Water quality measurements and data quality evaluation

Novatech

Lyon, France

June 23, 2013

Sovanna Tik, Janelcy Alferes, Peter Vanrolleghem

Presentation overview

- Context
- Need for advanced data quality evaluation
- Data quality assessment tools
- Results
- Conclusions and perspectives

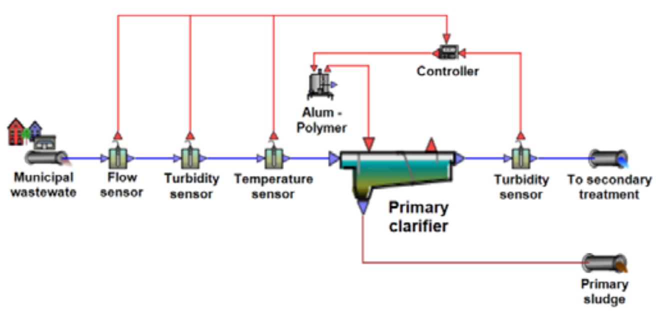
Context

- Real-Time Control systems development
  - Design
  - Operation
- Control strategies evolve:
  - quantity → quality → impact

➡ **Need for reliable online water quality data**

Context

- Process control system at inlet WWTP



Model configuration in WEST® (mikebydhi.com)

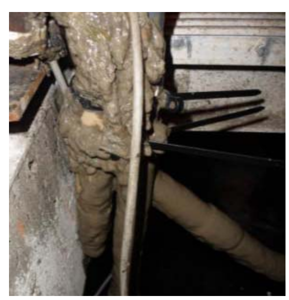
Context

- Online monitoring stations

Online sensors

Context

- Online sensors: challenging conditions



➡ **Representativeness??**

Need for advanced data quality evaluation

- In situ **monitoring stations**
  - Information-rich data sets √
  - Pollution dynamics √
  - Reduce costs √
  - Huge/complex data sets X
  - Errors and uncertainties X
  - Reliability of sensors insufficient X

➡ **Data evaluation/validation is crucial**

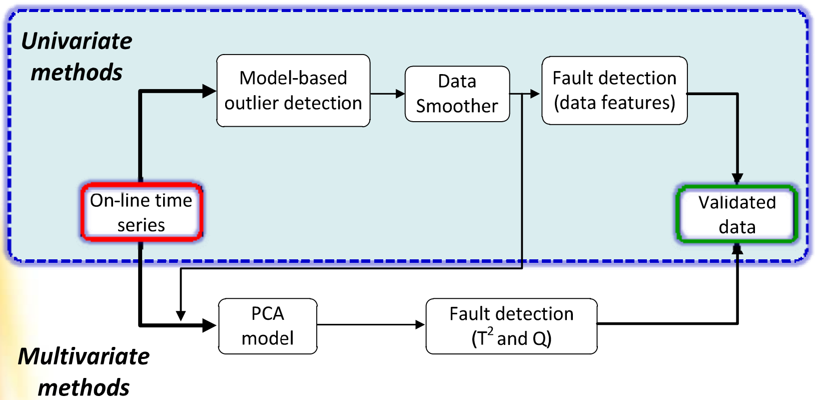
Need for advanced data quality evaluation

- Current practice : manual procedures
- Challenge in water systems monitoring ??

Automatic data quality evaluation

- Corrupted, doubtful, unreliable data
- Sensor faults

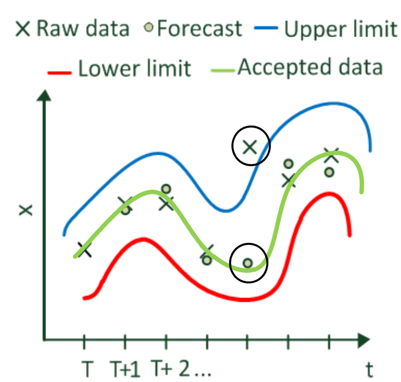
Data quality assessment tools



Data quality assessment tools

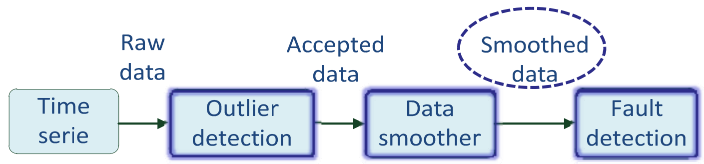
- Univariate methods
  - Outlier detection
  - Autoregressive models
  - At T forecasting (T+1):
    - variable $\hat{x}$
    - std of error $\hat{\sigma}_e$
  - Prediction interval:

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



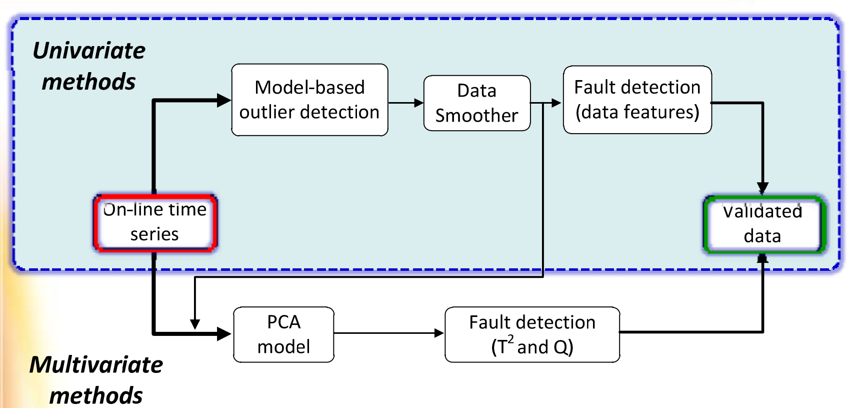
Data quality assessment tools

- Univariate methods



- Fault detection
  - % replaced data --» data goodness
  - Slope --» rate of change
  - Residuals correlation --» good fit of model
  - Residual std (RSD) --» variance

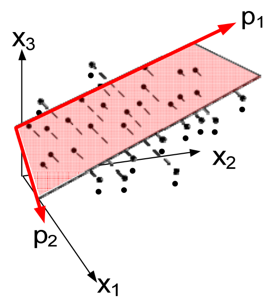
Data quality assessment tools



Data quality assessment tools

- Multivariate methods
  - Reduce dimension of data X (x_1, x_2..x_n)
  - New variables (p_1,p_2 … p_n) as linear combinations

$$p_1 = c_{11}x_1 + c_{12}x_2 + c_{13}x_3$$
$$p_2 = c_{21}x_1 + c_{22}x_2 + c_{23}x_3$$



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

Data quality assessment tools

- Multivariate methods
  - PCA model: $X = TP^T$ ⟶ **Model matrix** from correlation matrix C

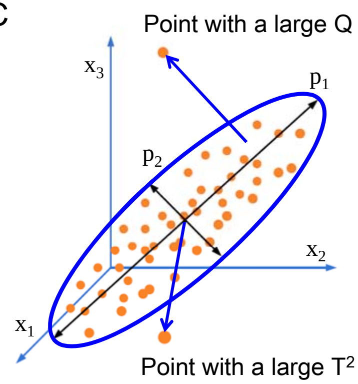
“components” “variances”

$$C = \begin{bmatrix} p_{11} & p_{12} & \cdots p_{1n} \\ p_{21} & p_{22} & \cdots \\ \vdots & & \\ p_{n1} & p_{n2} & \cdots p_{nn} \end{bmatrix} \bullet \begin{bmatrix} \lambda_1 & 0 & 0 \ldots 0 \\ 0 & \lambda_2 & 0 \quad 0 \\ \vdots & & \\ 0 & \cdots & \lambda_n \end{bmatrix} \bullet \begin{bmatrix} p_{11} & p_{12} & \cdots p_{1n} \\ p_{21} & p_{22} & \cdots \\ \vdots & & \\ p_{n1} & p_{n2} & \cdots p_{nn} \end{bmatrix}^T$$

... choosing the # of components --» largest variances

Data quality assessment tools

- Multivariate methods
  - Fault detection within the PC space
    - 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.



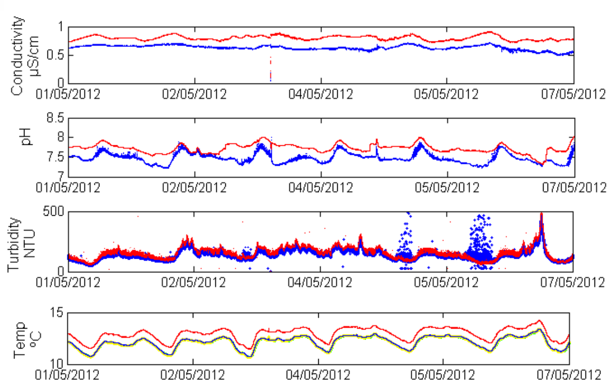
UNIVERSITÉ LAVAL

15

modelEAU

Results

- Case Study: WWTP, primary clarifier (Québec)



- Redundant turb, pH, temp, cond. sensors
- Sample time: 5 sec

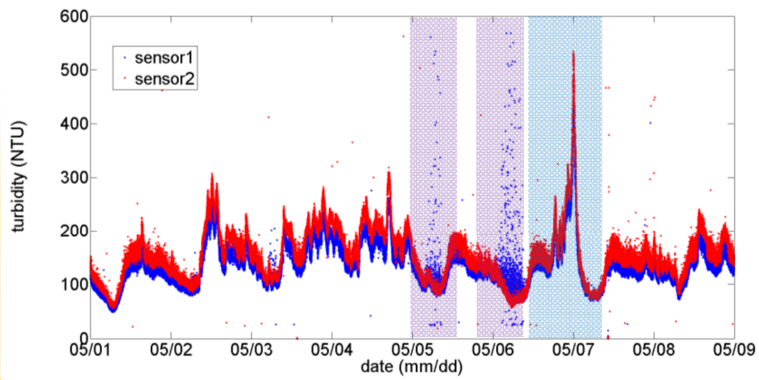
UNIVERSITÉ LAVAL

16

modelEAU

Results

- Raw turbidity data



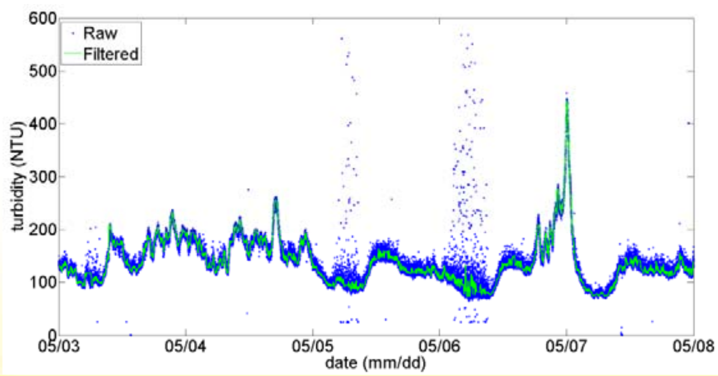
UNIVERSITÉ LAVAL

17

modelEAU

Data treatment results

- Univariate method – Filtered turbidity data



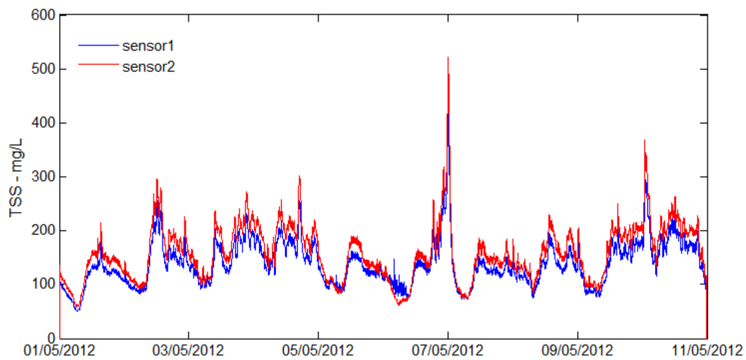
UNIVERSITÉ LAVAL

18

modelEAU

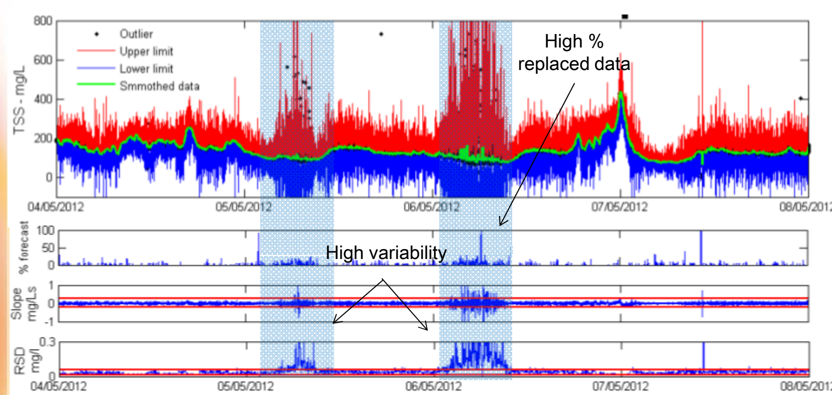
Results

- Univariate methods: Filtered data



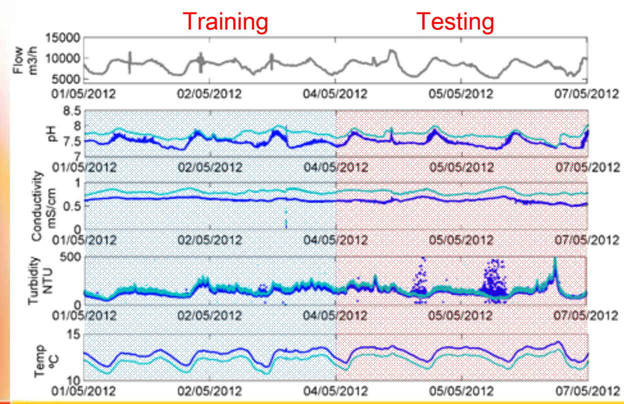
Results

- Univariate methods – Analysis sensor turb 1



Results

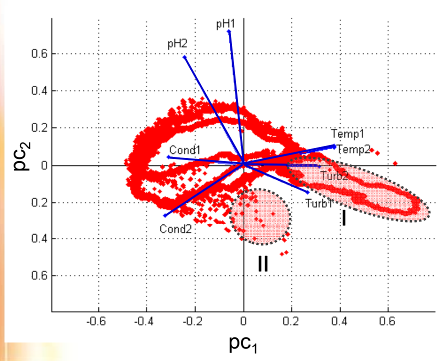
- Multivariate methods



- 8 variables
- First 3 PC > 90% variability
- Training: 3-day

Results

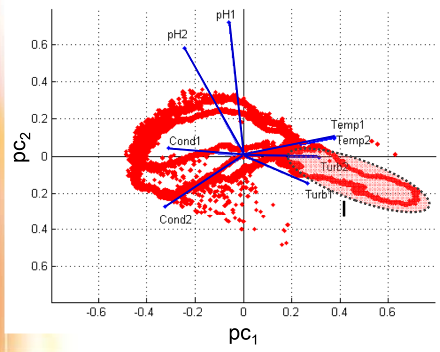
- Multivariate methods – PCA results

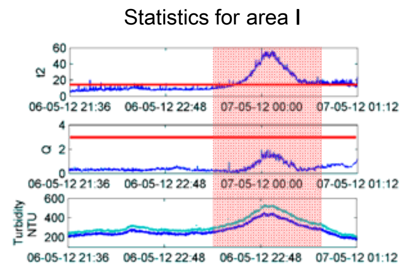


- Contribution of each parameter to pc_1 and pc_2
- Each point corresponds to a sample in the new space (scores)
- Visual detection of some abnormal periods

Results

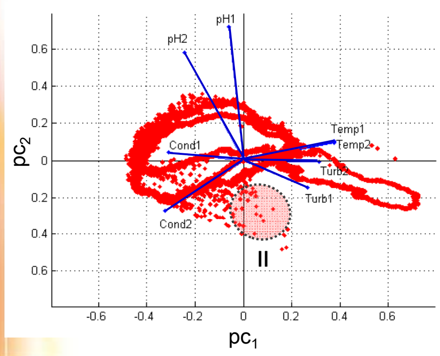
- Multivariate methods – PCA results

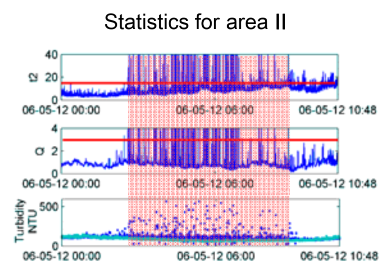




Results

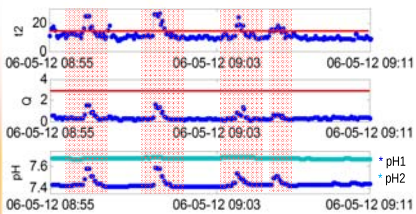
- Multivariate methods – PCA results





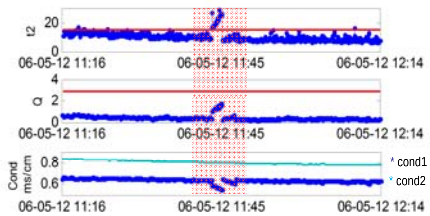
Results

- Multivariate methods – PCA results



Drift situations detected for pH sensor 1

Drift situations detected for conductivity sensor 1



Conclusions and perspectives

- Univariate methods
  - outliers removal, smoother time series
  - detection of individual faults
- Multivariate methods
  - dimension reduction
  - detection of multiple faults
- Next step: online implementation