Potential and limitations of white & black-box modelling concepts for process optimization of SBR WW treatment

Chemical Eng. Seminar

Ecole Polytechnique Montréal

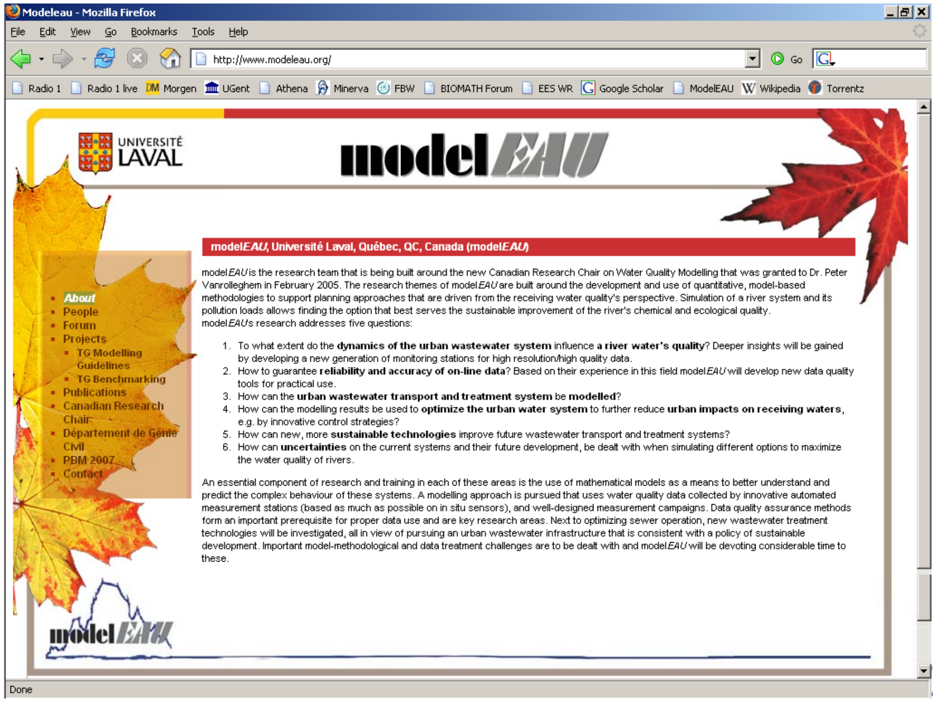
22 Feb 2006

Peter A. Vanrolleghem, Kris Villez & Gurkan Sin

UNIVERSITEIT GENT

UNIVERSITÉ LAVAL

modelEAU



Overview

- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

Introduction

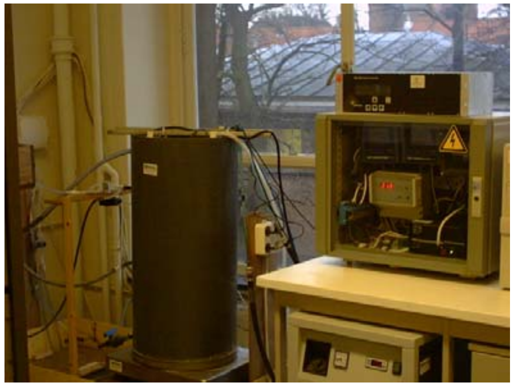
- Activated sludge systems:
  - Underpinning of the microbial community is *not fully* deciphered (yet…)
  - Involves many interacting processes …
  - Observed behaviour is dynamic & complex
  - *Mechanistic modelling* has proven useful for better understanding & improving operation…
  - *Data driven models* are promising techniques for process monitoring (FDI)

Overview

- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

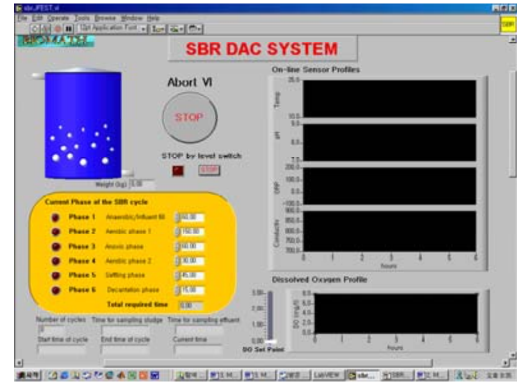
BIOMATH's pilot SBR

- Performs N & P removal
- Synthetic WW
- V_{max} 80 L
- 15 ^{o}C
- SRT 10 d
- HRT 12h



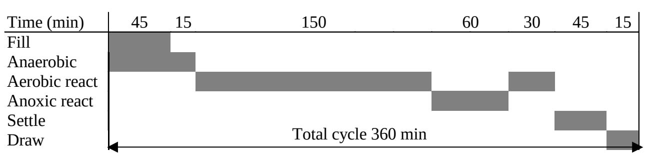
BIOMATH's pilot SBR

- Fully automated (LabView)
- 5 years of data
- Objectives
  - Stable sludge for Sedifloc project
  - Model-based Optimization of N & P removal
  - Fault detection and diagnosis

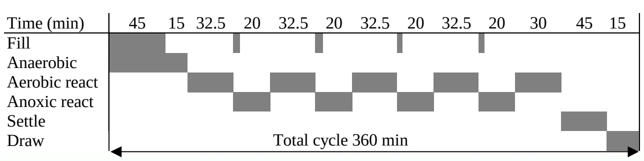


BIOMATH's pilot SBR

First operation (DO = 2 mg/l) (run for 2 years)

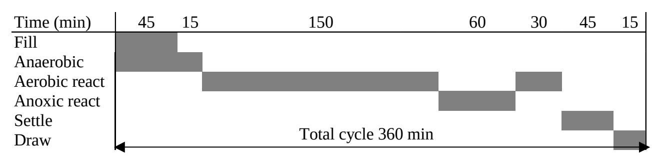


2nd operation (DO = 0.5 mg/l) (run for 3.5 months)

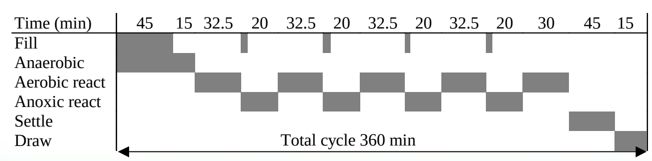


BIOMATH's pilot SBR

First operation (DO = 2 mg/l) (run for 2 years)



3rd operation (DO = 1 mg/l) (run for 1 year)



Overview

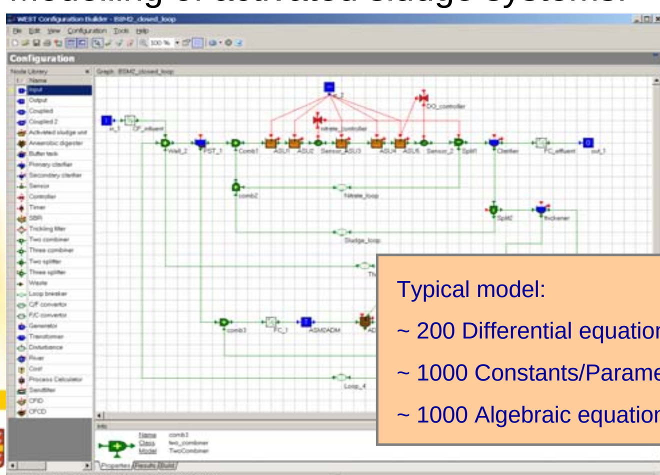
- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

Mechanistic modelling

- Modelling of activated sludge systems
  - Model structure: internal description of the system
  - Usually constructed using
    - available prior knowledge
    - observed system behavior
  - Selection of an appropriate model structure is *very important* to successfully model the system

Mechanistic modelling

- Modelling of activated sludge systems:



Typical model:

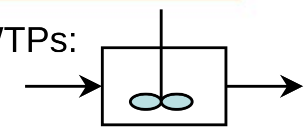
~ 200 Differential equations

~ 1000 Constants/Parameters

~ 1000 Algebraic equations

Mechanistic modelling

- Mechanistic models for WWTPs:
  - Mass balance for compound:

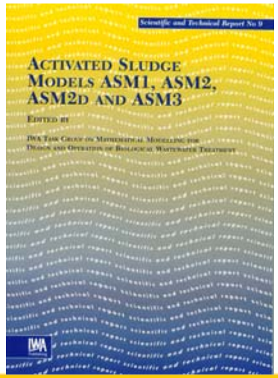


$$\frac{dM}{dt} = \frac{d(VC)}{dt} = \underbrace{Q_{in}C_{in} - Q_{out}C_{out}}_{\text{transport}} + \underbrace{rV}_{\text{conversion}}$$

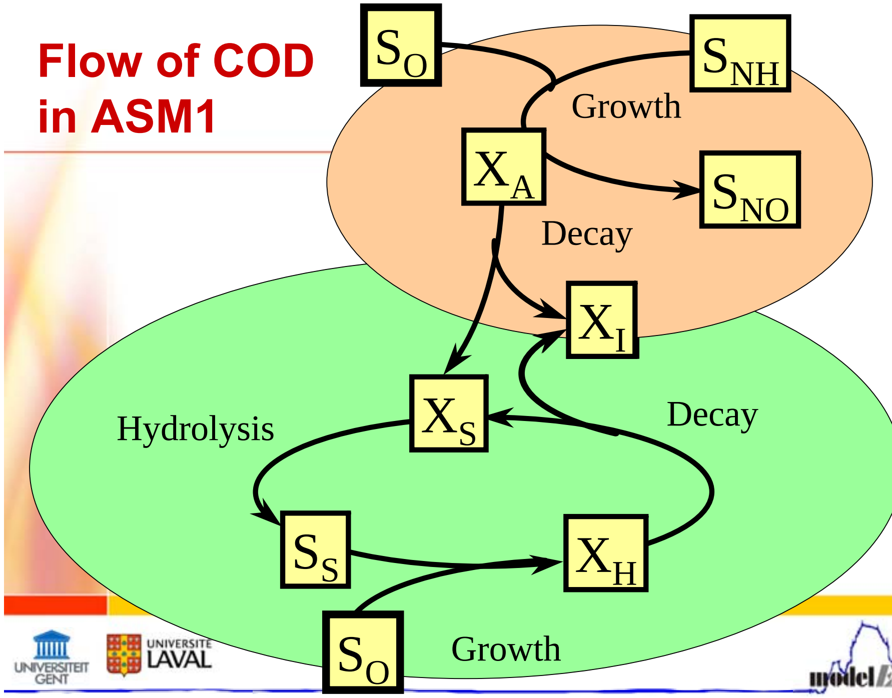
  - with
    - M: Mass of compound in system (g)
    - C: Concentration of compound (g/m3)
    - V: Volume of system (m3)
    - Q: flow rate (m3/h)
    - r: volumetric conversion rate (g/m3.h)

Mechanistic modelling

- Mechanistic models for WWTPs:
  - Henze, M., Gujer, W., Mino T., & van Loosdrecht, M. (2000)
  - Activated Sludge Models
    - ASM1
    - ASM2 & ASM2D
    - ASM3
  - Scientific and Technical Report No. 9
  - IWA Publishing, London.



Flow of COD in ASM1



Mechanistic modelling

- Mechanistic model structure for WWT:

Continuity →

Mass Balance ↓

Component → i	1 X_B	2 S_S	3 S_O	Process Rate, ρ_j $[ML^{-3}T^{-1}]$
j Process ↓				
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) $[M(-COD) L^{-3}]$	

Component → i		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
j	Process ↓	S_S	S_I	S_{NH4}	S_{NH3}	S_{NO2}	S_{NO3}	S_{HPO4}	S_{H2PO4}	S_{O2}	S_{CO2}	S_{HCO3}	S_{CO3}	S_H	S_{OH}	S_{Ca}	X_H	X_{N1}	X_{N2}	X_{ALG}	X_{CON}	X_S	X_I	X_P	X_{II}
(1a)	Aerobic Growth of Heterotrophs with NH4	-		?				?		-	+			?			1								
(1b)	Aerobic Growth of Heterotrophs with NO3	-					-	?		-	+			?			1								
(2)	Aerobic Resp. of Het.			+				+		-	+			-			-1						+		
(3a)	Anoxic Growth of Heterotrophs with NO3	-				+	-	?			+			?			1								
(3b)	Anoxic Growth of Heterotrophs with NO2	-				-		?			+			?			1								
(4)	Anoxic Resp. of Het.			+			-	+			+			-			-1						+		
(5)	Growth of 1st-stage Nitrifiers			-		+		-		-	-			+				1							
(6)	Aerobic Respiration of 1st-stage Nitrifiers			+				+			+			-				-1					+		
(7)	Growth of 2nd-stage Nitrifiers						-	-			-			-					1						
(8)	Aerobic Respiration of 2nd-stage Nitrifiers			+				+		-	+			-					-1				+		
(9a)	Growth of Algae with NH4			-						+				-						1					
(9b)	Growth of Algae with NO3						-			+				-						1					
(10)	Aerobic Resp. of Algae			+				+		-	+			-						-1			+		
(11)	Death of Algae			(+)				(+)		(+)	?			?						-1		+	+		
(12a)	Growth of Cons. on XALG			(+)				(+)		-	?			?						-	1	+			
(12b)	Growth of Cons. on XS			(+)				(+)		-	?			?							1	-			
(12c)	Growth of Cons. on XH			(+)				(+)		-	?			?			-				1				
(12d)	Growth of Cons. on XN1			(+)				(+)		-	?			?				-			1				
(12e)	Growth of Cons. on XN2			(+)				(+)		-	?			?					-		1				
(13)	Aerobic Resp. of Cons.			+				+		-	+			-							-1		+		
(14)	Death of Consumers			(+)				(+)		(+)	?			?							-1	+	+		
(15)	Hydrolysis	+		(+)				(+)		(+)	?			?								-1			
(16)	Eq. CO2 <-> HCO3										-1	1		+											
(17)	Eq. HCO3 <-> CO3											-1	1	+											
(18)	Eq. H2O <-> H + OH													1	1										
(19)	Eq. NH4 <-> NH3			-1	1									+											
(20)	Eq. H2PO4 <-> HPO4							1	-1					+											
(21)	Eq. Ca <-> CO3												+			1									
(22)	Ads. of Phosphate							-1																1	
(23)	Des. of Phosphate							1																-1	

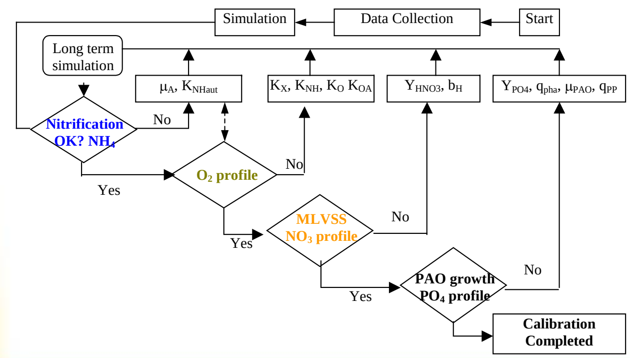
The horror matrix

Overview

- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

Model-based optimization

- BIOMATH calibration protocol was applied

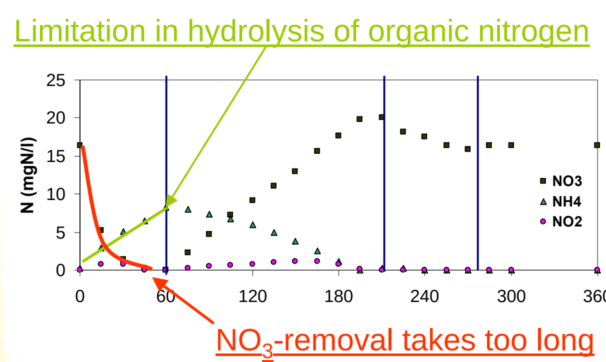


Model-based optimization (1)

- First iteration (basic operation)
  - Complete COD-removal
  - Complete nitrification
  - Incomplete denitrification (70 % N-removal)
  - 50 % P-removal (limited due to presence of NO_3)

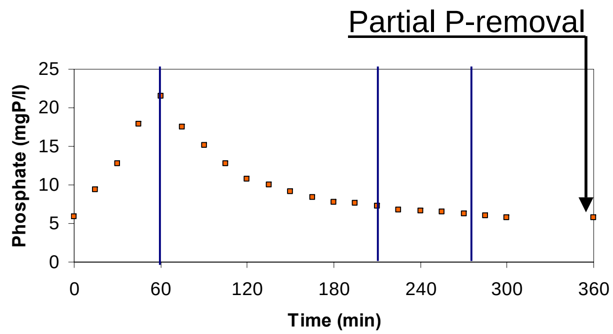
Model-based optimization (1)

- First iteration (basic operation)



Model-based optimization (1)

- First iteration (basic operation)



Model-based optimization (1)

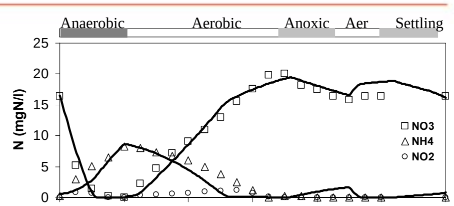
- First iteration (basic operation)

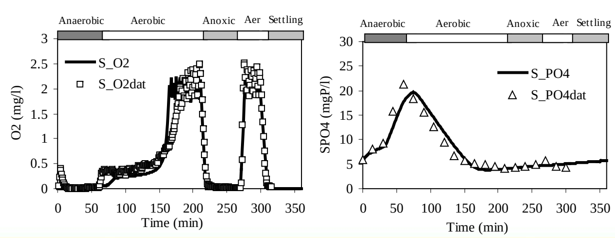
Prior knowledge (e.g. ASM2d)
+
System's observation
(Long-term performance + intensive in-cycle data)

Model structure change needed (ASM2dN)
(ASM2d is extended with ASM1 hydrolysis model)

Model-based optimization (1)

- First iteration fitting results:
  - NH_4, NO_3, NO_2
  - DO PO$_4$

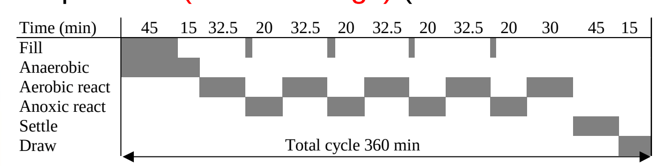




Model-based optimization (2)

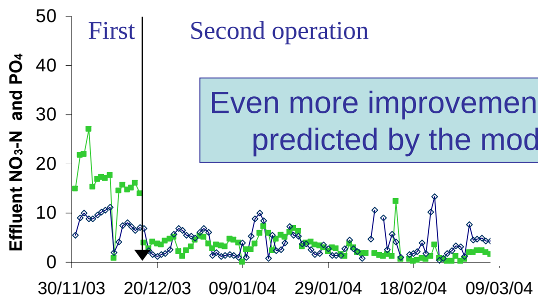
- With the calibrated ASM2dN model, process operation was optimized:
  - Low DO; Step-feed; short aerobic & anoxic phases…

2nd operation (DO = 0.5 mg/l) (run for 3.5 months)

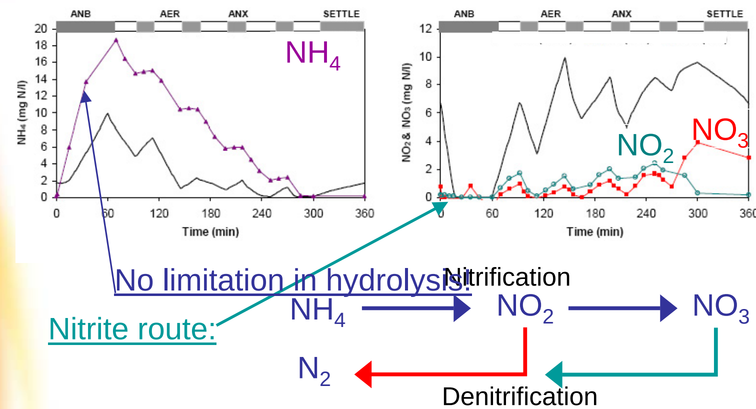


Model-based optimization (2)

- N-removal improved by 86%
- P-removal improved by 65%



Model-based optimization (2)



Model-based optimization (2)

- During 2nd operation ASM2dN fails :
  - No longer a limitation of hydrolysis of organic N
  - 2-step nitrification occurs
  - Nitrite route takes place
  - Nitrate removal is better (observed !)
  - No more inhibition of P-removal (observed !)

Model-based optimization (2)

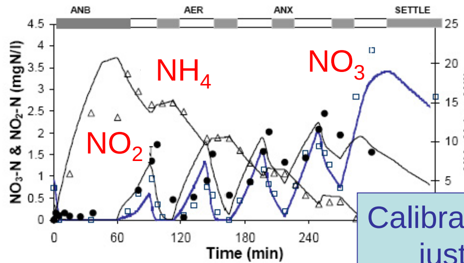
- Necessary model modifications → ASM2d2N

Modified processes of ASM2d (stoichiometry)

Processes	S_{NH}	S_{NO3}	S_{NO2}	S_{N2}	X_H	X_{NH}	X_{NO}
NH_4 oxidation	$-i_{NBM} - 1/Y_{NH}$		$1/Y_{NH}$			1	
NO_2 oxidation	$-i_{NBM}$	$1/Y_{NO}$	$-1/Y_{NO}$				1
NO_3 reduction	$-i_{NBM}$	$-(1-Y_{HNO3})/(1.14\ Y_{HNO3})$	$(1-Y_{HNO3})/(1.14\ Y_{HNO3})$		1		
NO_2 reduction	$-i_{NBM}$		$-(1-Y_{HNO2})/(1.72\ Y_{HNO2})$	$(1-Y_{HNO2})/(1.72\ Y_{HNO2})$	1		
Lysis of X_{NH}							-1
Lysis of X_{NO}						-1	

Model-based optimization (2)

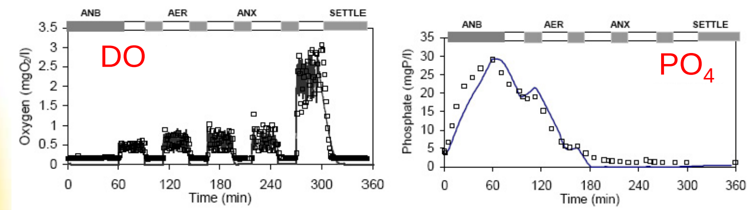
- ASM2d2N fitting results to in-cycle measurements:



Calibration is again just perfect !

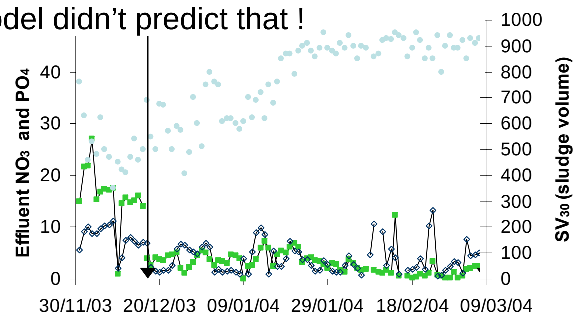
Model-based optimization (2)

- ASM2d2N fitting results to in-cycle measurements:



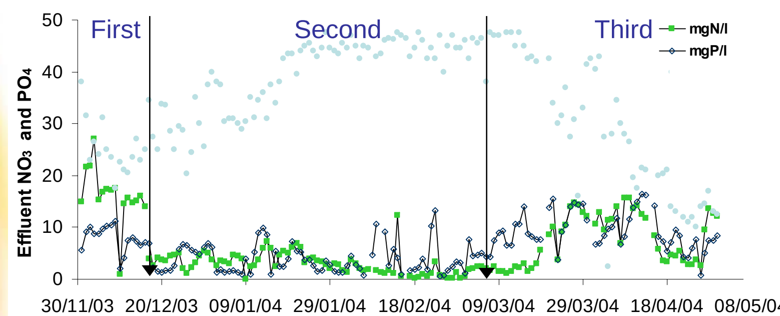
Model-based optimization (2)

- Nutrient removal is very good, but …
  - Severe sludge bulking occurs !
  - Model didn't predict that !



Model-based optimization (3)

- 3[rd] scenario: DO ↑ (from 0.5 to 1.0 mg/l) & Ca^{2+} ↑
- bulking solved, but N,P-removal worse again

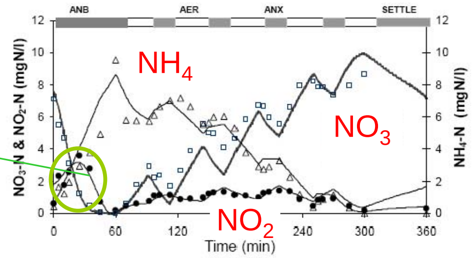


Model-based optimization (3)

- ASM2d2N performs well
- but re-calibration was needed !

Sequential denitrification

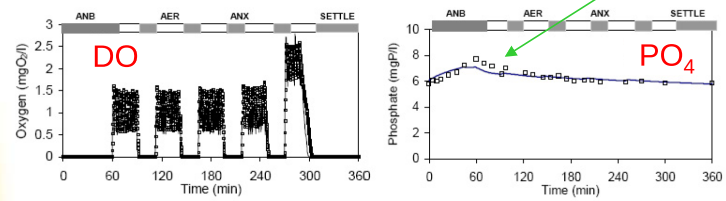
NO_2 build-up



Model-based optimization (3)

- ASM2d2N performs well
- but re-calibration was needed !

Loss of bio-P activity



Model-based optimization (3)

- Validation of the ASM2d2N showed:
  - The model structure remained valid, but …
  - Parameters of the model had to be changed

Conclusions

- The 3 models predicted the system dynamics to some extent :
  - Parallel to system changes (operation):
  - Model structure had to be changed twice
    - Hydrolysis
    - NO_2 route (nitrification & denitrification)
  - Parameters had to be changed every time
- Poor predictive power of mechanistic models,
- Not to mention prediction of bulking…

Conclusions

- The underlying reasons remain unclear, but could be :
  - Unaccounted input disturbances
  - Imperfect model structure
  - …
  - Or perhaps the system is too complex to *mechanistically* model!
- Biology was proven by DGGE analysis to change significantly after operation changes

Conclusions

- Models that validly describe system behaviour under a wide range of conditions are not available yet
- But models appear valid within certain (narrow?) boundaries, e.g. under certain operation conditions…
- and models help to understand the system and point to optimization approaches

Overview

- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

Black box modelling: Intro

Data drowning…

Black box modelling: Intro

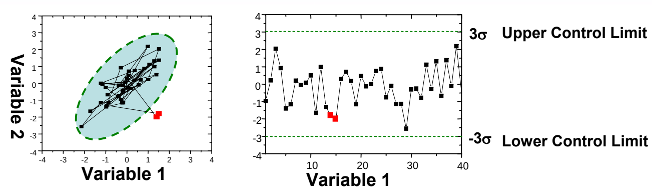
- Many data-driven approaches !
- Here we only consider multivariate statistical analysis:
  - Principal Component Analysis (PCA)
    → process monitoring (fault detection/diagnosis)
  - Partial Least Squares (PLS)
    → prediction in view of control
- Applied to the BIOMATH pilot SBR

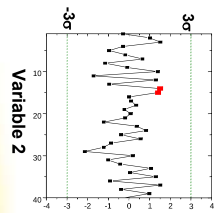
Process monitoring (PCA)

- Monitoring the state of the process: Statistical Process Control (SPC)
- Traditional SPC = Univariate SPC
  - One variable at a time, not efficient
  - Problem of correlation between variables

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

Process monitoring (PCA)





- The deviation is not detected unless the variables are combined
- Most variables are correlated
- The key to early fault detection is the correlation structure, not the original variables

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

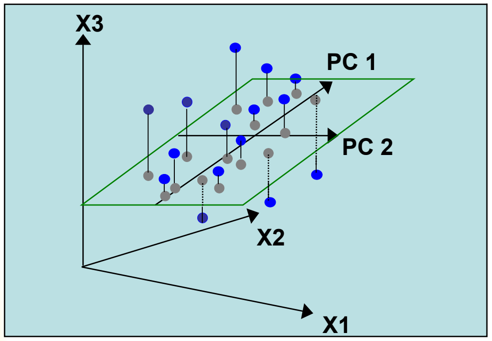
Process monitoring (PCA)

- Monitoring the state of the process: Statistical Process Control (SPC)
- Traditional SPC = Univariate SPC
  - One variable at a time, not efficient
  - Problem of correlation between variables
- Multivariate SPC
  - Account for interactions among variables
  - Detect upsets and find assignable causes

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

Process monitoring (PCA)

- Geometrical interpretation

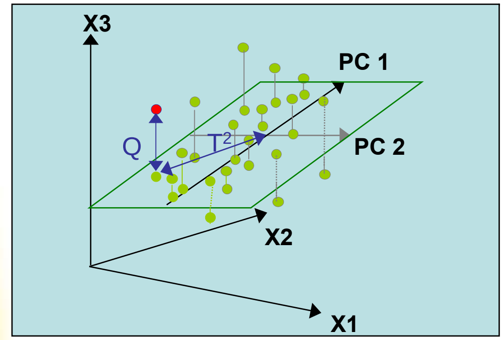


Loads (P): linear combinations of original process variables

Scores (T): projected coordinates of samples

Process monitoring (PCA)

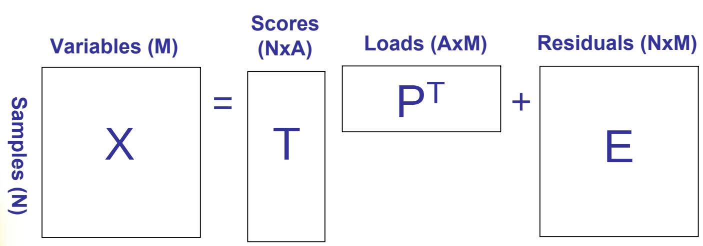
- Lack of Model Fit Statistics



Q: distance between the model plane and a sample

T^2: distance within model plane from a sample to the origin

Process monitoring (PCA)

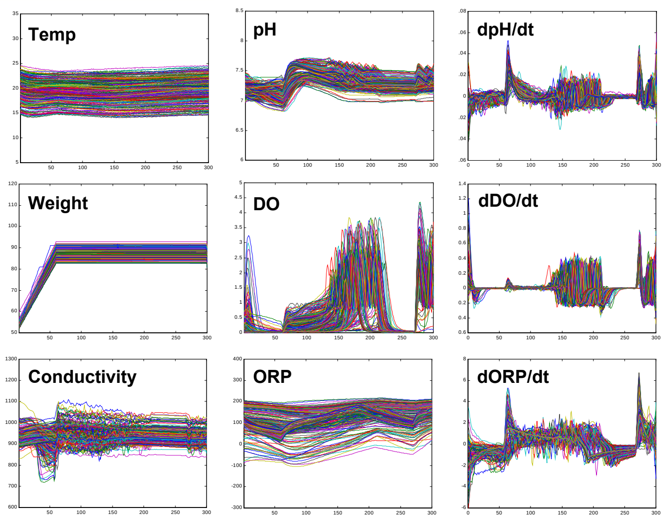


A: the number of principal components (A << M)

Process monitoring (PCA)

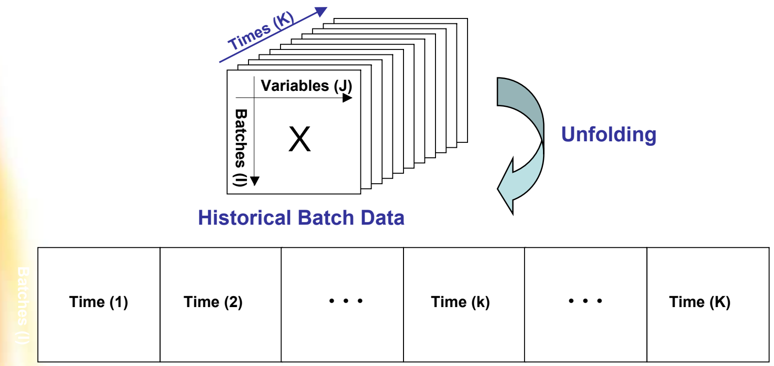
- Applying PCA models to the BIOMATH SBR
- Objective: Develop real-time monitoring
  - Detect the major sources of process disturbances
  - Useful to keep the sludge as stable as possible
- On the basis of simple on-line data, e.g.
  - pH, temperature, weight
  - conductivity
  - dissolved oxygen (DO)
  - oxidation reduction potential (ORP)

Process monitoring (PCA)



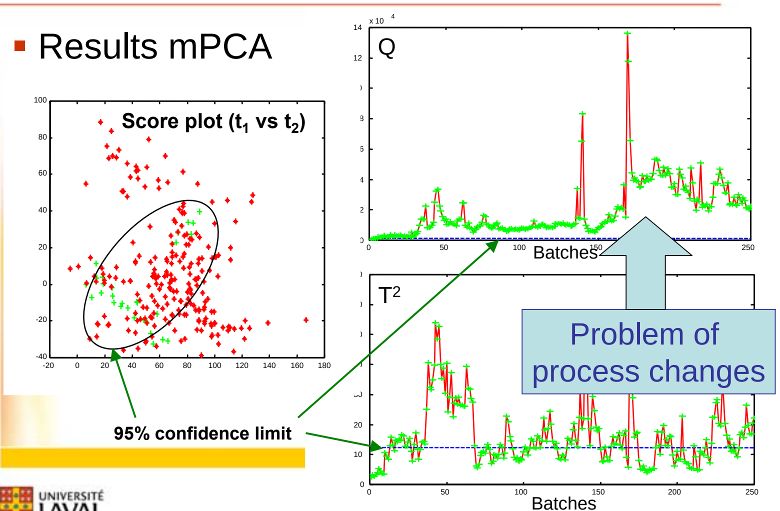
Process monitoring (PCA)

- Multi-way PCA model



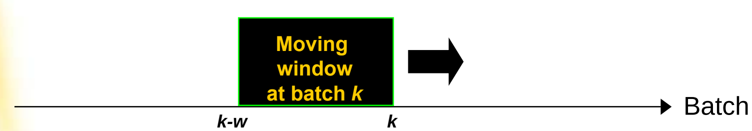
Process monitoring (PCA)

- Results mPCA



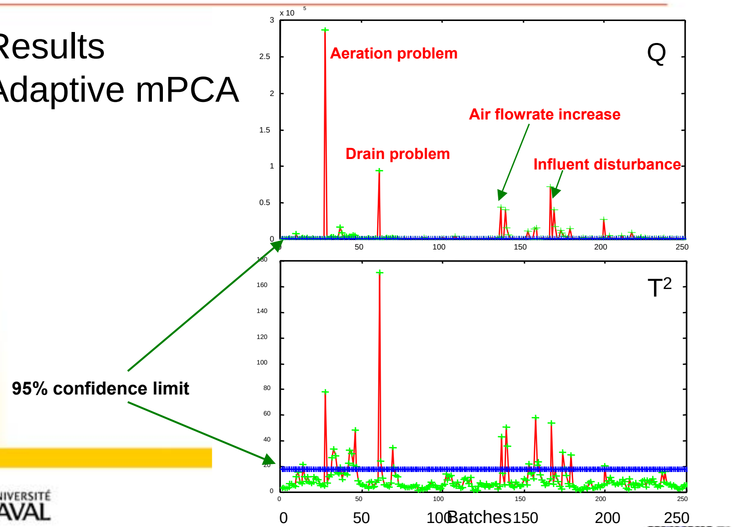
Process monitoring (PCA)

- To overcome the problem of changing process conditions:
- *Adaptive* multi-way PCA



Process monitoring (PCA)

- Results
  Adaptive mPCA



UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

Conclusions

- Adaptive Multi-way PCA
  provides more information than adaptive PCA
- Critical process disturbances
  are well captured in Q & T^2 plots
- Adaptive Multi-way PCA is a
  powerful tool for monitoring SBR processes

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

Overview

- Introduction
- BIOMATH's pilot SBR
- White box modelling
  - Modelling approach
  - Calibration/Validation in an optimization loop
- Black box modelling
  - Background on PCA/PLS multivariate analysis
  - Performance evaluation
- Conclusions

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

PLS modelling

- Why PLS ?
  - We want Y=f(X)
  - When dealing with collinear inputs (X), multivariate linear regression (MLR)

$$y = \mathbf{B} \,.\, \mathbf{x} = b_1 .\, x_1 + b_2 \,.\, x_2 + \ldots + b_n . x_n$$

will lead to the unbiased regression vector
but the estimated regression vectors will have
a high variance (very accurate, low precision)

UNIVERSITEIT GENT · UNIVERSITÉ LAVAL · modelEAU

PLS modelling

- PLS is one way to overcome this problem
  - It trades a bias
    with a decreased variance of the solution
    by reducing the dimension of the input space
    while minimizing the prediction error.

$$\mathbf{x} = \mathbf{t} \,.\, \mathbf{p}^t = t_1.\mathbf{p_1}^t + t_2.\mathbf{p_2}^t + \ldots + t_n\, \mathbf{p_n}^t$$
$$\mathbf{y} = \mathbf{u} \,.\, \mathbf{q}^t = u_1.\mathbf{q_1}^t + u_2.\mathbf{q_2}^t + \ldots + u_n\, \mathbf{q_n}^t$$

where: $u_i = b_i \,.\, t_i + h_i$; b_i : inner relation coefficient
h_i : inner model error

t_i (u_i) present the transformed (lower dimensional) input (output) data

PLS modelling

- Advantages:
  - Dimension reduction of input space
  - More robust estimates of regression vector(s)
- Disadvantages
  - Limited to linear regression conditions

Neural Net PLS modelling

- PLS is a linear method by definition and thus fails when the relationship between inputs (X) and outputs (Y) is non-linear in nature

Neural Net PLS modelling

- NNPLS tackles this problem by replacing the (linear) inner relationship coefficient in PLS by a 3-layer network (1 hidden layer).

$$\mathbf{x} = \mathbf{t} \,.\, \mathbf{p}^t = t_1.\mathbf{p_1}^t + t_2.\mathbf{p_2}^t + \ldots + t_n\, \mathbf{p_n}^t$$
$$\mathbf{y} = \mathbf{u} \,.\, \mathbf{q}^t = u_1.\mathbf{q_1}^t + t_2.\mathbf{q_2}^t + \ldots + u_n\, \mathbf{q_n}^t$$

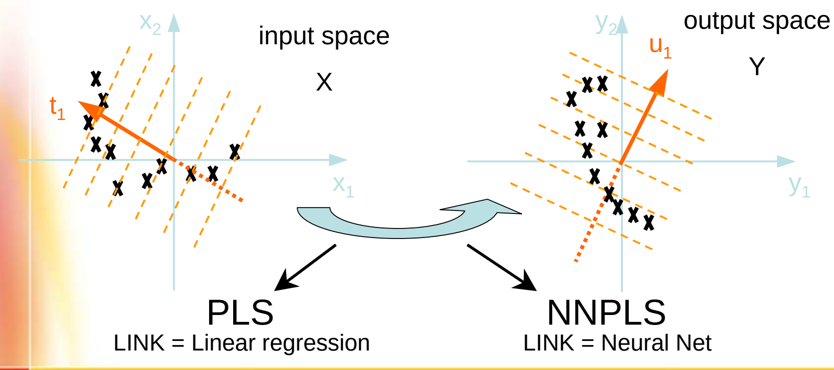
where: $u_i = NN_i(t_i) + h_i$; NN_i : inner neural net
h_i : inner model error

Neural Net PLS modelling

- Advantages:
  - Not restricted to linear regression problems
- Disadvantages
  - Additional parameters (number of nodes in hidden layer)
  - => model is more complex

Neural Net PLS modelling

- Comparison NNPLS / PLS



Kernel PLS modelling

- KPLS also tackles the problem of nonlinearity
- not by looking for a nonlinear relation Y=f(t)
- but by transforming the input space (X), prior to PLS modelling

Kernel PLS modelling

- The transformation is chosen such that the input data become "more linear" :

$$\mathbf{f} = \mathbf{\Phi(x)}$$

$$\mathbf{f} = \mathbf{t} \,.\, \mathbf{p}^t = t_1.\mathbf{p_1}^t + t_2.\mathbf{p_2}^t + \ldots + t_n\, \mathbf{p_n}^t$$

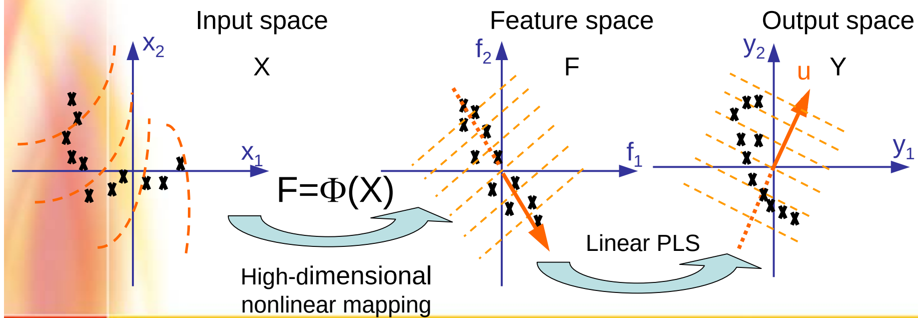
$$\mathbf{y} = \mathbf{u} \,.\, \mathbf{q}^t = u_1.\mathbf{q_1}^t + t_2.\mathbf{q_2}^t + \ldots + u_n\, \mathbf{q_n}^t$$

where: $u_i = b_i \,.\, t_i + h_i$; b_i : inner relation coefficient

h_i : inner model error

Kernel PLS modelling

- The transformation is chosen such that the input data become "more linear" :



Kernel PLS modelling

- In this work the Gaussian kernel function was applied to transform the input data:

$$f_{i,j} = \mathbf{\Phi}(\mathbf{x}_i,\mathbf{x}_j) = \mathbf{k}(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2 / d)$$

where: d = width of the Gaussian kernel function
(= extra tuning or meta-parameter)

Kernel PLS modelling

- Advantages:
  - nonlinear collinearity within the X-space is dealt with
  - No non-linear optimisation required
- Disadvantages
  - Larger computational demand (x10 – x100)
  - Models are hard to interpret
    (as the transformed inputs are hard to interpret)

PLS modelling: Results

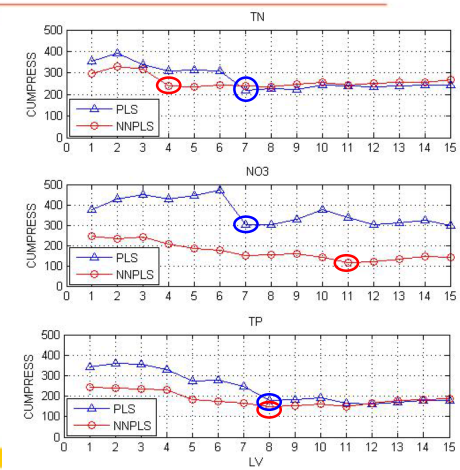
- <u>Objective:</u> predict SBR effluent quality
  - Total nitrogen
  - NO_3
  - PO_4
- using on-line data (1600 batches; ΔT=1 min)
  - pH, temperature, weight
  - conductivity
  - dissolved oxygen (DO)
  - oxidation reduction potential (ORP)

PLS modelling: Results

- For PLS and NNPLS:
  - Degree of freedom: # of latent variables
  - Selection based on cross-validation CUMPRESS = Sum of SSE values

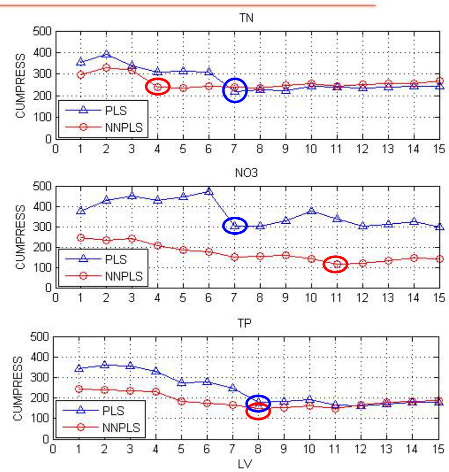
PLS modelling: Results

- # LV's for prediction of
  - Total N
  - NO_3
  - PO_4



PLS modelling: Results

- # LV's for prediction of
  - Total N
  - NO_3
  - PO_4
- NNPLS gives:
  - A lower dimension of the model (TN)
  - Better prediction (NO_3, PO_4)

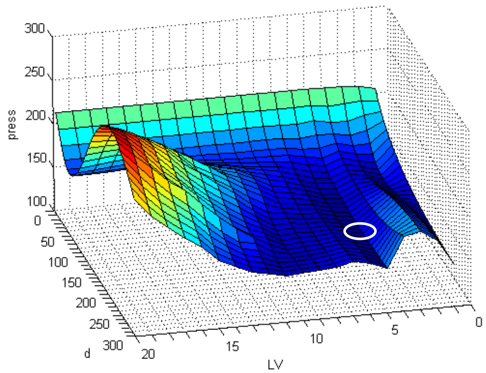


PLS modelling: Results

- For Kernel PLS :
  - Degree of freedom:
    - # of latent variables
    - width of kernel function d
  - Selection based on cross-validation CUMPRESS = Sum of SSE values

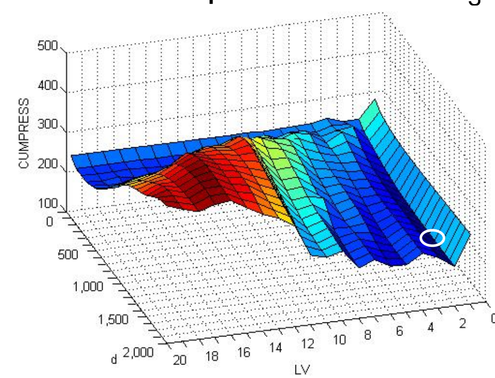
PLS modelling: Results

- # LV's and d for prediction of Total N



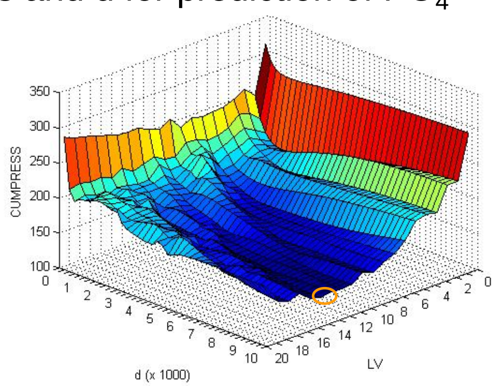
PLS modelling: Results

- # LV's and d for prediction of NO_3

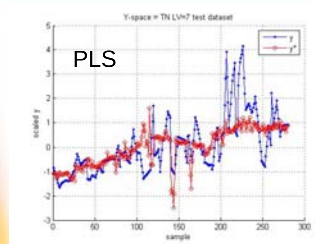


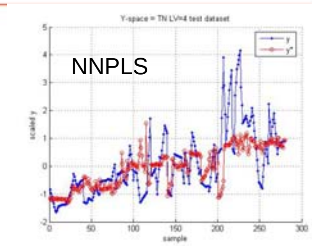
PLS modelling: Results

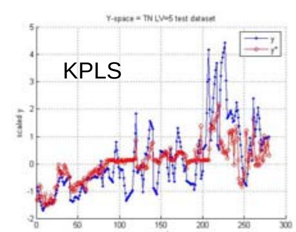
- # LV's and d for prediction of PO_4



PLS modelling: Results (Total N)

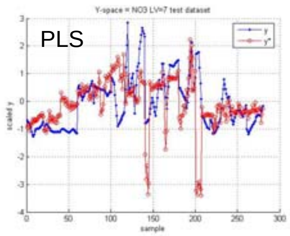


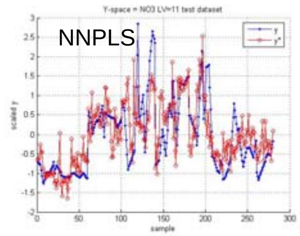


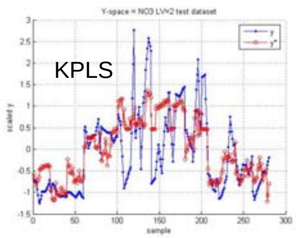


NO MODEL IS SATISFYING

PLS modelling: Results (NO_3)



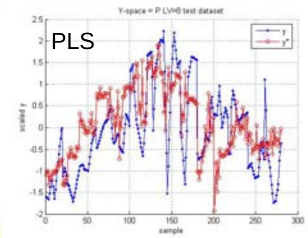


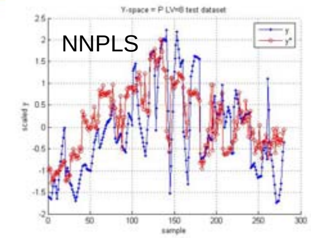


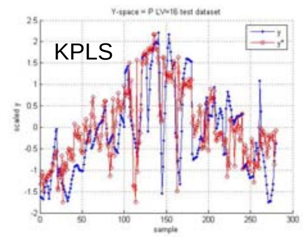
NNPLS catches dynamics best

KPLS delivers poor prediction

PLS modelling: Results (PO_4)







NNPLS delivers slight improvement

KPLS performs better but at a cost of 16 LV's, PLS and NNPLS : 8 LV's

PLS modelling: Summary

	output	PLS	NNPLS	KPLS
quality	TN	-	-	-
	NO3	-	++	-
	P	+	+	++
cumpress	TN	220	235	147
	NO3	303	116	198
	P	179	150	142
LV's	TN	7	4	5
	NO3	7	11	2
	P	8	8	16

- TN: unacceptable (dynamics)
- NO_3: NNPLS is only satisfactory model
- PO_4: NNPLS is selected (KPLS model needs too many LV's: 16)

Conclusions

- Still, none of the models is really satisfying
- What did we miss?
  - The inputs data do not describe the process
  - The data are treated as independent observations
    → In fact, they represent a time series
    Autocorrelation should be accounted for
  - The data stem from a large time window (14 months)
    → Equipment, operation and biological changes may not permit a unique (overall) model

Peter Vanrolleghem
Canada Research Chair in Water Quality Modelling

model*EAU*, Département de Génie Civil
Pavillon Pouliot, Université Laval
Québec G1k 7P4, QC, Canada

Tel: +1 418 656 5085
Fax: +1 418 656 2928
E-mail: peter.vanrolleghem@modelEAU.org

and of course also, please visit us at :
www.model*EAU*.org