Accuracy of different on-line sensors for the estimation of pollutant concentrations (TSS, COD total and dissolved) in wastewater, stormwater and all water

Mathieu Lepot ¹,²
Jean-Luc Bertrand-Krajewski ¹
Jean-Baptiste Aubin ¹

¹ University of Lyon, INSA Lyon, LGCIE, 34 avenue des arts, F-69621 Villeurbanne cedex, France, mathieu.lepot@insa-lyon.fr, jean-luc.bertrand-krajewski@insa-lyon.fr, jean-baptiste.aubin@insa-lyon.fr
² Sepia Conseils, 53 rue de Turbigo, 75003 Paris, France.
Summary

- Context
- Objectives
- Materials & Methods
- Results
- Conclusions
Context

- *On-line* monitoring for pollutants concentrations

- No sensors able to measure directly TSS or COD / Need of sets of samples to calibrate correlation functions

- Many surrogate sensors

- No study to compare them
Objectives

- Compare different surrogate sensors to estimate pollutants concentrations
- Find most accurate sensors for on-line monitoring
- Compare different methods to remove outliers
Materials & Methods

- Materials
  - Grab sampling (inlet of a WWTP – combined sewer)
    - 94 samples collected during dry weather
    - 44 samples collected during wet weather
    - 165 samples for all weather (dry and wet)
Materials & Methods

- **Materials**
  - 7 sensors
    - Single wavelength (IR: 880 nm) turbidity meters: *Krohne, Endress-Hauser*,
    - Double wavelength (UV: 254 and Vis: 560 nm) turbidity meter: *EFS*,
    - UV/vis spectrometer: *s::can*
    - Microwave sensor: *Metso Automation*
    - pH and conductivity meters: *Yokogawa*
Materials & Methods

- **Materials**
  - Connexion of sensors

![Diagram of connexion of sensors](image)
Materials & Methods

Methods

1. Sensors calibration
Materials & Methods

Methods

2. Lab analyses and recorded data:
   • Triplicates
   • TSS: NF T90-105-2 (1997)
   • Total and dissolved COD: NF T90-101 (2001)
   • $p$ and $u(p)$ for each pollutant and each sample
   • 60 values given by sensors ($x$ and $u(x)$) or 10-15 spectra for UV-vis
Materials & Methods

Methods

3. Outliers detection

3.a. Pretreatment of recorded spectra (for each sample)

- removal of « most extreme » spectra
- search of « the most in the middle » one, based of data depth theory (Lopez-Pintado et Romo, 2006)
- evaluation of standard uncertainty for each wavelength
Materials & Methods
Materials & Methods

Methods

3.b. Three steps method (for each dataset, Bertrand Krajewski et al, 2007):

- outliers in TSS, total and dissolved COD in triplicates
- too uncertain values (10 % for TSS, 20 % for COD, 5 % for sensors values)
- abnormal samples:
  - Mahalanobis distance (DM)
  - Euclidean distance (DE), to evaluate if one sample could strongly affect the correlation function
Materials & Methods

Methods

4. Calculation of correlation functions:
   - For univariate models: Williamson method
     \[ S = \sum_{i=1}^{M} \left[ \frac{1}{u^2(x_i)} \cdot (X_i - x_i)^2 + \frac{1}{u^2(y_i)} \cdot (P_i - p_i)^2 \right] \]
   - For spectral regression: PLS
     (Torres & Bertrand-Krajewski, 2008)
   - For multivariate models:
     \[ S_{EXTENDED} = \sum_{i=1}^{M} \left\{ \sum_{j=1}^{N_{y}} \left[ \frac{1}{u^2(x_{j,i})} \cdot (X_{j,i} - x_{j,i})^2 \right] + \frac{1}{u^2(p_i)} \cdot (P_i - p_i)^2 \right\} \]
Materials & Methods

Methods

- Evaluation of accuracy
  - RMSEP criterion (Dantas Filho et al, 2005)

\[
RMSEP = \sqrt{\frac{\sum_{i=1}^{M} (P_i - p_i)^2}{M}}
\]
## Results

<table>
<thead>
<tr>
<th>Set</th>
<th>Data</th>
<th>Methods</th>
<th>n-tuple</th>
<th>Initial number</th>
<th>Validated number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>4-20 mA output</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>94</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Physical units</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>94</td>
<td>68</td>
</tr>
<tr>
<td>Dry</td>
<td>4-20 mA output</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>47</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Spectra</td>
<td>DM</td>
<td>spectra</td>
<td>94</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Spectra</td>
<td>DE</td>
<td>spectra</td>
<td>94</td>
<td>78 - 77 - 79</td>
</tr>
<tr>
<td>Wet</td>
<td>4-20 mA output</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>44</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Physical units</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Spectra</td>
<td>DM</td>
<td>spectra</td>
<td>44</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Spectra</td>
<td>DE</td>
<td>spectra</td>
<td>44</td>
<td>34 - 36 - 35</td>
</tr>
<tr>
<td>All</td>
<td>4-20 mA output</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>165</td>
<td>144</td>
</tr>
<tr>
<td>All</td>
<td>Physical units</td>
<td>DM</td>
<td>TSS, total COD, dissolved COD</td>
<td>165</td>
<td>100</td>
</tr>
<tr>
<td>All</td>
<td>Spectra</td>
<td>DM</td>
<td>spectra</td>
<td>165</td>
<td>130</td>
</tr>
<tr>
<td>All</td>
<td>Spectra</td>
<td>DE</td>
<td>spectra</td>
<td>165</td>
<td>136 - 136 - 138</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Set</th>
<th>Pollutant</th>
<th>$RMSEP$ (mg/L) with DE</th>
<th>$RMSEP$ (mg/L) with DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>TSS, one wavelength</td>
<td>40.0</td>
<td>91.3</td>
</tr>
<tr>
<td>Dry</td>
<td>Total COD, one wavelength</td>
<td>61.7</td>
<td>129.1</td>
</tr>
<tr>
<td>Dry</td>
<td>Dissolved COD, two wavelengths</td>
<td>28.3</td>
<td>35.3</td>
</tr>
<tr>
<td>Dry</td>
<td>Dissolved COD, one wavelength</td>
<td>28.0</td>
<td>30.6</td>
</tr>
<tr>
<td>Wet</td>
<td>TSS, one wavelength</td>
<td>16.5</td>
<td>31.6</td>
</tr>
<tr>
<td>Wet</td>
<td>Total COD, one wavelength</td>
<td>41.9</td>
<td>41.9</td>
</tr>
<tr>
<td>Wet</td>
<td>Dissolved COD, two wavelengths</td>
<td>6.9</td>
<td>8.2</td>
</tr>
<tr>
<td>Wet</td>
<td>Dissolved COD, one wavelength</td>
<td>12.1</td>
<td>12.2</td>
</tr>
</tbody>
</table>
Results (for TSS)

Kajaani TS 4-20 mA output
DW: 10th RMSE = 189.2 mg/L (P3)
WW: Not tested
AW: Not tested

EFS 4-20 mA output(s)
DW: 6th RMSE = 34.6 mg/L (P1 - 550 nm)
WW: Not tested
AW: Not tested

TK - TEH TURBIDITIES
DW: 1st RMSE = 18.4 mg/L (T1)
WW: 6th RMSE = 25.7 mg/L (T1)
AW: 2nd RMSE = 20.1 mg/L (T1)

TK - TEH 4-20 mA outputs
DW: 3rd RMSE = 19.9 mg/L (T1)
WW: 4th RMSE = 20.5 mg/L (T1)
AW: 3rd RMSE = 21.2 mg/L (T2)

TK TURBIDITY
DW: 2nd RMSE = 19.6 mg/L (P2)
WW: 2nd RMSE = 18.8 mg/L (P2)
AW: 1st RMSE = 18.5 mg/L (P3)

TK 4-20 mA output
DW: 4th RMSE = 21.1 mg/L (P1)
WW: 3rd RMSE = 18.9 mg/L (P3)
AW: 4th RMSE = 21.2 mg/L (P3)

TEH TURBIDITY
DW: 5th RMSE = 60.1 mg/L (P1)
WW: 7th RMSE = 26.7 mg/L (P3)
AW: 6th RMSE = 51.2 mg/L (P1)

TEH 4-20 mA output
DW: 8th RMSE = 46.3 mg/L (P3)
WW: 6th RMSE = 25.6 mg/L (P2)
AW: 7th RMSE = 48.8 mg/L (P2)

s: spectra lyser, ABSORBANCE(S)
DW: 7th RMSE = 39.6 mg/L (1,1 : 722.5 nm)
WW: 1st RMSE = 16.5 mg/L (1,1 : 715 nm)
AW: 6th RMSE = 38.1 mg/L (1,1 : 380 nm)

s: spectra lyser TSS
DW: 5th RMSE = 29.9 mg/L (P1)
WW: 6th RMSE = 52.3 mg/L (P1)
AW: 5th RMSE = 35.1 mg/L (P1)
Results (for TSS)
Results (for total COD)

- For each data set
  - DW: 2 wavelengths (EFS)
  - WW: $T_{EH} - T_K - \text{Cond}$
  - AW: $T_{EH} - T_K - \text{Cond}$

- Global conclusions
  - Nephelometric turbidity meter with conductivity meter provides accurate model
  - Spectrometer (local $>$ spectral)
Results

Total COD (mg/L) = b_0 + b_1 \cdot TK (A) + b_2 \cdot Cond (A)

RMSEP = 82.5 mg/L.
Results (for dissolved COD)

- Spectral calibration provides always the best models (200 – 230 nm)
Conclusions (Methods)

- DE method give more accurate models than DM method
- A PLS calibration is not always more accurate than a local calibration
- 4-20 mA outputs and physical units provide models with comparable accuracy
Conclusions (Sensors)

- No sensor is the most accurate one for all pollutants and all data sets.
- Microwave sensor was not adapted to our range of concentrations.
- Use of conductivity seems to improve the accuracy to predict total COD concentrations.
Thanks for your attention

Questions? ...
Models tested

- **Polynomials:** 
  \[ [P] = c_0 + \sum_{i=1}^{N_{\text{MAX}}} c_i \times X^i \]

- **Multivariate models:**
  \[ [P] = c_0 + c_1 \times X + c_2 \times Y + c_3 \times Z \]
  \[ [P] = c_0 + c_1 \times X + c_2 \times Y + c_3 \times X^2 \]
  \[ + c_4 \times Y^2 + c_5 \times X \times Y \]
  \[ [P] = c_0 + c_1 \times X + c_2 \times Y + c_3 \times X \]
  \[ + c_4 \times X^2 + c_5 \times Y^2 + c_6 \times Z^2 \]
  \[ + c_7 \times X \times Y + c_8 \times X \times Z + c_9 \times Y \times Z \]

- **PLS models:** 
  \[ [P] = c_0 + \sum_{i=1}^{N_{\text{YI}}} c_i \times Abs_i \]
Results