Continuous monitoring of river quality using in situ monitoring stations: Efficient data quality assessment

M3 Workshop

Cologne, Germany

14-15 06 2012

Janelcy Alferes, Pascal Poirier and Peter Vanrolleghem

UNIVERSITÉ LAVAL

modelEAU

Problem definition

Water Framework Directive (WFD)

↓

Ecological status of water bodies

↓

River Basin Management Plans (RBMP)

↓

Integrated evaluation

↓

Need for monitoring and assessment

Problem definition

- In situ **monitoring stations**
  - WFD investigative monitoring.
  - High sampling frequency, huge data sets
- Reliability of sensors insufficient
- Real data are mostly noisy

➡ Data quality assessment crucial for practical use: ex.“modelling”

In situ monitoring stations

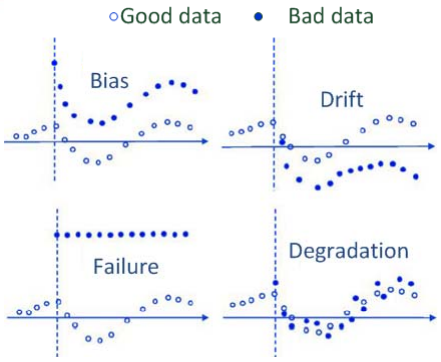
- Urban river (QC, Canada)
- Water quality variables:
  - pH, °C, µS/cm, O_2, NH_4, NO_3, K^+, TSS, COD…
- Sample time: 5-60 sec

  More than 1 million measurements per variable!

Representative data??

In situ monitoring stations

- Some observations:
  - Faulty probes
  - Practical issues
  - Maintenance
  - Low water level

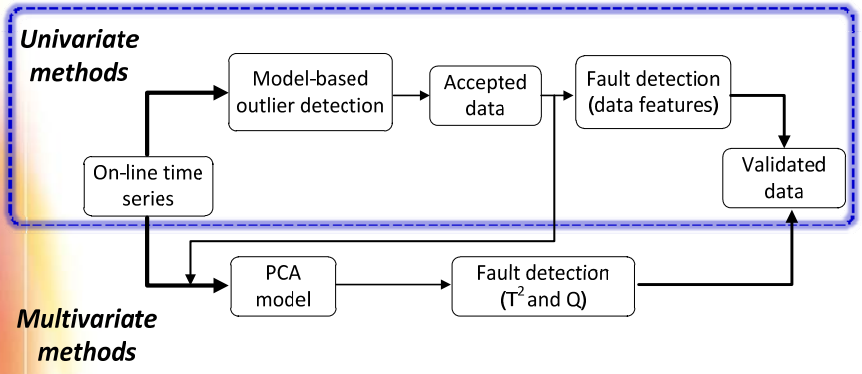


Data reliability can decrease

Data quality assessment tools

- Software tools
  for automatic data quality evaluation
- Using time series information

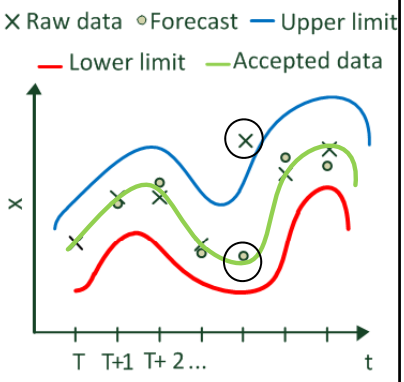
Data quality assessment tools



Data quality assessment tools

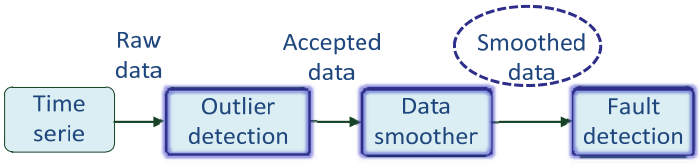
- Univariate methods
  - Outlier detection
  - Forecasting of variable $\hat{x}$ and std of error $\hat{\sigma}_e$
  - Prediction interval:

$$x_{\lim} = \hat{x} \pm K \cdot \hat{\sigma}_e$$



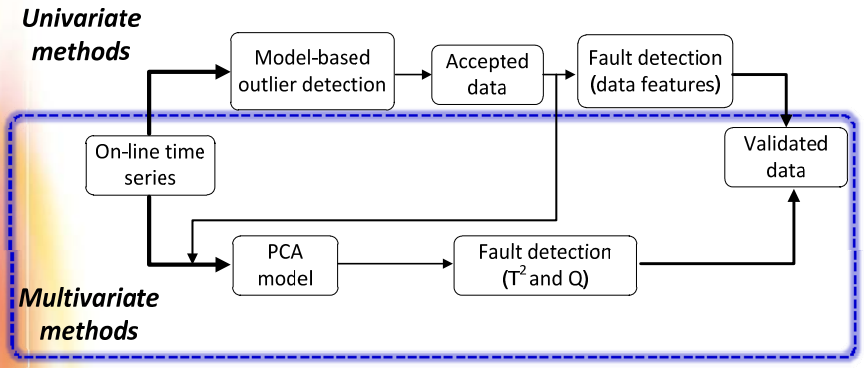
Data quality assessment tools

- Univariate methods



  - Fault detection --» data features and their limits:
    - % replaced data
    - Slope
    - Residual standard deviation (RSD)

Data quality assessment tools



Data quality assessment tools

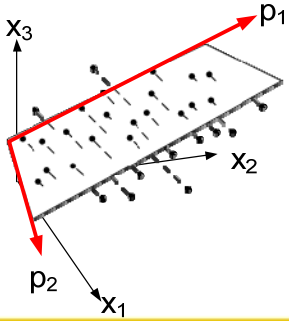
- Multivariate methods
  - Large multidimensional dataset X, redundant and correlated.
  - Allow to reduce the dimension of X by identifying “key” variables.

➡ New coordinated system

Data quality assessment tools

- Multivariate methods
  - Let be *n* variables ($x_1 \ldots x_n$) and *m* samples
  - New variables ($p_1..p_a$) as linear combinations:

$$p_1 = c_{11}x_1 + c_{12}x_2 + \ldots + c_{1n}x_n$$
$$p_2 = c_{21}x_1 + c_{22}x_2 + \ldots + c_{2n}x_n$$
$$p_a = c_{n1}x_1 + c_{n2}x_2 + \ldots + c_{nn}x_n$$



  - Axes of a new coordinate system
  - Directions of max. variability

Data quality assessment tools

- Multivariate methods
  - Final model:

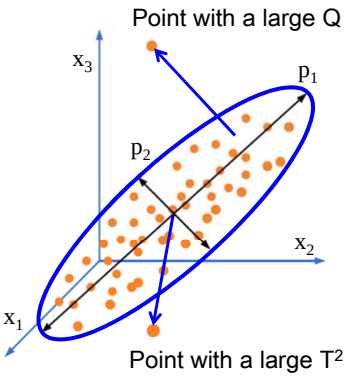
$$T = XP$$

transformed data "**scores**" | original data | transf. matrix

  - Fault detection using the model by monitoring two statistics (T^2, **Q**)

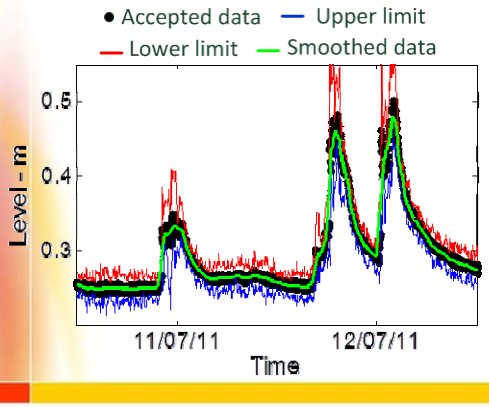
Data quality assessment tools

- Multivariate methods
  - T^2: normalized sum of scores: variations within the model
  - **Q**: sum of squared residuals: goodness of fit of samples to the model
  - Detection limits are defined.



Some preliminary results

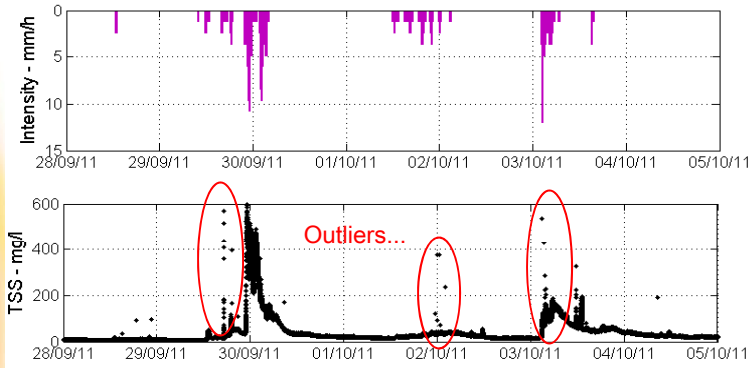
- Univariate methods (Level)



- Variable prediction interval
- Adaptation to time-varying hydraulics

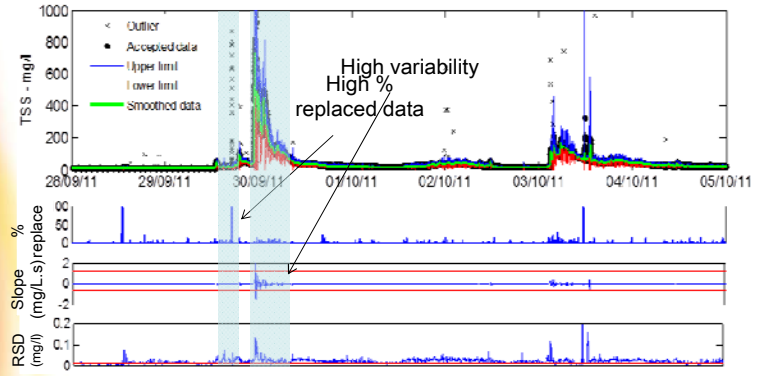
Some preliminary results

- Univariate methods (TSS)



Some preliminary results

- Univariate methods (TSS)



Some preliminary results

- Multivariate methods
  - Dataset with 8 variables
    - NTU, NO_3, TOC, DOC, pH, K^+, NH_4 and °C
  - Obtaining the model
    - Two new variables (p_1,p_2) capture the max variability.
    - New space and transformed data ("scores"):

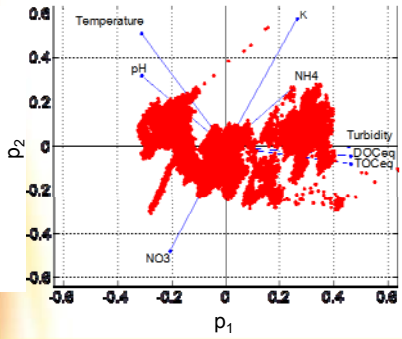
$$T = XP$$

  - Monitoring of T^2 and Q
    - Define acceptability limits

Some preliminary results

- Multivariate methods

Representation of data in the new space

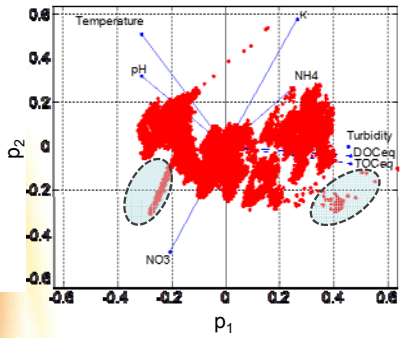


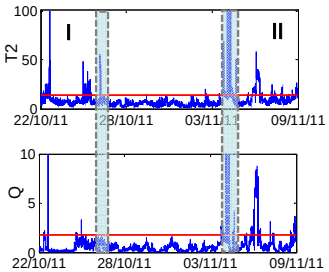
- Vectors represent the variables and their contribution to p_1 and p_2
- Each point corresponds to a sample in the new space (scores)

Some preliminary results

- Multivariate methods

Representation of data in the new space

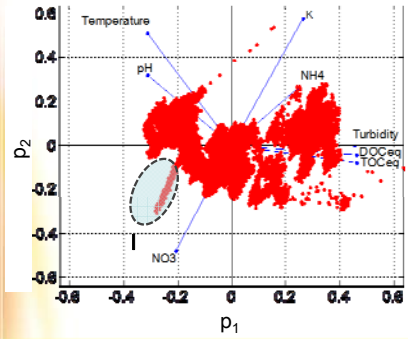


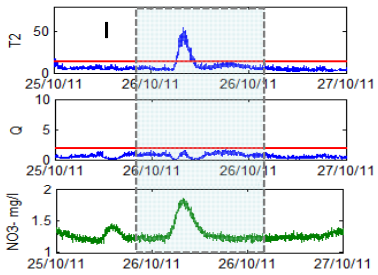


Some preliminary results

- Multivariate methods

Representation of data in the new space

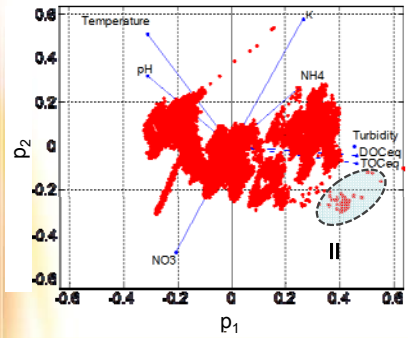


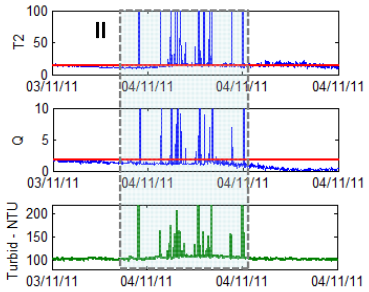


Some preliminary results

- Multivariate methods

Representation of data in the new space





Conclusions

- Data quality assessment tools satisfactory validated
- Methods allowed the detection of individual and multiple faults
- Validated data can be used within the WFD (river models, basin management...)
- Next steps: data coming from different stations

Acknowledgement

Canada Research Chair in Water Quality Modeling

Primodal
Earth Essential Technologies


Thank you for your attention!
Questions are welcome...
For more information:
Janelcy.Alferes@gci.ulaval.ca
UNIVERSITÉ LAVAL
25
modelEAU