Les données massives pour le système d'eaux urbaines: Opportunités, défis et dangers

Objectif numérique 2017

Québec, QC

1 novembre 2017

Peter VANROLLEGHEM

Canada Research Chair in Water Quality Modelling

UNIVERSITÉ LAVAL

modelEAU

Water Quality (WQ) data – use?

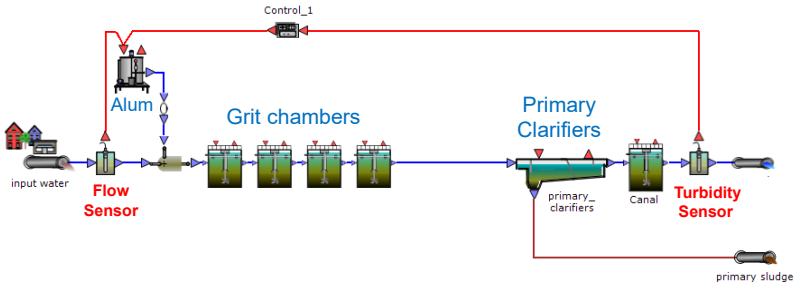
- Example from primary treatment Québec City

UNIVERSITÉ LAVAL

modelEAU

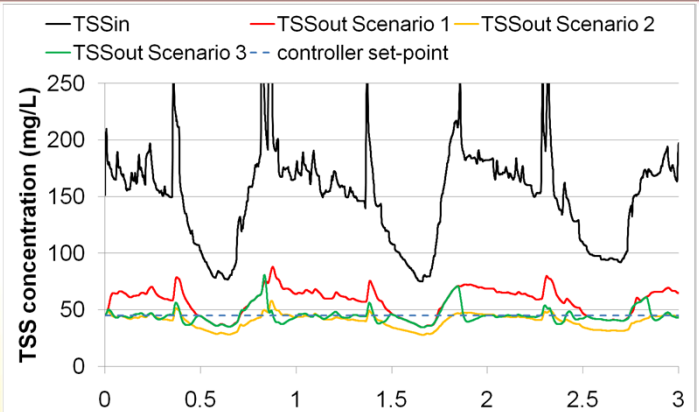
WQ data use – A motivating example

- Chemically Enhanced Primary Treatment
  - Alum/polymer addition based on effluent turbidity and influent flow rate



Tik et al. (2013) ICA2013, Narbonne, France

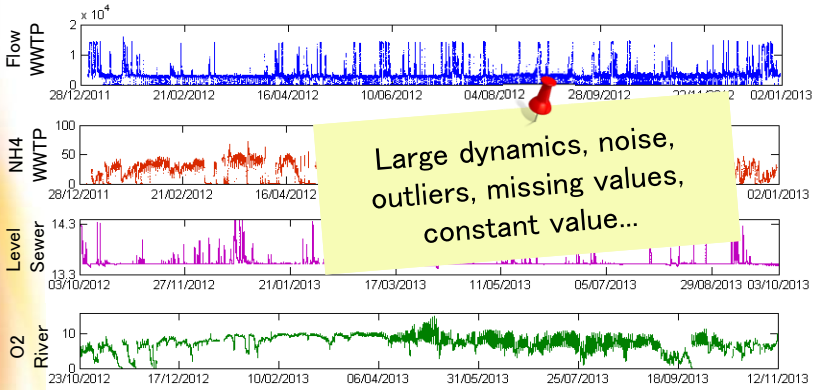
WQ data use – A motivating example



Scenario 3 uses 30% less alum than scenario 2

Data collection – status

... Time series



Large dynamics, noise, outliers, missing values, constant value...

Data collection – status

Shift from **not enough data,**

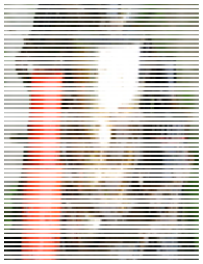
but with typically sufficient accuracy

to

Data graveyards
with unknown accuracy

Data quality – Problem definition

- Current challenges



- Development sensors
- Automated in-situ monitoring
- Information-rich data sets

Objective: practical tools for identification of unreliable data

UNIVERSITÉ LAVAL

7

modelEAU

Data quality – Problem definition

- Current methods?
  - Water quality time series are difficult to analyse
  - Manual procedures
    - Time consuming, inefficient
  - Automatic procedures
    - Effective data quality monitoring

Extract information from individual time series

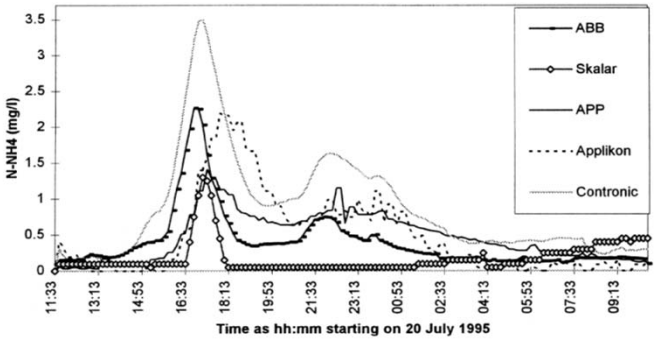
UNIVERSITÉ LAVAL

8

modelEAU

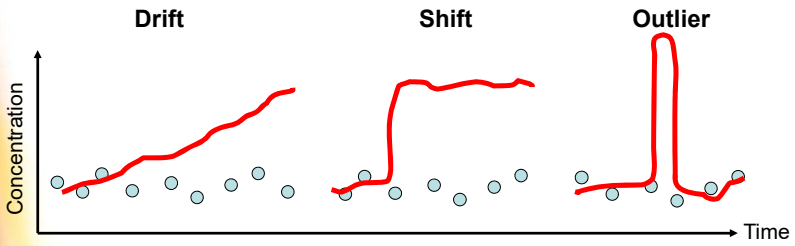
Data quality

- Wacheux et al. (1996) – Ammonia sensors



Data quality

- Systematic measurement errors

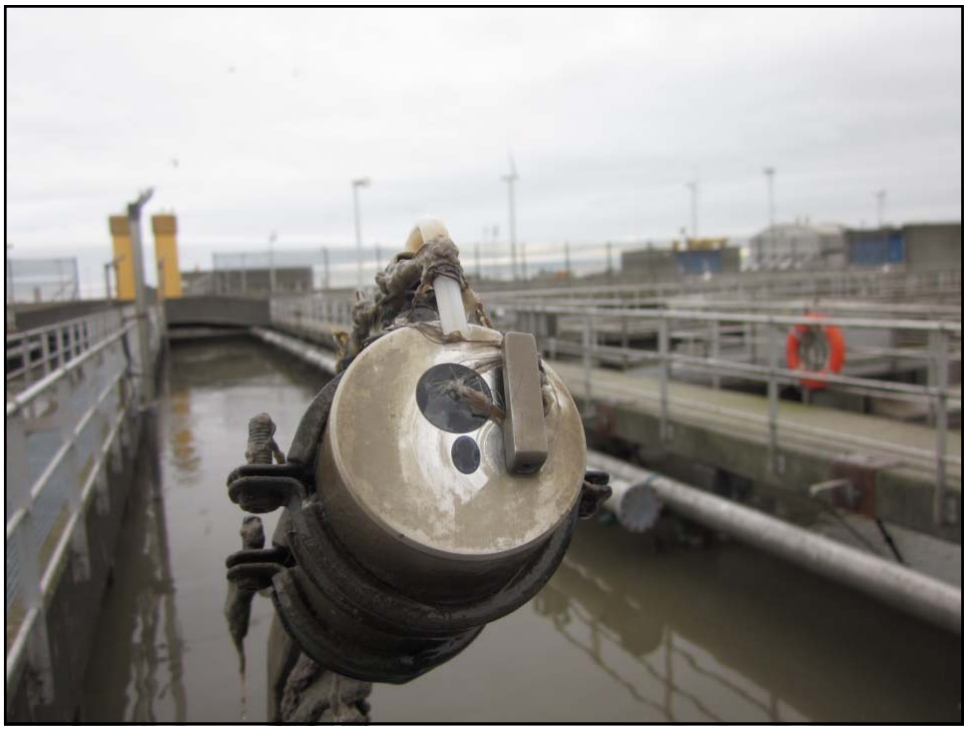


Leiv Rieger

Data collection :
Weekly maintenance + Air Cleaning

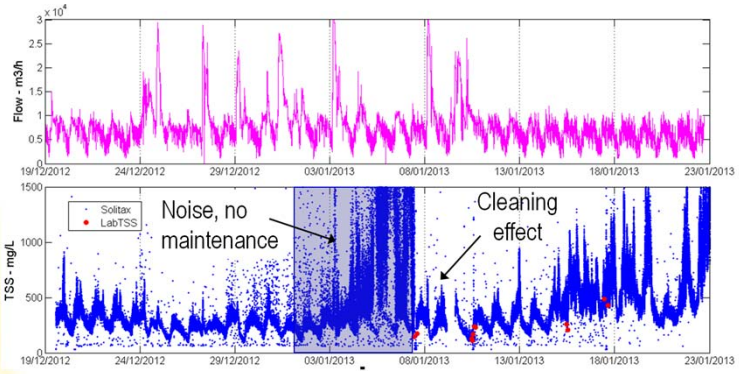
- Increase cleaning frequency
  until time has no effect on data quality


TSS



Data collection : Weekly maintenance + Wiper

- Effect of hair on wiper (raw data at PC inlet)



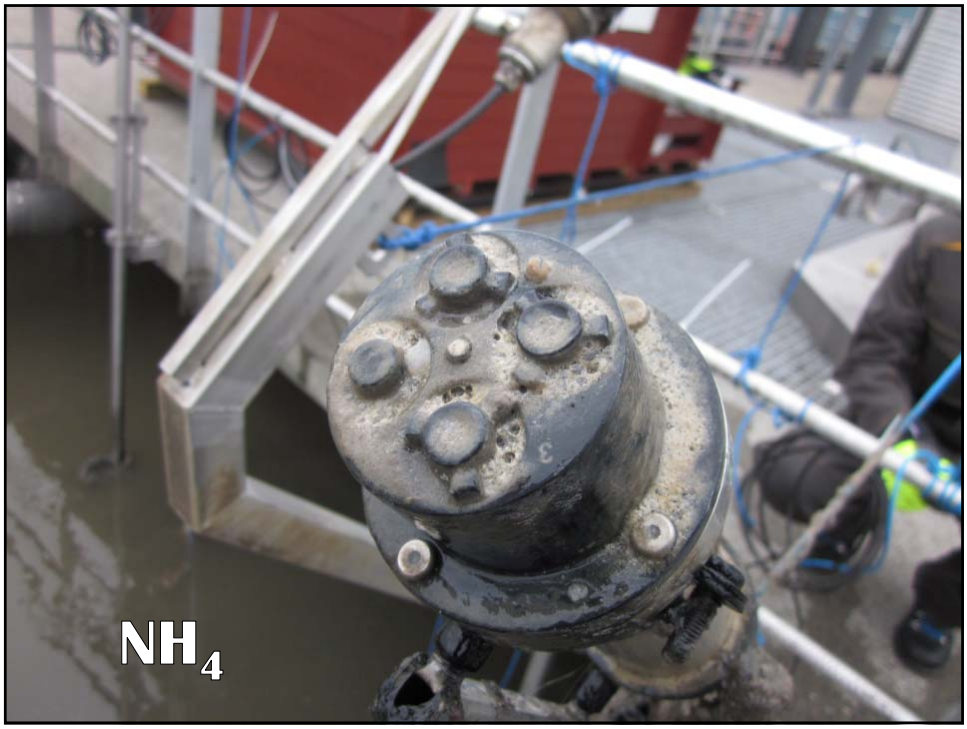
UNIVERSITÉ LAVAL

15

modelEAU

Data collection : Weekly maintenance + Air Cleaning

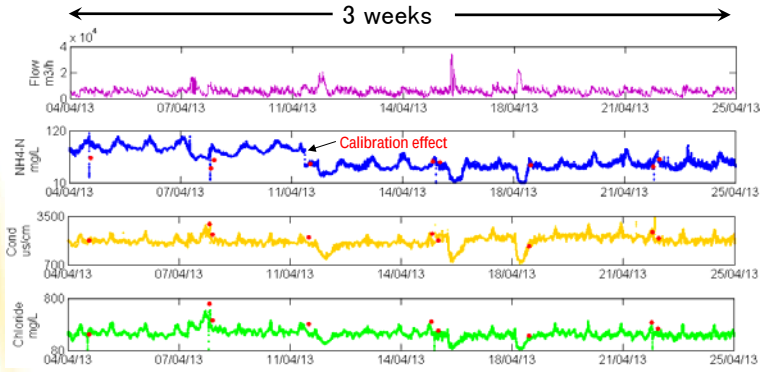
Data collection :
Weekly maintenance + Air Cleaning



Data collection :
Weekly maintenance + Air Cleaning

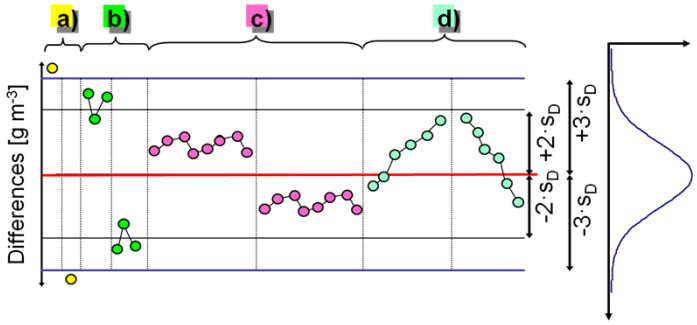
Data quality assessment – I

- Quality control measurements – recalibration



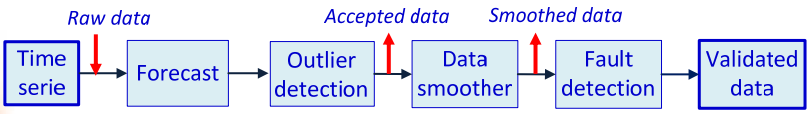
Data quality assessment – I

- Shewhart control charts
  (comparison of sensor and sample data)



Data quality assessment – II

- Method of Alferes *et al.* (2012)



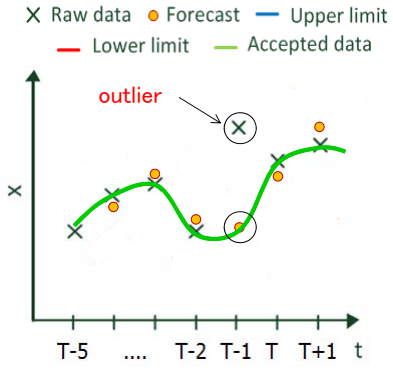
1. Outlier detection → Smoothed data
   Outlier elimination + smoothing
2. Fault detection → Validated data
   Evaluation of data features in outlier-cleaned and smoothed data → fault elimination

"An outlier is a raw value that deviates notably from other normal observations"

Data quality assessment – II

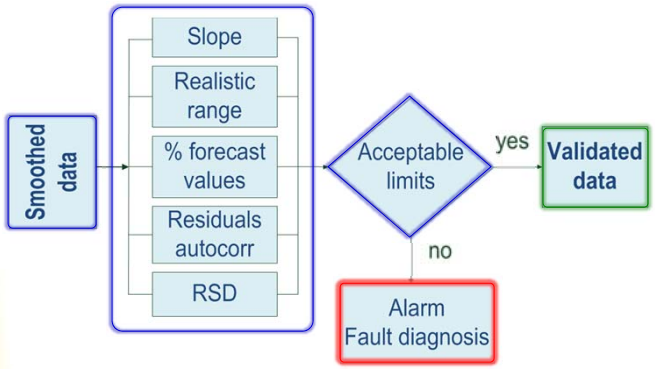
- Outlier detection
  - Autoregressive models
  - At T forecasting (T+1):
    - Variable $\hat{x}_T$
    - Std of error $\hat{\sigma}_{e,T}$
  - Prediction interval

$$xlim_T = \hat{x}_T \pm K\hat{\sigma}_{e,T}$$



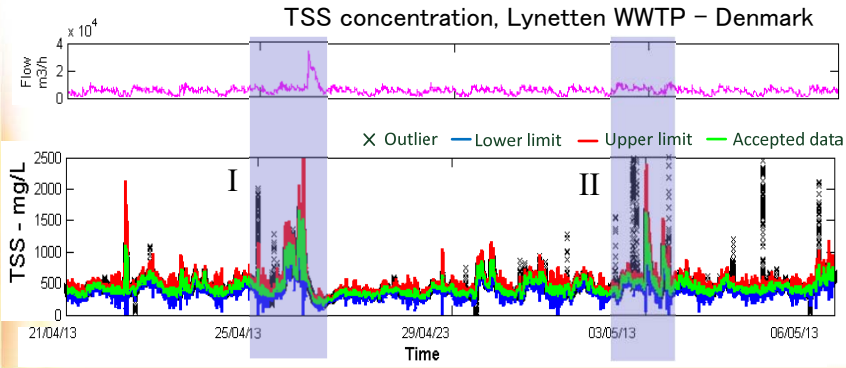
Data quality assessment – II

- Data features for fault detection



Data quality assessment – II

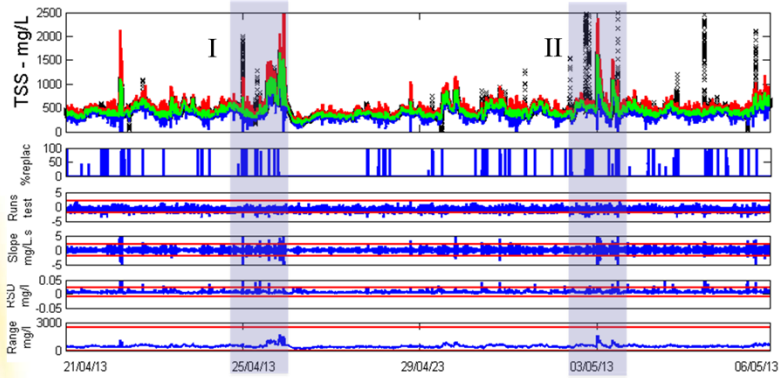
- An example



... some outliers and doubtful data identified

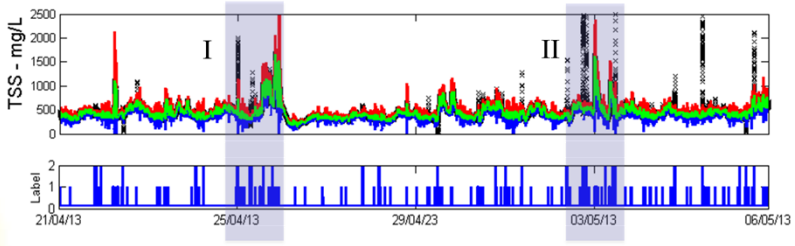
Data quality assessment – II

- An example



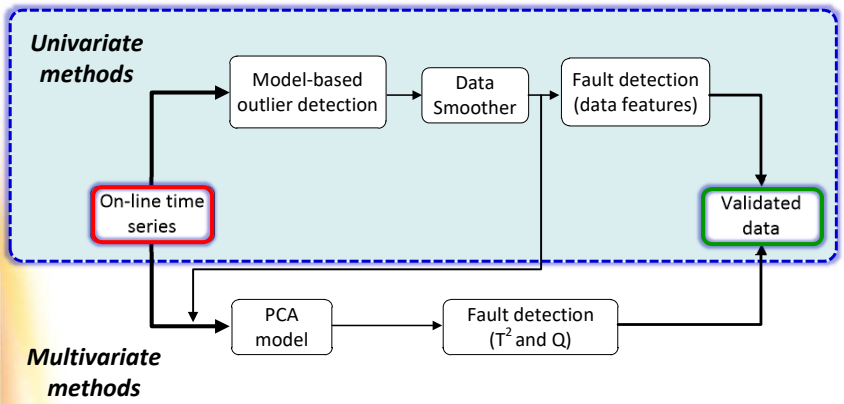
Data quality assessment – II

- An example



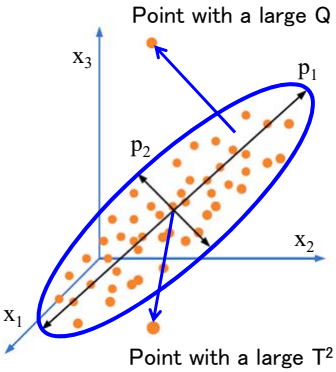
About 8% of data is considered as doubtful or not valid
(typically between 5 and 50% data loss)

Data quality assessment – II



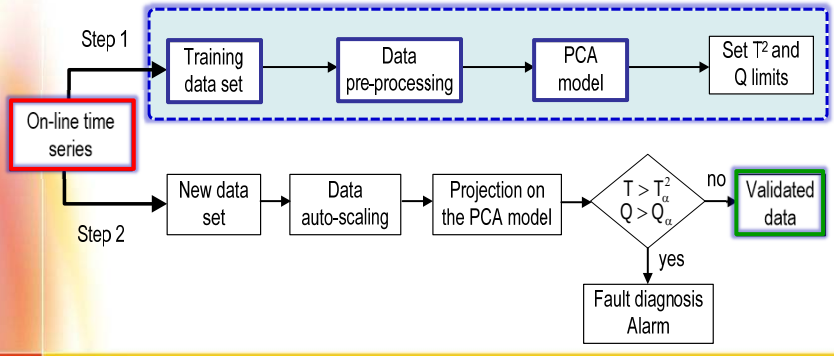
Data quality assessment – II

- Principle Component Analysis
  - Fault detection with statistics:
    - **T^2**: normalized sum of scores: variations within the model
    - **Q**: sum of squared residuals: goodness of fit of samples to the model
    - Fault detection limits are defined on the basis of "normal data"



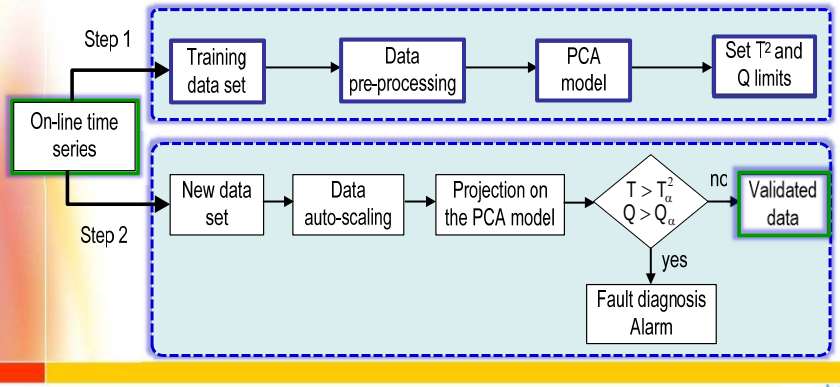
Data quality assessment – II

- Multivariate methods



Data quality assessment – II

- Multivariate methods



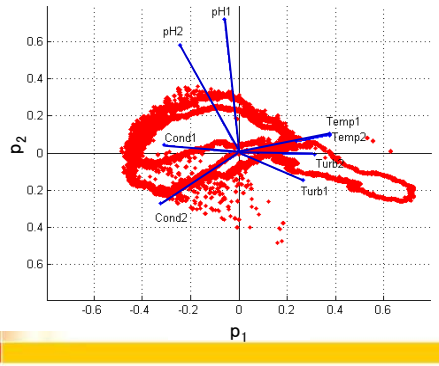
Data quality assessment – II

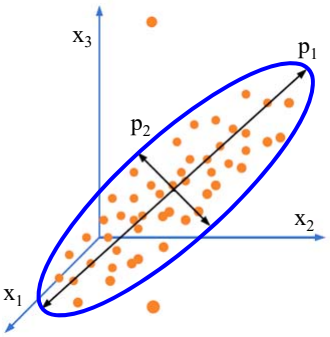
- Multivariate methods (WWTP, Quebec)
  - Dataset with 8 variables (redundant, 1 w/ air clean)
    - pH_1, pH_2, $Cond_1$, $Cond_2$, $Turb_1$, $Turb_2$, $Temp_1$, $Temp_2$
  - Training: 3-day of normal data to build the model

Data quality assessment – II

- Multivariate methods (WWTP, Quebec)

Data in the new PCA space – first 2 components

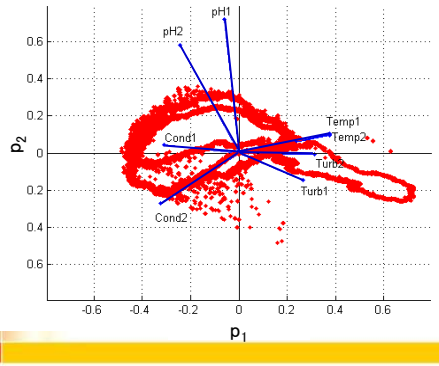




Data quality assessment – II

- Multivariate methods (WWTP, Quebec)

Data in the new PCA space – first 2 components

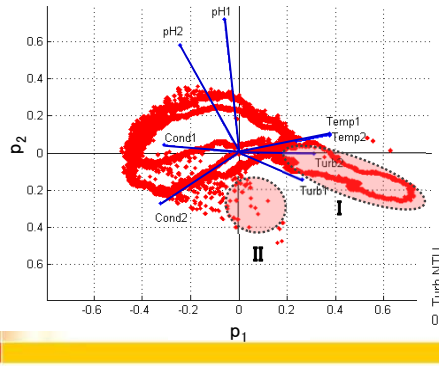


- Each point corresponds to a sample in the new space
- Vectors represent variables and contributions to p_1 and p_2
- Divergences between vectors represent bias between redundant sensors

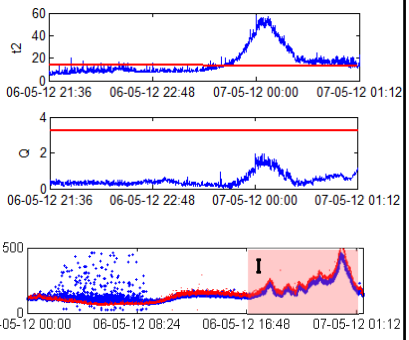
Data quality assessment – II

- Multivariate methods (WWTP, Quebec)

Data in the new space



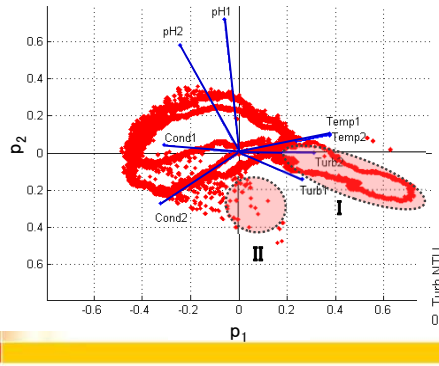
Statistics period I



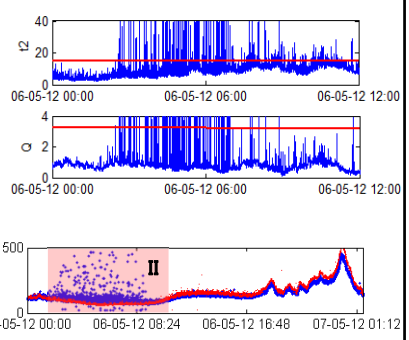
Data quality assessment – II

- Multivariate methods (WWTP, Quebec)

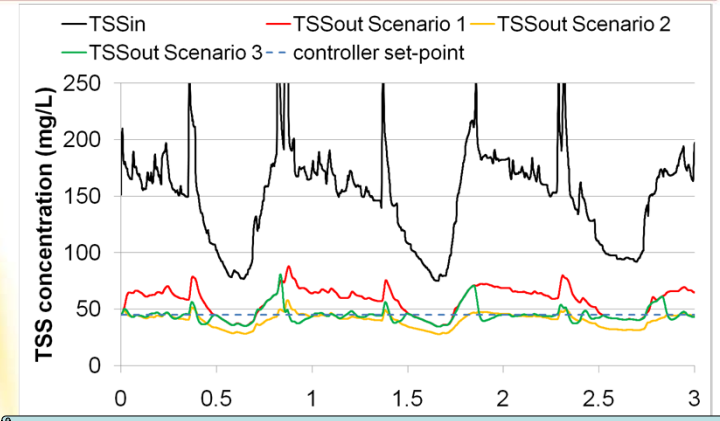
Data in the new space



Statistics period II



Take home messages – we want this!



Scenario 3 uses 30% less alum than scenario 2

Take home messages – not this!

Take home messages

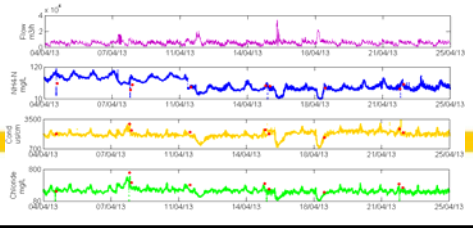
- Our set of water quality sensors is growing
- Our set of things we can do with them too

- We have more of them
- We can use them better

Take home messages

- We use them better by:
  - Better installation
  - Better sensor self-diagnosis
  - Better automatic cleaning systems

  - Automatic outlier removal and fault detection
  - More maintenance work, <u>when needed</u>



Acknowledgments

Canada Research Chair in Water Quality Modeling

Primodal

VILLE DE QUÉBEC

UMRSU

BIONEST Wastewater Treatment Solutions™

Développement durables des Infrastructures publiques

Fondation canadienne pour l'innovation
Canada Foundation for Innovation

Waterschap De Dommel

Acknowledgments (2006-2017)