Data quality improvement in automated water quality measurement stations

Presented by:
Atefeh Saberi (M.Sc. Student /Electrical engineering)

Co-Author:
Janelcy Alferes Castano

Supervisors:
André Desbiens, Peter Vanrolleghem

Stats & Control Meeting 2014

École Polytechnique Montréal

May 22-23 2014

Problem definition

Hydrological systems

Effective management of water bodies

Reliable water quality information

Trustable further application

Modeling, Decision making, Control

Problem definition

- Better description of water systems with fast dynamics

Application of In-situ Automated Continuous Monitoring Stations!

Problem definition

- In situ automated continuous monitoring stations:
  - Information-rich data sets √
  - Capturing pollutions dynamics √
  - Reduce costs √
  - Huge/complex data sets X
  - Errors and uncertainties X
  - Insufficient sensor reliability X

➡ **Data evaluation/validation is crucial**

In situ monitoring stations

- Municipal treatment plant
  (Copenhagen, Denmark)
- Water quality variables:
  - pH, TSS, NH_4, Turbidity,...
- Sample time: 5-60 sec
- Practical issues:
  - Maintenance, fouling, clogging...

Representative data?!

Data quality assessment methods

- Manual procedures
  - Tedious and time consuming
- Automatic data quality evaluation procedures

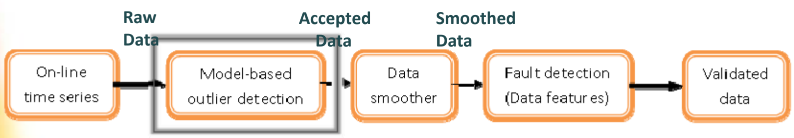
Using time series information!

  - Corrupted, doubtful, unreliable data
  - Noise
  - Sensor faults
  - Outliers

Automated data quality assessment tools

- Univariate method developed at model*EAU,*

Département de génie civil et de génie des eaux, Université Laval



Current Model-based outlier detection method

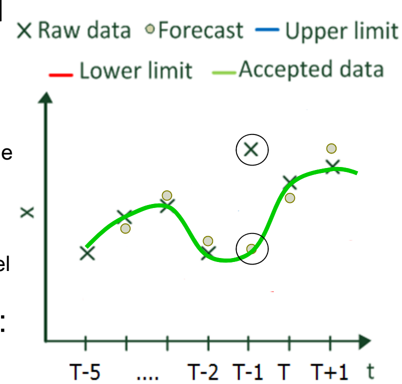
Exponential smoothing model

- At T forecasting T+1:
  - Variable $\hat{x}$
    - 3[rd] order exponential smoothing mode

$$x_t = a + bt + \frac{1}{2}ct^2$$

  - STD of forecast error $\hat{\sigma}_e$
    - 1[st] order exponential smoothing model
- Dynamic Prediction interval:

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



Model-based outlier detection

- Alternative method: Desired properties
  - Real-time, on-line applicable
  - Systematic
  - Appropriate for the system under study
  - Automatically applicable
  - Not complicated

Autoregressive moving average with integrator (ARIMA) applied to a moving window data

UNIVERSITÉ LAVAL

9

modelEAU

ARIMA model

- Proposed ARIMA model : $$y(k) = \frac{C(z^{-1})}{A(z^{-1})(1-z^{-1})} e(k)$$

  with: $$A(z^{-1}) = 1 + a_1 z^{-1} + \ldots + a_{na} z^{-na}$$ $$C(z^{-1}) = 1 + c_1 z^{-1} + \ldots + c_{nc} z^{-nc}$$

- Calculation of j-step ahead forecast value according to data available at time k:

$$\hat{y}(k+j\,|\,k) = \frac{F_j}{C(z^{-1})} y(k)$$

With F_j calculated according to Diophantine equation:

$$\frac{C(z^{-1})}{A(z^{-1})} = E_j + z^{-j} \frac{F_j}{A(z^{-1})}$$

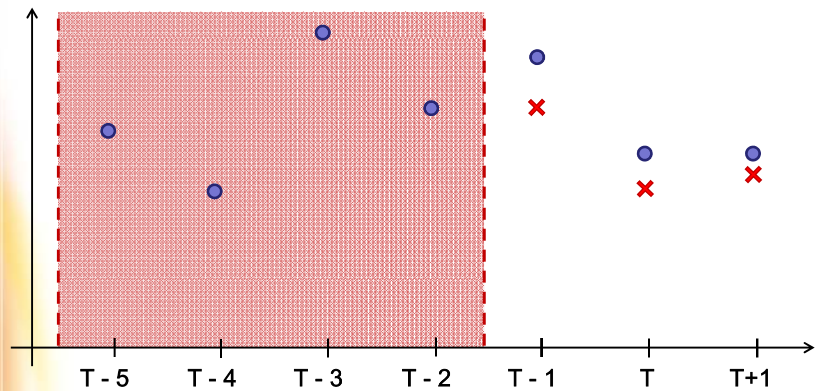
From degree j-1

UNIVERSITÉ LAVAL

10

modelEAU

Moving Window approach



UNIVERSITÉ LAVAL

11

modelEAU

ARIMA model with Moving Window

- Algorithm for window size N:
  1. Consider N data values in window
  2. Identify ARIMA model parameters
  3. Forecast of j-step ahead values
  4. Move data window one step further
  5. Repeat steps 2 to 4

UNIVERSITÉ LAVAL

12

modelEAU

Challenge!

- Specification of Window Size (WS) ?
  - If WS too small ⇨ More sensitive to noise
  - If WS too large ⇨ Averaging the dynamic variations
- Solution ?

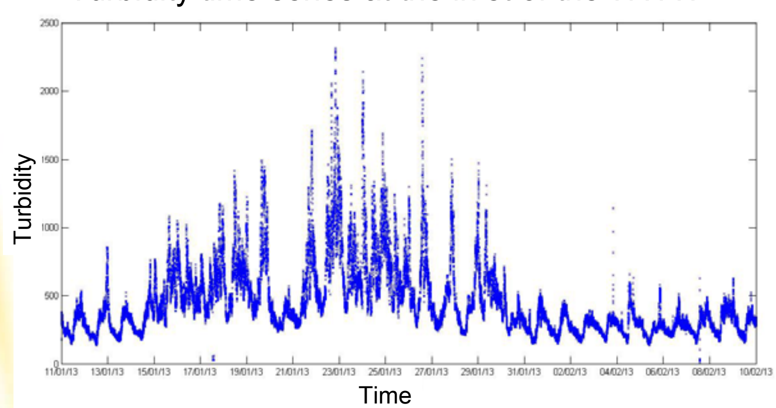
Selection of WS that considers the dynamics of the variable

UNIVERSITÉ LAVAL

modelEAU

Raw data

Turbidity time series at the inlet of the WWTP

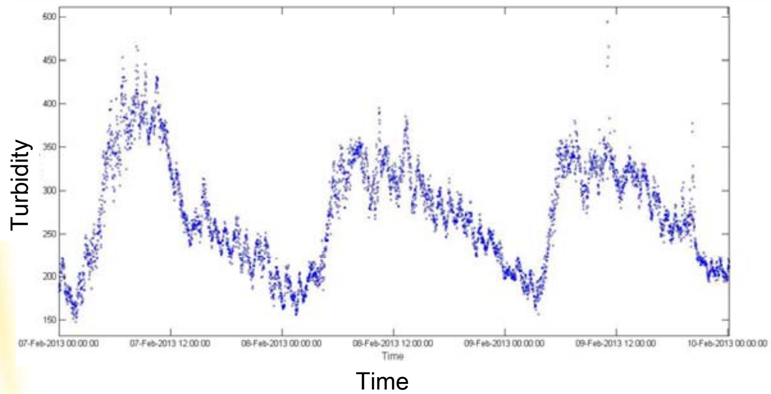


UNIVERSITÉ LAVAL

modelEAU

Raw data

Turbidity time series at the inlet of the WWTP



UNIVERSITÉ LAVAL

Application to the system

- Window size of 30 data points for 1 min sampling time
- Estimation of ARIMA model parameters with one pole and one zero

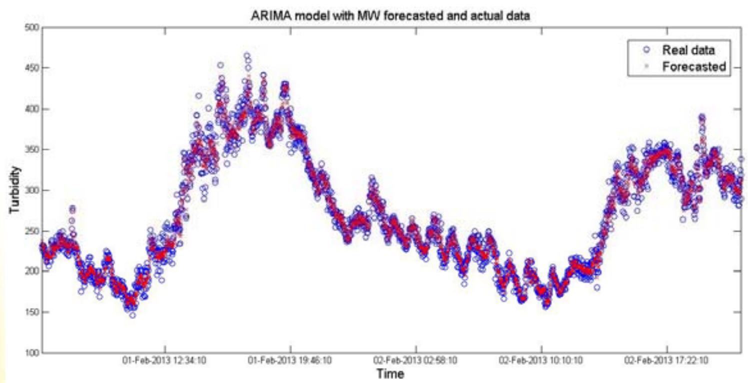
$$y(k) = \frac{(1 + c_1 z^{-1})}{(1 + a_1 z^{-1})(1 - z^{-1})} e(k)$$

- Calculation of one step ahead forecast value (j=1)
- Calculation of dynamic forecast error and prediction interval $x_{\lim} = \hat{x} \pm K \cdot \hat{\sigma}_e$
- Detection and replacement of outlier

UNIVERSITÉ LAVAL

Results

1 step-ahead forecast for a dry period

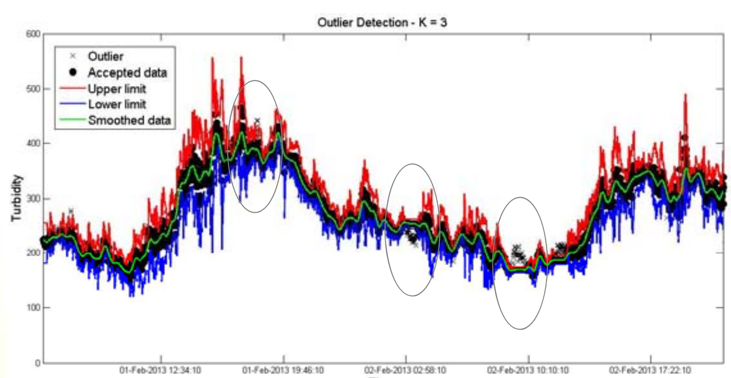


UNIVERSITÉ LAVAL

17

Results

Outlier detection

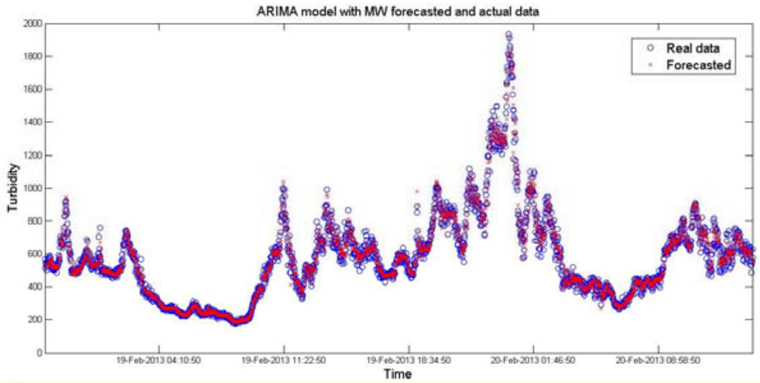


UNIVERSITÉ LAVAL

18

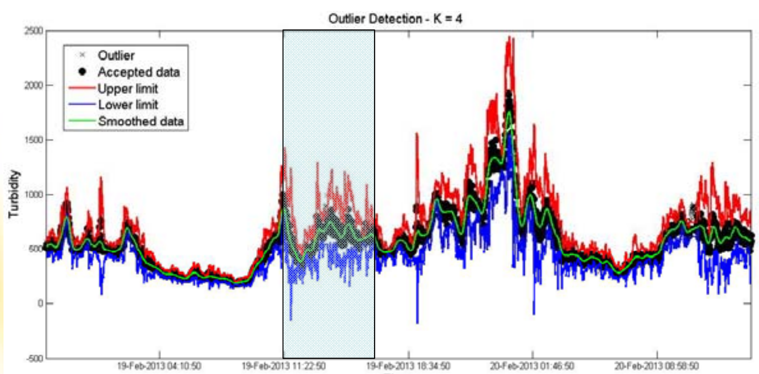
Results

1 step-ahead forecast for a wet period



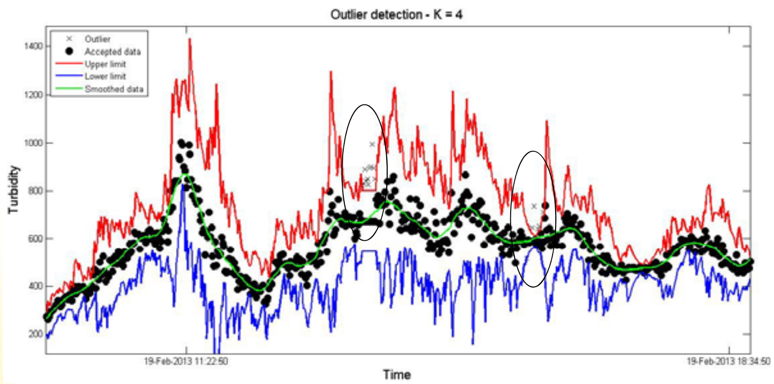
Results

Outlier detection



Results

Outlier detection

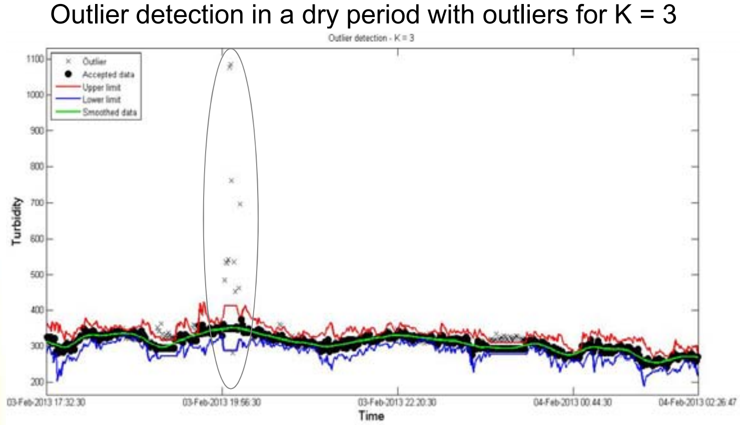


UNIVERSITÉ LAVAL

21

Results

Outlier detection in a dry period with outliers for K = 3



UNIVERSITÉ LAVAL

22

Conclusion

- ARIMA model with moving window is fitted to water quality time series to produce one step-ahead forecast
- The outliers are detected by considering prediction intervals calculated according 1^{st} order exponential smoothing model

Future works

- Application of the Multi Model Filtering Algorithm (MMFA) to the system under study
  - Identification of set of models according to different modes of system's behavior
  - Designing Kalman filters for each of the different modes of behavior
  - Calculating conditional probability of each of the models to represent the observed system behavior