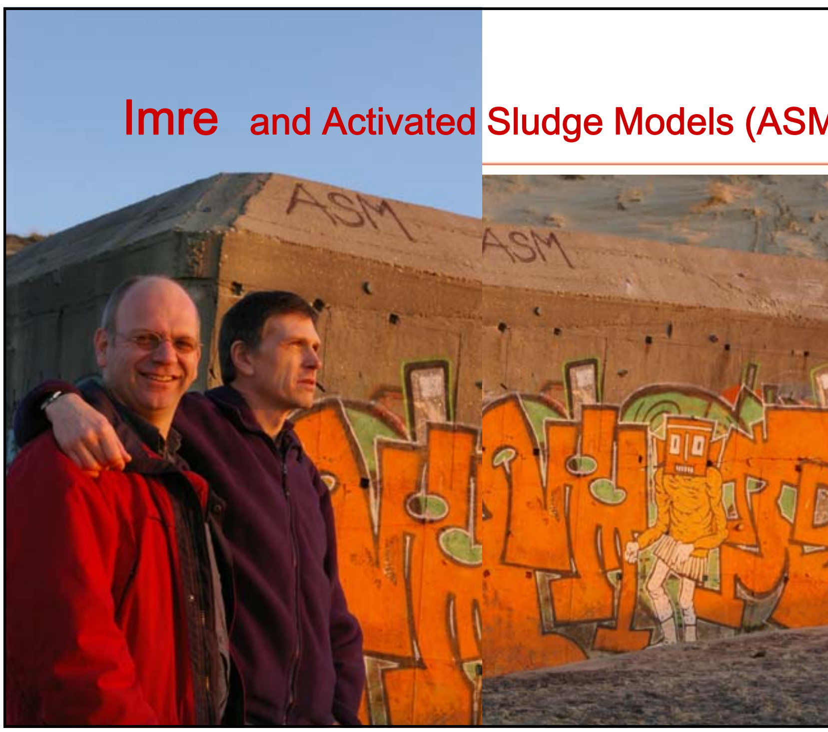
Modelling WWT is easy !

Thanks to Bill Gates & Imre
(& many others…)



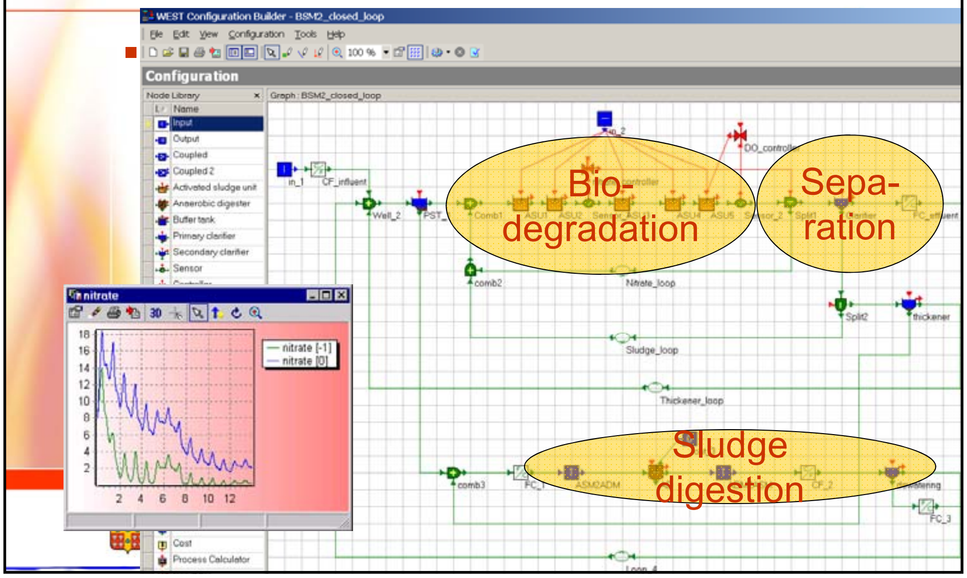
3

Imre and Activated Sludge Models (ASM)…



4

modelEAU Software: WEST



The Sky's the Limit!

Process Modeling Applications

1. Plant-Wide Mass Balance
2. Design Criteria for New Processes
3. Evaluate Plant Capacity
4. Aeration Optimization
5. Sensitivity Analyses
6. What-if Scenarios
7. Develop Control Algorithms
8. Training
9. Compare Process Alternatives
10. Process Design for Standard Processes

WEFTEC®.05

Modelling Wastewater Treatment

- Standard activity in consulting companies
- Design and optimisation of treatment plants
- Standardized « consensus » models are used
  - IWA's ASM (1, 2, 2d, 3 & 3P), ADM1
    - IWA: International Water Association
    - ASM: Activated Sludge Models
    - ADM: Anaerobic Digestion Model
    - Implemented in different softwares : GPS-X, WEST, Simba
  - BioWin model
    - Same basic structure but proprietary, EnviroSim, Hamilton

UNIVERSITÉ LAVAL 7 modelEAU

Modelled processes in WWT

- Hydraulics & Mixing (CSTR)
- Mass transfer (gas-liquid, diffusion)
- Sedimentation (particle separation)
- Biodegradation

is clearly getting most attention in WWT models

$$\frac{d\overline{C}}{dt} = \frac{Q}{V}\left(\overline{C_{in}} - \overline{C}\right) + \overline{r} \qquad \overline{r} = \left(Stoichiometry\right)^T \overline{\rho}$$

UNIVERSITÉ LAVAL 8 modelEAU

The Gujer matrix presentation

- Compact presentation of biokinetic model

Continuity →

Mass Balance ↓

Component → i / j Process ↓	1 X_B	2 S_S	3 S_O	Process Rate, ρ_j $[ML^{-3}T^{-1}]$
1 Growth	1	$-\frac{1}{Y}$	$-\frac{1-Y}{Y}$	$\frac{\hat{\mu} S_S}{K_S + S_S} X_B$
2 Decay	-1		-1	bX_B
Observed Conversion Rates $ML^{-3}T^{-1}$	$r_i = \sum_j r_{ij} = \sum_j \nu_{ij}\rho_j$			Kinetic Parameters: Maximum specific growth rate: $\hat{\mu}$ Half-velocity constant: K_S Specific decay rate: b
Stoichiometric Parameters: True growth yield: Y	Biomass [M(COD) L^{-3}]	Substrate [M(COD) L^{-3}]	Oxygen (negative COD) [M(−COD) L^{-3}]	

UNIVERSITÉ LAVAL

modelEAU

« Standard models » in use today

Model	Year	WWTP#	Processes	Components	Parameters*
'ASM0'	1987	C	2	3	4
ASM1	1987	C,N	9	13	19
ASM2	1994	C,N,P	20	19	65
ASM2d	1999	C,N,P	22	19	67
ASM3	1999	C,N	13	13	36
ASM3P	2001	C,N,P	24	17	71
ADM1	2002	X	28	36	96
BioWin	2007	C,N,P,X	51	62	470

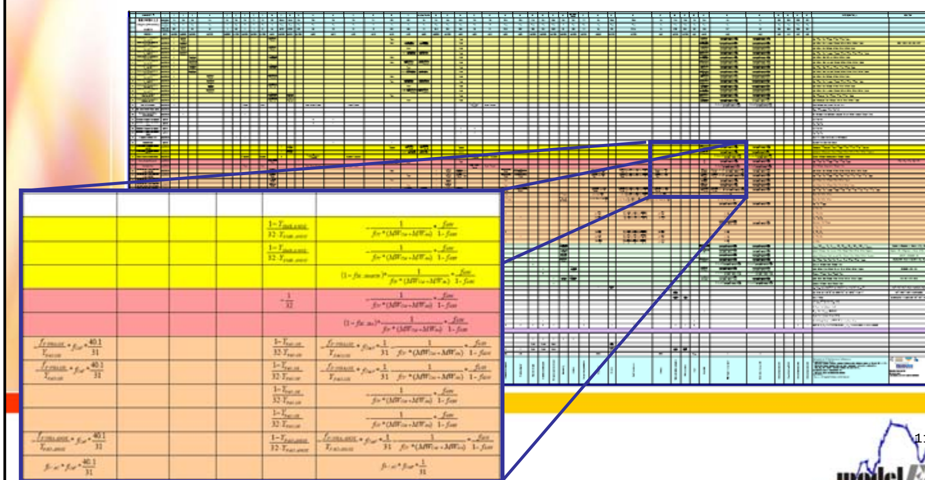
#C= COD-removal, N= N-removal, P= P-removal, X = Sludge digestion
**Parameters: stoichiometric, kinetic & composition parameters*

UNIVERSITÉ LAVAL

modelEAU

The Gujer matrix presentation

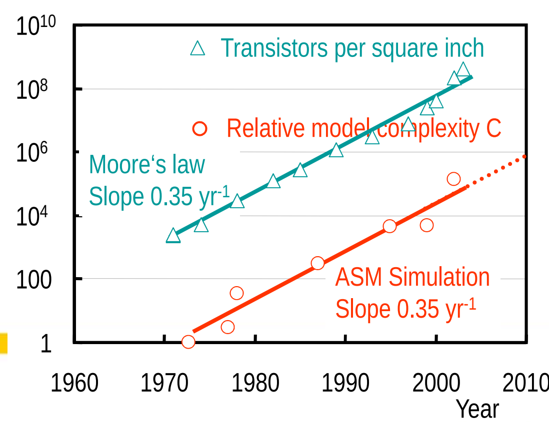
- BioWin model matrix



Gujer's law

- Model complexity measure:

  Complexity = $\#_{components}$ x $\#_{processes}$ x $\#_{compartments}$



UNIVERSITÉ LAVAL

Gujer's law

- We model what we can simulate in a period:
  - a coffee break
  - a night
  - a weekend
- What are the virtuel experiments we do?
  - Optimization
  - Scenario analysis
  - Sensitivity analysis
  - Uncertainty analysis

Embarrasingly Parallel Computational Problems
(Justin Babendreier, EPA)

UNIVERSITÉ LAVAL

13

Gujer's law

We can break this law!

Grid computing/clusters/distribution over idle PCs

EPA's SuperMUSE cluster (2002)

UNIVERSITÉ LAVAL

14

Gujer's law

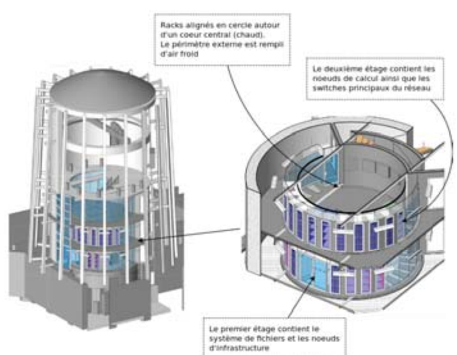
- Cluster computing @ Laval
- Université Laval's Colosse
  - Cluster computer
    97th in world Top 500 (Jun11)
  - 8000 compute cores
    24 TB RAM
    HDD: 4 DVD/s (17 GB/s)
  - In an old particle accelerator
  - Water & Energy ...

UNIVERSITÉ LAVAL

15

Gujer's Law

- Cluster computing @ Laval



UNIVERSITÉ LAVAL

16

Gujer's Law

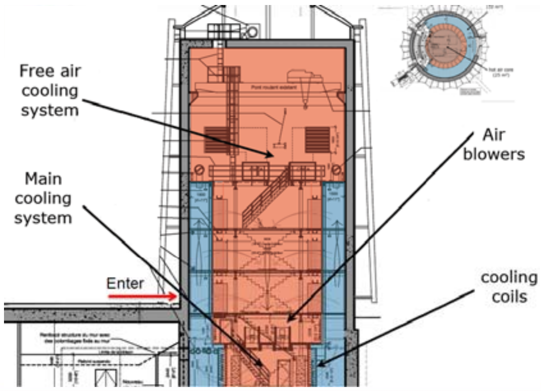
- Cluster computing @ Laval

UNIVERSITÉ LAVAL

17

Gujer's Law

- Cluster computing @ Laval



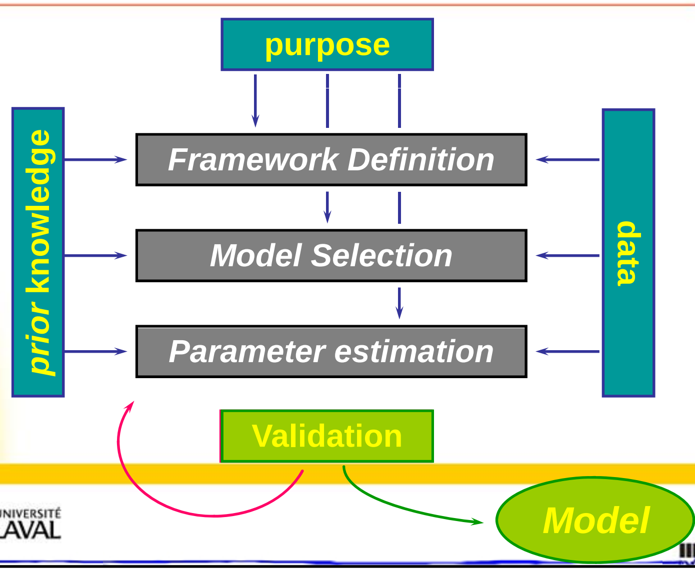
UNIVERSITÉ LAVAL

18

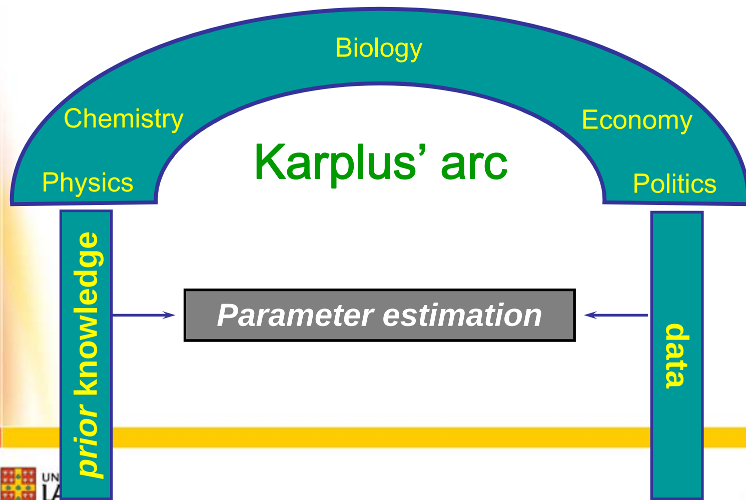
Contents

- Background on WWT modelling
- The overparametrization problem
- A motivating example
- Solving the identifiability problem
  - 4 approaches
  - use of prior information on (some of) the parameters
- The overparametrization problem revisited

The model building exercise



The model building exercise



21

The Identifiability Problem

- Structural identifiability
  - Given perfect data, can I get parameter values?

$$y = a.x_1 + b.x_2 + c.(x_1+x_2)$$

For measurement set $\{y,x_1,x_2\}$: $\theta_{ident} = \{a+c,b+c\}$

  - <u>Result:</u> A set of identifiable parameter combinations
  - One wants values for all parameters
    ==> Give values to some of them (priors, defaults)
  - Nonlinear analysis, only feasible for simple models
    Petersen et al (2003) Simplified method for Gujer matrix models

22

The Identifiability Problem

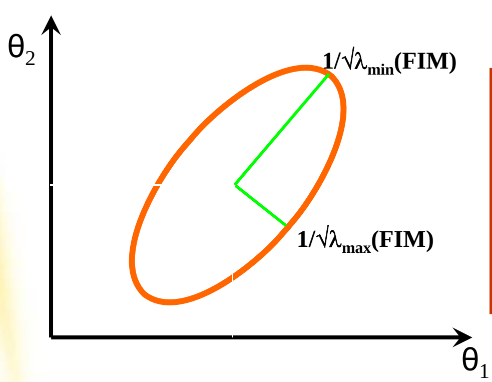
- Practical identifiability
  - Given noise-corrupt data, can I get parameter values?
  - Fisher Information Matrix analysis
    = inverse of Parameter Estimation Covariance Matrix

$$FIM = \left[\left(\frac{\partial y}{\partial \theta} \right)^T Q \left(\frac{\partial y}{\partial \theta} \right) \right]$$

= weighted sensitivity of all measured variables
to the parameters to be estimated

The Identifiability Problem

- Interpretation of the Fisher Information Matrix:



FIM-properties

D:	Determinant = Volume
modE:	Condition Nr = Correlation
E:	Longest axis

Contents

- Background on WWT modelling
- The overparametrization problem
- A motivating example
- Solving the identifiability problem
  - 4 approaches
  - use of prior information on (some of) the parameters
- The overparametrization problem revisited

The Identifiability Problem: Monod

- Monod equation is everywhere in biokinetics!

$$\rho_i = \mu \cdot \frac{S}{K+S} \cdot \frac{K}{K+S} \cdot X_{BM}$$

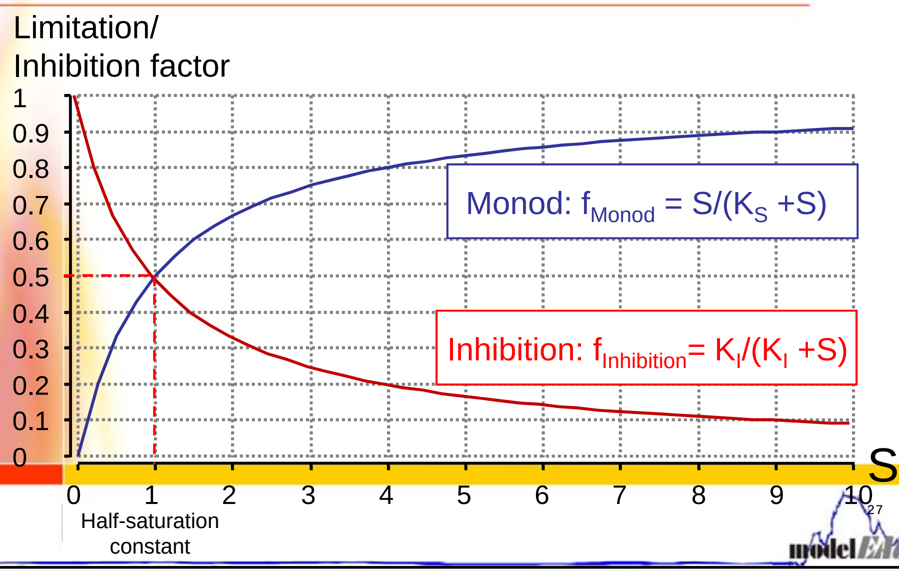
Maximum rate

Monod term (limitation)

Inhibition term

Biomass concentration

The Identifiability Problem: Monod



27

The Identifiability Problem: Monod

- Monod equation is everywhere in biokinetics!

$$\rho_i = \mu \cdot \frac{S}{K+S} \cdot \frac{K}{K+S} \cdot X_{BM}$$

Maximum rate

Monod term (limitation)

Inhibition term

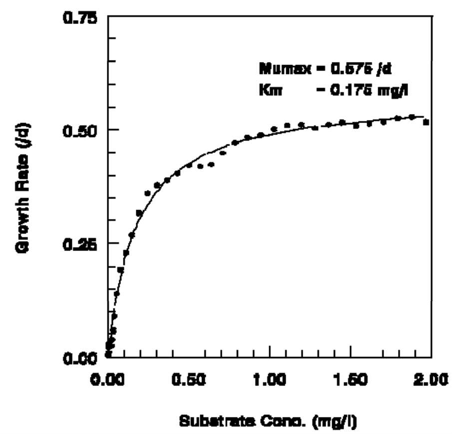
Biomass concentration

- Holmberg (1982) : Identifiabiliy problem !

UNIVERSITÉ LAVAL

28

The Identifiability Problem: Monod



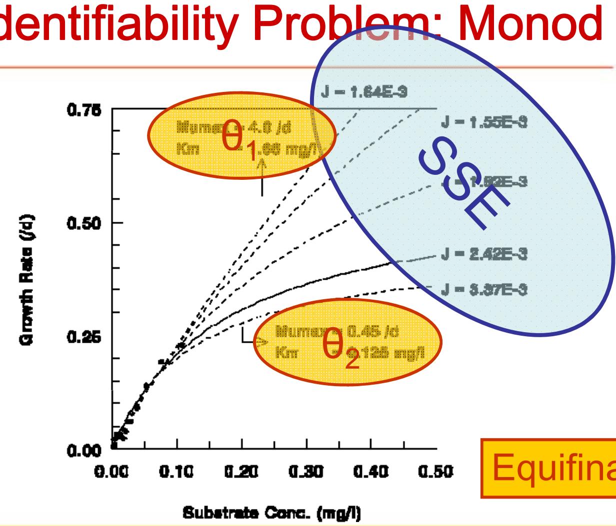
Unique parameter values

Dochain & Vanrolleghem (2001)

UNIVERSITÉ LAVAL

29

The Identifiability Problem: Monod



Equifinality !

Dochain & Vanrolleghem (2001)

UNIVERSITÉ LAVAL

30

The Identifiability Problem: Monod

- Same experimental data, acceptable fit with a large set of equifinal parameter sets
- μ_{max} = 4.0 d^{-1}, K_m = 1.9 mg.L^{-1}
  - Fast growth rate, small reactor volumes allowed
  - High effluent pollutant concentrations
- μ_{max} = 0.4 d^{-1}, K_m = 0.1 mg.L^{-1}
  - Slow growth rate, large reactor volumes necessary
  - Low effluent pollutant concentrations

UNIVERSITÉ LAVAL

31

modelEAU

The Identifiability Problem

- What to do ?
- Which parameter set to select?
- Important differences in decision depending on the selected parameter set!
- Consulting engineers make decisions with these models. How do they handle this?

UNIVERSITÉ LAVAL

32

modelEAU

Contents

- Background on WWT modelling
- The overparametrization problem
- A motivating example
- Solving the identifiability problem
  - 4 approaches
  - use of prior information on (some of) the parameters
- The overparametrization problem revisited

Dealing with the problem

- Four different approaches are observed:
  1. Just live with the problem and estimate all θ
  2. Reduce the model
  3. Get additional, informative data
  4. Select identifiable parameters

Approach 1: Just live with it

- It's a « NO-problem »
- « Our optimization algorithms are very capable and will find the best parameter values »
- Practice shows, they don't!
  - Local minima, no convergence
  - Equifinal parameter sets
- But the decisions are different (see example)!

UNIVERSITÉ LAVAL

35

modelEAU

Approach 1: Just live with it (cont'd)

- No use of prior information on the parameters

→ **Parameter estimation** ← data

- So, how can we integrate prior knowledge still?
- Grau et al. (2007)
  - Add a penalty term to the objective function
  - punishing deviations from initial estimates (« priors ») that do not improve the fit of the model to the data

UNIVERSITÉ LAVAL

36

modelEAU

Approach 1: Just live with it (cont'd)

- No use of prior information on the parameters

→ **Parameter estimation** ← data

- So, how can we integrate prior knowledge still?
- Omlin & Reichert (1997)
  « On the usefulness of overparameterized ecological models »
  - Bayesian approach to estimate the parameters
  - Disadvantage: Computational burden

UNIVERSITÉ LAVAL

37

modelEAU

Approach 2: Reduce the model

- Apply the parsimony principle to find a model of which all parameters are identifiable given the available data
- A lot of prior knowledge is thrown away !
  - Mechanisms
  - Interrelationships
  - Parameter values/ranges (prior distributions)
- What about decision-making ?

UNIVERSITÉ LAVAL

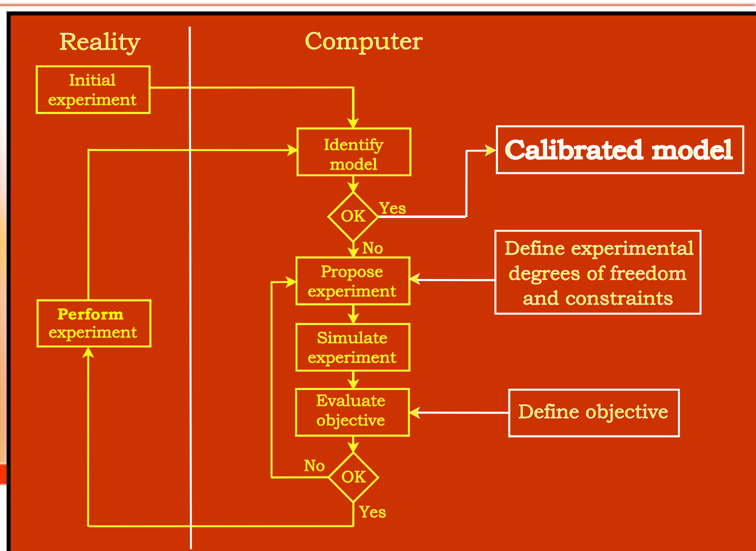
38

modelEAU

Approach 3: Add informative data

- Add the missing data
  that allow identifying all model parameters
- Many proposals for dedicated experiments
- Too much work though to perform them all
- Optimal Experimental Design methodology is applied, based on the Fisher Information Matrix

Approach 3: Add informative data



Approach 4: Select identifiable θ's

- Don't reduce the model
- Keep the whole parameter set
- Don't get new, expensive data
- Use prior knowledge on parameter values

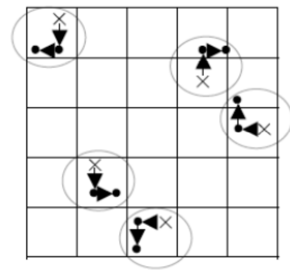
- Select the identifiable parameters
- Estimate their values from the limited data
- Keep the other parameters at their prior values

Approach 4: Select identifiable θ's

- Four methods
  for selection of identifiable parameter set
  - Weijers & Vanrolleghem (1997): Systems Analysis
  - Brun et al. (2002): Regression Theory
  - van Griensven et al. (2006): Sensitivity Analysis
  - Ruano et al. (2007): Expert Selection

Approach 4: Select identifiable θ's

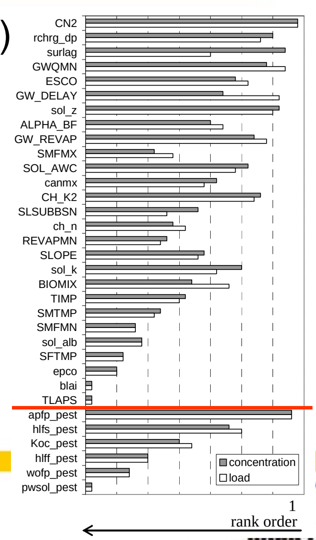
- van Griensven et al. (2006)
  - LHS of parameter space
  - Perform sensitivity analysis in each point
  - Average out local sensitivities
  - Rank parameters according to averaged sensitivities
  - Select the first *N*



43

Approach 4: Select identifiable θ's

- van Griensven et al. (2006)
  - Typical result
  - Holvoet et al. (2006)
  - SWAT-model for simulation of pesticides
    - Hydrological parameters
    - Quality parameters



44

Approach 4: Select identifiable θ's

- Weijers & Vanrolleghem (1997)
  - Select *N* parameters based on average sensitivities
  - Calculate the D- and modified E criteria of the FIM for all possible combinations of parameters in a parameter set of a certain size *P*
  - Rank parameter subsets
  - Perform this exercise for different set sizes *P*
  - Select the parameter subset with acceptable D and modified E criteria values

UNIVERSITÉ LAVAL

45

Approach 4: Select identifiable θ's

- Weijers & Vanrolleghem (1997): ASM1

Determinant	Cond. Number	n	Parameter subset
1.44 10^4	4.795	2	fX_I, μ_A20
11.0 10^4	35.1	3	fX_I, fS_I, μ_A20
50.2 10^4	110	4	$b_H20, fX_I, fS_I, \mu_A20$
104.1 10^4	295	5	$b_H20, fX_I, fS_I, \mu_A20, \eta_g$
61.5 10^4	634	6	$b_H20, fX_{BH}, fX_I, fS_I, \mu_A20, \eta_g$
24.8 10^4	1356	7	$b_H20, fX_{BH}, fX_I, fS_I, \mu_A20, \mu_H20, \eta_g$
10.5 10^4	2400	8	$b_H20, fX_{BH}, fX_I, fS_I, \mu_A20, \mu_H20, Y_A, \eta_g$

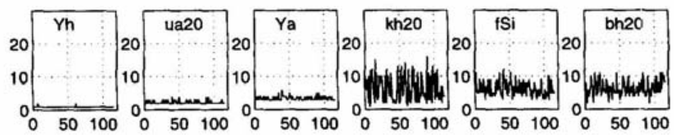
  - Practically, 5 parameters could be estimated from typical noise corrupted data collected at the plant

UNIVERSITÉ LAVAL

46

Approach 4: Select identifiable θ's

- Weijers & Vanrolleghem (1997)
  - Sensitivities are a local property
  - Ranking of parameters based on sensitivities as well
  - Example for ASM1
    - 115 LHS samples
    - Parameter rank for each set of parameter values:

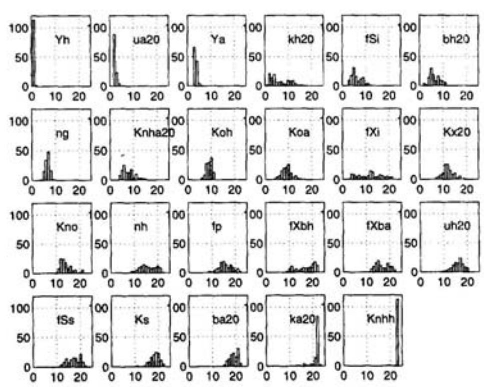


UNIVERSITÉ LAVAL

47

Approach 4: Select identifiable θ's

- Weijers & Vanrolleghem (1997)
  - Histogram of parameter sensitivity ranks



UNIVERSITÉ LAVAL

48

Approach 4: Select identifiable θ's

- Weijers & Vanrolleghem (1997)
  - Effect on identifiable parameter subset?
  - In 100 out of 115 cases,
    the same subset would have been selected
  - Subset composition is
    not substantially changed by prior estimates

UNIVERSITÉ LAVAL

modelEAU 49

Approach 4: Select identifiable θ's

- Brun et al. (2002)
  - Calculate relative sensitivities s_{ij} for all parameters
  - Calculate sensitivity measure δ^{msqr}

$$\delta_j^{msqr} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} s_{ij}^2}$$

  - Calculate collinearity index γ (λ_K is eigenvalue of $S^T{}_K \cdot S_K$)

= FIM !

$$\gamma_K = \frac{1}{\sqrt{\min \tilde{\lambda}_K}} \qquad 1 < \lambda_K < \textit{infinity}$$

  - Select largest parameter subset with γ_K smaller than 10

UNIVERSITÉ LAVAL

modelEAU 50

Approach 4: Select identifiable θ's

- Brun et al. (2002)
- Evaluation by Ruano et al. (2007)
  - Full-scale data
  - ASM2 model
  - 69 parameters to be estimated
  - 30 parameters selected on the basis of δ^{msqr}

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\ PHA}$
3					AUT
4					$K_{O2\ AUT}$
5	1.43E+05	0.333	60.0	1.13	α_{SF}, α_{SI}, $k_{la\ min}$, $\theta_{q\ fe}$, i_{NXS}
6	5.94E+05	1	44.2	1.43	α_{SF}, $\theta_{b\ PAO}$, α_{SI}, $k_{la\ min}$, $k_{la\ carr}$, $K_{O2\ AUT}$
7	2.04E+06	3	29.7	1.72	α_{SF}, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
8	5.85e+06	9	18.0	1.96	α_{SF}, $\theta_{b\ AUT}$, α_{SI}, q_{PHA}, $\theta_{K\ X}$, $k_{la\ min}$, $k_{la\ carr}$, i_{NXS}
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\ PP}$, α_{SI}, $\theta_{K\ h}$, b_H, $k_{la\ min}$, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}
10	3.00E+07	39	4.5	3.36	α_{SF}, $\theta_{q\ PP}$, $\theta_{b\ H}$, α_{XI}, $\theta_{b\ PP}$, $k_{la\ min}$, b_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
11	5.46E+07	67	2.2	3.53	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{b\ H}$, $\theta_{b\ PP}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$ b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
12	8.65E+07	106	0.9	3.86	α_{SF}, $\theta_{q\ PP}$, Y_H, α_{XI}, $\theta_{K\ X}$, $k_{la\ min}$, K_{PS}, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
13	1.20E+08	150	0.39	5.23	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{\mu\ PAO}$, $\theta_{b\ H}$, $\theta_{b\ AUT}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}

Parameter subset size

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\ PHA}$
3	4.06E+03	0.010	87.2	1.03	α_{SF}, $\theta_{\mu\ PAO}$, $K_{NH4\ AUT}$
4	2.74E+04	0.083	71.3	1.08	α_{SI}, $k_{la\ min}$, Y_{PAO}, $K_{O2\ AUT}$
5	1.43E+05	0.333	60.0	1.13	α_{SF}, α_{SI}, $k_{la\ min}$, $\theta_{q\ fe}$, i_{NXS}
6	5.94E+05	1	44.2	1.43	α_{SF}, $\theta_{b\ PAO}$, α_{SI}, $k_{la\ min}$, $k_{la\ carr}$, $K_{O2\ AUT}$
7					in, Y_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
8					PHA, $\theta_{K\ X}$, $k_{la\ min}$, $k_{la\ carr}$,
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\ PP}$, α_{SI}, $\theta_{K\ h}$, b_H, $k_{la\ min}$, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}
10	3.00E+07	39	4.5	3.36	α_{SF}, $\theta_{q\ PP}$, $\theta_{b\ H}$, α_{XI}, $\theta_{b\ PP}$, $k_{la\ min}$, b_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
11	5.46E+07	67	2.2	3.53	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{b\ H}$, $\theta_{b\ PP}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$ b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
12	8.65E+07	106	0.9	3.86	α_{SF}, $\theta_{q\ PP}$, Y_H, α_{XI}, $\theta_{K\ X}$, $k_{la\ min}$, K_{PS}, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
13	1.20E+08	150	0.39	5.23	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{\mu\ PAO}$, $\theta_{b\ H}$, $\theta_{b\ AUT}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}

Total number of subset combinations tested

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\ PHA}$
3	4.06E+03	0.010	87.2	1.03	α_{SF}, $\theta_{\mu\ PAO}$, $K_{NH4\ AUT}$
4	2.74E+04	0.083	71.3	1.08	α_{SI}, $k_{la\ min}$, Y_{PAO}, $K_{O2\ AUT}$
5					fe, i_{NXS}
6					a min, $k_{la\ carr}$, $K_{O2\ AUT}$
7	2.04E+06	3	29.7	1.72	α_{SF}, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
8	5.85e+06	9	18.0	1.96	α_{SF}, $\theta_{b\ AUT}$, α_{SI}, q_{PHA}, $\theta_{K\ X}$, $k_{la\ min}$, $k_{la\ carr}$, i_{NXS}
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\ PP}$, α_{SI}, $\theta_{K\ h}$, b_H, $k_{la\ min}$, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}
10	3.00E+07	39	4.5	3.36	α_{SF}, $\theta_{q\ PP}$, $\theta_{b\ H}$, α_{XI}, $\theta_{b\ PP}$, $k_{la\ min}$, b_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
11	5.46E+07	67	2.2	3.53	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{b\ H}$, $\theta_{b\ PP}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$ b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
12	8.65E+07	106	0.9	3.86	α_{SF}, $\theta_{q\ PP}$, Y_H, α_{XI}, $\theta_{K\ X}$, $k_{la\ min}$, K_{PS}, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
13	1.20E+08	150	0.39	5.23	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{\mu\ PAO}$, $\theta_{b\ H}$, $\theta_{b\ AUT}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}

The computational time (d)

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\,PHA}$
3	4.06E+03	0.010	87.2	1.03	α_{SF}, $\theta_{\mu\,PAO}$, $K_{NH4\,AUT}$
4	2.74E+04	0.083	71.3	1.08	α_{SI}, $k_{la\,min}$, Y_{PAO}, $K_{O2\,AUT}$
5	1.43E+05	0.333	60.0	1.13	α_{SF}, α_{SI}, $k_{la\,min}$, $\theta_{q\,fe}$, i_{NXS}
6	5.94E+05	1	44.2	1.43	α_{SF}, $\theta_{b\,PAO}$, α_{SI}, $k_{la\,min}$, $k_{la\,carr}$, $K_{O2\,AUT}$
7	2.04E+06	3	29.7	1.72	α_{SF}, α_{SI}, $\theta_{K\,h}$, $k_{la\,min}$, Y_{PAO}, i_{NXS}, $K_{NH4\,AUT}$
8	5.85e+06	9	18.0	1.96	α_{SF}, $\theta_{b\,AUT}$, α_{SI}, q_{PHA}, $\theta_{K\,X}$, $k_{la\,min}$, $k_{la\,carr}$, i_{NXS}
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\,PP}$, α_{SI}, $\theta_{K\,h}$, b_H, $k_{la\,min}$, b_{PAO},
10					α_{XI}, $\theta_{b\,PP}$, $k_{la\,min}$, b_{PAO}, i_{NXS},
11	5.46E+07	67	2.2	3.53	α_{SF}, $\theta_{q\,PP}$, $\theta_{\mu\,AUT}$, $\theta_{b\,H}$, $\theta_{b\,PP}$, α_{SI}, $\theta_{K\,h}$, $k_{la\,min}$ b_{PAO}, $K_{O2\,AUT}$, i_{NXS}, $K_{NH4\,AUT}$
12	8.65E+07	106	0.9	3.86	α_{SF}, $\theta_{q\,PP}$, Y_H, α_{XI}, $\theta_{K\,X}$, $k_{la\,min}$, K_{PS}, Y_{PAO}, b_{PAO}, $K_{O2\,AUT}$, i_{NXS}, $K_{NH4\,AUT}$
13	1.20E+08	150	0.39	5.23	α_{SF}, $\theta_{q\,PP}$, $\theta_{\mu\,AUT}$, $\theta_{\mu\,PAO}$, $\theta_{b\,H}$, $\theta_{b\,AUT}$, α_{SI}, $\theta_{K\,h}$, $k_{la\,min}$, Y_{PAO}, b_{PAO}, $K_{O2\,AUT}$, i_{NXS}

The ratio of identifiable subsets to the total number of subsets

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\,PHA}$
3	4.06E+03	0.010	87.2	1.03	α_{SF}, $\theta_{\mu\,PAO}$, $K_{NH4\,AUT}$
4	2.74E+04	0.083	71.3	1.08	α_{SI}, $k_{la\,min}$, Y_{PAO}, $K_{O2\,AUT}$
5	1.43E+05	0.333	60.0	1.13	α_{SF}, α_{SI}, $k_{la\,min}$, $\theta_{q\,fe}$, i_{NXS}
6	5.94E+05	1	44.2	1.43	α_{SF}, $\theta_{b\,PAO}$, α_{SI}, $k_{la\,min}$, $k_{la\,carr}$, $K_{O2\,AUT}$
7	2.04E+06	3	29.7	1.72	α_{SF}, α_{SI}, $\theta_{K\,h}$, $k_{la\,min}$, Y_{PAO}, i_{NXS}, $K_{NH4\,AUT}$
8	5.85e+06	9	18.0	1.96	α_{SF}, $\theta_{b\,AUT}$, α_{SI}, q_{PHA}, $\theta_{K\,X}$, $k_{la\,min}$, $k_{la\,carr}$, i_{NXS}
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\,PP}$, α_{SI}, $\theta_{K\,h}$, b_H, $k_{la\,min}$, b_{PAO}, $K_{O2\,AUT}$, i_{NXS}
10	3.00E+07	39	4.5	3.36	α_{SF}, $\theta_{q\,PP}$, $\theta_{b\,H}$, α_{XI}, $\theta_{b\,PP}$, $k_{la\,min}$, b_{PAO}, i_{NXS}, $K_{NH4\,AUT}$
11					$\theta_{b\,H}$, $\theta_{b\,PP}$, α_{SI}, $\theta_{K\,h}$, AUT, i_{NXS}, $K_{NH4\,AUT}$
12	8.65E+07	106	0.9	3.86	α_{SF}, $\theta_{q\,PP}$, Y_H, α_{XI}, $\theta_{K\,X}$, $k_{la\,min}$, K_{PS}, Y_{PAO}, b_{PAO}, $K_{O2\,AUT}$, i_{NXS}, $K_{NH4\,AUT}$
13	1.20E+08	150	0.39	5.23	α_{SF}, $\theta_{q\,PP}$, $\theta_{\mu\,AUT}$, $\theta_{\mu\,PAO}$, $\theta_{b\,H}$, $\theta_{b\,AUT}$, α_{SI}, $\theta_{K\,h}$, $k_{la\,min}$, Y_{PAO}, b_{PAO}, $K_{O2\,AUT}$, i_{NXS}

The minimum collinearity index

Approach 4: Select identifiable θ's

Size	Combinations	*CT* (d)	γ<10 (%)	γ_{min}	Parameters subset for γ_{min}
2	4.35E+02	0.003	95.6	1.00	α_{SF}, $\theta_{q\ PHA}$
3	4.06E+03	0.010	87.2	1.03	α_{SF}, $\theta_{\mu\ PAO}$, $K_{NH4\ AUT}$
4	2.74E+04	0.083	71.3	1.08	α_{SI}, $k_{la\ min}$, Y_{PAO}, $K_{O2\ AUT}$
5	1.43E+05	0.333	60.0	1.13	α_{SF}, α_{SI}, $k_{la\ min}$, $\theta_{q\ fe}$, i_{NXS}
6	5.94E+05	1	44.2	1.43	α_{SF}, $\theta_{b\ PAO}$, α_{SI}, $k_{la\ min}$, $k_{la\ carr}$, $K_{O2\ AUT}$
7	2.04E+06	3	29.7	1.72	α_{SF}, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
8	5.85e+06	9	18.0	1.96	α_{SF}, $\theta_{b\ AUT}$, α_{SI}, q_{PHA}, $\theta_{K\ X}$, $k_{la\ min}$, $k_{la\ carr}$, i_{NXS}
9	1.43E+07	19	9.6	2.81	α_{SF}, $\theta_{q\ PP}$, α_{SI}, $\theta_{K\ h}$, b_H, $k_{la\ min}$, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}
10	3.00E+07	39	4.5	3.36	α_{SF}, $\theta_{q\ PP}$, $\theta_{b\ H}$, α_{XI}, $\theta_{b\ PP}$, $k_{la\ min}$, b_{PAO}, i_{NXS}, $K_{NH4\ AUT}$
11	5.46E+07	67	2.2	3.53	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{b\ H}$, $\theta_{b\ PP}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$ b_{PAO}, $K_{O2\ AUT}$, i_{NXS}, $K_{NH4\ AUT}$
12	8.65E+07	106	0.9	3.86	[illegible] $k_{la\ min}$, K_{PS}, Y_{PAO}, [illegible] 4 AUT
13	1.20E+08	130	0.39	3.23	α_{SF}, $\theta_{q\ PP}$, $\theta_{\mu\ AUT}$, $\theta_{\mu\ PAO}$, $\theta_{b\ H}$, $\theta_{b\ AUT}$, α_{SI}, $\theta_{K\ h}$, $k_{la\ min}$, Y_{PAO}, b_{PAO}, $K_{O2\ AUT}$, i_{NXS}

The 'best' identifiable parameter subset

Approach 4: Select identifiable θ's

- Ruano et al. (2007)
  - Parameter subsets used by « experts »
  - How would they have performed for this case study?

Protocol	Parameter subset size	γ
Melcer *et al.*, (2003)	7	3.7
Takács (2006)	4	1.2
Makinia *et al.*, (2006)	13	46.6
Insel *et al.*, (2006)	13	51.1
Hulsbeek *et al.*, (2002)*	11	26.4
García-Usach *et al.*, (2006)	8	15.3

Not identi-fiable

UNIVERSITÉ LAVAL

modelEAU

Approach 4: Select identifiable θ's

- Comparison
  - Only sensitivity (OAT: One-at-a-Time)
  - Integrated sensitivity measures (FIM)
    - Variance
    - Correlation/Collinearity
  - Nonlinear models: Sensitivity to parameter values
    - Averaged over the entire parameter space
    - Evaluation of local results on parameter selection
- Overall result:
  Parameter estimates conditional on « priors »

UNIVERSITÉ LAVAL

59

modelEAU

Contents

- Background on WWT modelling
- The overparametrization problem
- A motivating example
- Solving the identifiability problem
  - 4 approaches
  - use of prior information on (some of) the parameters
- The overparametrization problem revisited

UNIVERSITÉ LAVAL

60

modelEAU

Aren’t we forgetting something?

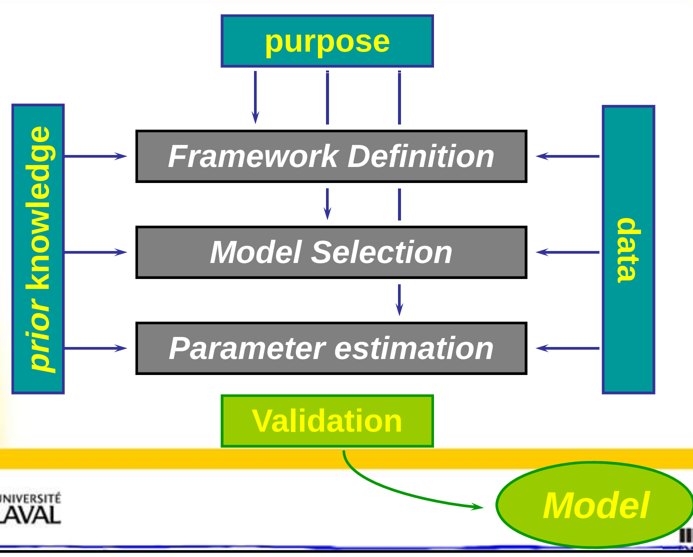
- I’ve been talking about:

UNIVERSITÉ LAVAL

61

modelEAU

Aren’t we forgetting something?



UNIVERSITÉ LAVAL

62

modelEAU

Aren't we forgetting something?

- Most often we're not interested in the parameter values themselves
- We want to make model-based decisions
- We should estimate those parameters that affect those decisions
- We should apply the methods to select them, and use prior estimates if data are insufficient

UNIVERSITÉ LAVAL

63

Take home messages

- Background on WWT modelling
- The overparametrization problem
- A motivating example
- Solving the identifiability problem
  - 4 approaches
  - use of prior information on (some of) the parameters
- The overparametrization problem revisited

UNIVERSITÉ LAVAL

64

Acknowledgements

Canada Research Chair
in Water Quality Modelling

BIOMATH – Ghent University, Belgium
(Gurkan Sin, Dirk De Pauw, Ann van Griensven)

CEIT – San Sebastian, Spain
(Paloma Grau, Eduardo Ayesa)

65