Transformation of continuous turbidity measurements into pollutants concentrations in urban drainage systems: detangling global regressions

S. Sandoval[1]*, B. Spahni[1] & F. Favre[1]

[1] *University of applied Sciences of Western Switzerland (HES-SO), HEIA-Fr, ITEC, Boulevard de Pérolles 80, 1700 Fribourg, Switzerland*

**Corresponding author email: santiago.sandoval@hefr.ch*

Highlights

- Turbidity continuous measurements is a well-recognized technique for pollutants monitoring
- Grouping campaigns lead to lower uncertainties in the Turbidity-Concentrations regression(s)
- Findings show a significant impact over uncertainties of environmental indicators

Introduction

Online monitoring has emerged over last decades as a well-recognized technique for water quality monitoring in urban Hydro-Systems. However, sensors employed for this purpose (e.g. turbidimeters, flowmeters, conductivity or water level sensors) installed in the field represent important challenges related to the appropriate treatment and interpretation of the acquired data (Métadier and Bertrand-Krajewski, 2012). This last, given intrinsic aspects of the measured signals such as complex noise structures, outliers and missing values. Aimed to tackle most of these issues, uncertainty-based data treatment techniques have received increasing interest. For example, new methods have been developed in the field of environmental signal processing covering aspects such as: - Signal filtering by linear and non-linear uncertainty-based filters (Kalman Filters, Particle Filters or Bayesian Filters), - Time series uncertainty-based interpolation techniques (Lepot *et al.*, 2017) - Transformation of noisy signals by calibration functions (*e.g.* water levels to flowrate) (Horner *et al.*, 2018). On the other hand, heteroscedasticity and autocorrelation of fitting errors in these calibrated functions are reported to have a significant impact over the propagation of uncertainties in *e.g.* the calculation of environmental indicators (*e.g.* total loads, means concentrations or total runoff volumes). The present study proposes a simplified methodology aimed to reduce the effect of autocorrelation and heteroscedasticity of errors in regressions employed to transform turbidity signals into pollutants concentrations (*e.g.* Total Suspended Solids TSS). The proposed methods are applied to estimate pollutant loads time series from online monitoring and sampling campaigns in two experimental treatment systems, collecting stormwater from a train railway in Switzerland.

Methodology

Experimental site and data description

The monitored system was designed to treat runoff collected from a 9055 m^2 area that drains stormwater from a train railway in Renens, Switzerland. Stormwater from the draining area is collected into a pumping chamber by gravity, where it is pumped into a main detention tank. The pumps are activated when the water level in the pumping chamber is above 90 cm. Then, the inlet to the main tank (IN) is produced as pumping pulses with a duration between 3 to 10 min. The main tank has two identical outlet pipes, which were equipped with two different experimental treatment modules, aimed to treat and release the stormwater to the receiving water body. Three monitoring points were located at IN and at the outlet of the two modules (HEPIA and CFF hereafter), measuring online flowrate and turbidity during 10 months (January to October 2020) with a time-step of 10 seconds. The three (IN, HEPIA and CFF) turbidity and flow rate time series were aggregated to a one-minute time step. A median filter was applied to these time series and outliers were also removed. Standard uncertainties of time series were attributed based on manufacturers recommendations of the sensors. Missing data in time series for gaps shorter than two hours were interpolated, accounting for standard uncertainties and the autocorrelation of the time series (method in Lepot *et al.*, 2017). Stormwater elemental samples were collected by three automated 24 bottles samplers (one per point IN, HEPIA and CFF) during nine rainfall events. For each sample (bottle), TSS and heavy metals (Zn, Cu, Fe, Ni and Cr) laboratory analyses were carried out, as well as Turbidity measurements. Standard uncertainties of concentrations *C* for TSS and heavy metals were attributed as a function of the analytical methods employed and the intra-sample variability from duplicate analyses (about 10 % of the measured concentrations). All turbidimeters (three *in situ* and one in laboratory) were calibrated with standard solutions, propagating uncertainties from: (i) standard uncertainties of the original turbidity measurements and (ii) errors from the fitted regression employed for the calibration (*e.g.* Métadier and Bertrand-Krajewski, 2012).

Turbidity-Concentration relation from detangling global regressions

The transformation of the turbidity signals into *C* (TSS or heavy metals) is proposed to be carried out from the pairs of Turbidity-*C* values and their uncertainties obtained from the elemental samples. However, employing a global Turbidity-*C* regression for each IN, HEPIA and CFF monitoring point (*i.e.* with the totality of the Turbidity-*C* pairs from all the samples collected during the nine events) might introduce biases and heteroscedasticity in the errors of the regression (Hannouche *et al.*, 2017). This last, principally from not accounting for the temporal variability of the Turbidity-*C* relation. On the other hand, adopting a different regression for each event might lead to an insufficient number of points (< 20) to fit the regressions. Therefore, the present study proposes a simplified solution to the aforementioned drawbacks. A grouping strategy of samples from monitored campaigns is aimed to establish an intermediate number of Turbidity-*C* regressions (between the event-based and the global case). This method is based on grouping the Turbidity-*C* data from sampling campaigns by chronological order. All Turbidity-*C* values from campaign *i* are grouped with all Turbidity-*C* values from campaign *i* + 1 if: the number of Turbidity-*C* couples for campaign *i* is lower than 30 and (ii) if the elapsed time between both campaigns *i* and *i* + 1 is lower than 30 days. Regressions were fitted by Ortogonal Distance Regressions ODR (Boggs and Rogers, 1990), considering uncertainties in Turbidity, as well as in *C* values. The best order for the fitted regression was chosen for each group of campaigns among a linear, second- and third-degree polynomial, based on the Bayesian Information Criteria (BIC) (Ward, 2008). An example of a fitted regression is shown in Figure 1 (left) for all pairs Turbidity-Iron from a group of sampling campaigns at CFF.

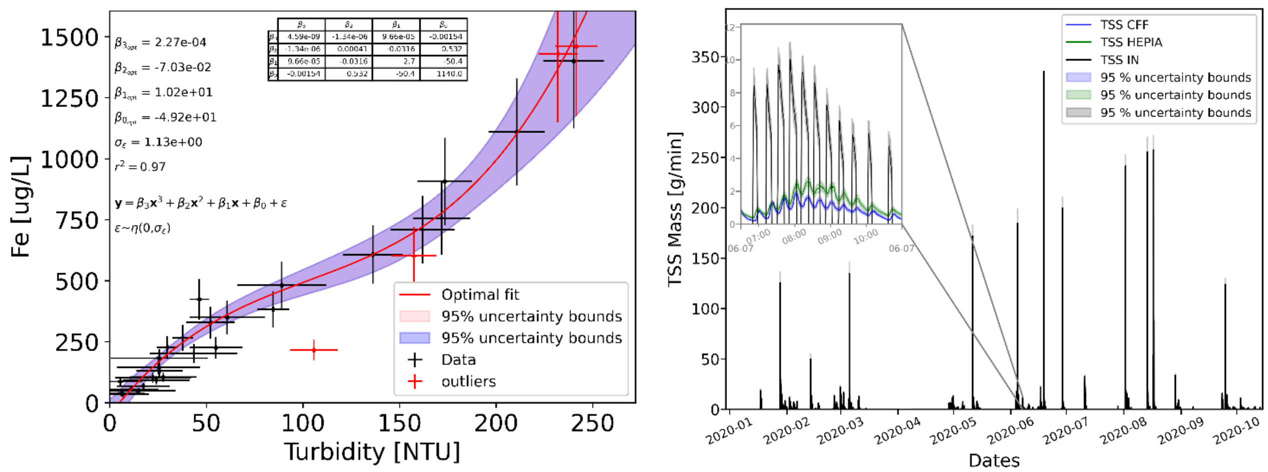

Figure 1. Fitted Regression Turbidity-Iron with data from a group of sampling campaigns at HEPIA (left) and TSS load time series for IN, HEPIA and CFF.

Established regressions are used to transform continuous Turbidity time series from each point (IN, HEPIA and CFF) to *C* (TSS or heavy metals). For each Turbidity value at *t,* the transformed *C* value is calculated as a weighted average from the two regressions fitted for the preceding and subsequent group of campaigns to *t*. Uncertainties in *C* time series are obtained from the propagation of: (i) standard uncertainties of calibrated turbidity time series and (ii) errors from the two neighbour Turbidity-*C* regressions employed for the transformation. Then *C* time series can be used to calculate loads *L* time series as $L = Q \cdot C$. For example, TSS loads are obtained from time series of TSS concentrations (Figure 1).

Results and discussion

Turbidity time series at IN, HEPIA and CFF were transformed into TSS and Iron concentrations by the proposed methodology. Results for the standard uncertainties compared to TSS concentrations over time are shown in Figure 2 for HEPIA (left) and CFF (right) monitoring points.

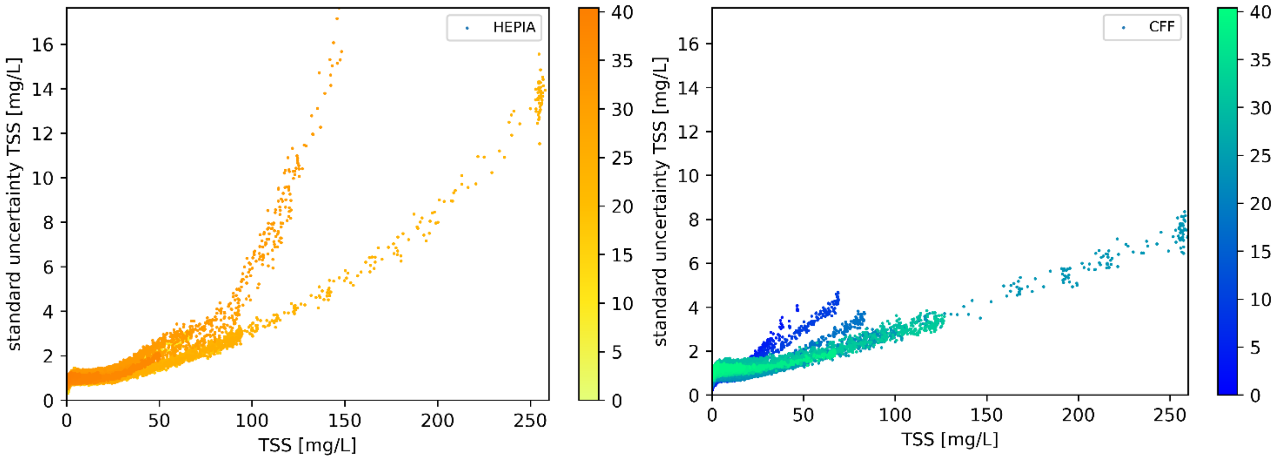

Figure 2. TSS concentrations versus standard uncertainties over time (in colour) for HEPIA (left) and CFF (right)

Three different trends can be identified from Figure 2, due to the three different Turbidity-TSS regressions employed over time (one for each group of campaigns). From the proposed methodology applied to IN, HEPIA and CFF, relative uncertainties in TSS concentrations were reported to be generally lower than 10 %, compared to values of about 14 % obtained from global regressions. The proposed approach brings to a certain extent more homoscedastic, less-autocorrelated residuals in the employed Turbidity-*C* regressions, in contrast to global regressions. These improvements could be expected to be more significant for larger datasets, including a higher heteroscedasticity in the errors of the regression (Hannouche *et al.*, 2017).

Conclusions and future work

The present study proposes a methodology to improve the quality of the transformation of continuous turbidity time series into pollutants concentrations *C* (*e.g.* TSS). A strategy based on grouping events or campaigns is aimed to obtain more adaptable error structures (homoscedastic and less-autocorrelated) in the Turbidity-*C* regression(s) used for this purpose, compared to global regressions. Future work can be oriented to formalize and optimize the trade-off between: the autocorrelation and heteroscedasticity in the residuals (global) / the amount of data included in the regression (event-based). The enhancement of Turbidity-*C* regressions encourages the reduction of uncertainties and biases in the estimation of environmental indicators (*e.g.* total loads, removal efficiencies).

References

Boggs, P. T. and Rogers J. E. (1990). "Orthogonal Distance Regression," in "Statistical analysis of measurement error models and applications: proceedings of the AMS-IMS-SIAM joint summer research conference held June 10-16, 1989," Contemporary Mathematics, vol. 112, pg. 186.

Hannouche, A., Joannis, C. and Chebbo, G. (2017). Assessment of total suspended solids (TSS) event load and its uncertainties in combined sewer system from continuous turbidity measurements. Urban Water Journal, 14(8), 789-796.

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H. K. and Pierrefeu, G. (2018). Impact of stage measurement errors on streamflow uncertainty. Water Resources Research, 54(3), 1952-1976.

Lepot, M., Aubin, J. B. and Clemens F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. Water, 9(10), 796

Métadier, M. and Bertrand-Krajewski J. L. (2012). The use of long-term on-line turbidity measurements for the calculation of urban stormwater pollutant concentrations, loads, pollutographs and intra-event fluxes. Water research, 46(20), 6836-6856.

Ward, E. J. (2008). A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecological Modelling, 211(1-2), 1-10.