Science of the Total Environment 625 (2018) 336–343

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Online monitoring and conditional regression tree test: Useful tools for a better understanding of combined sewer network behavior

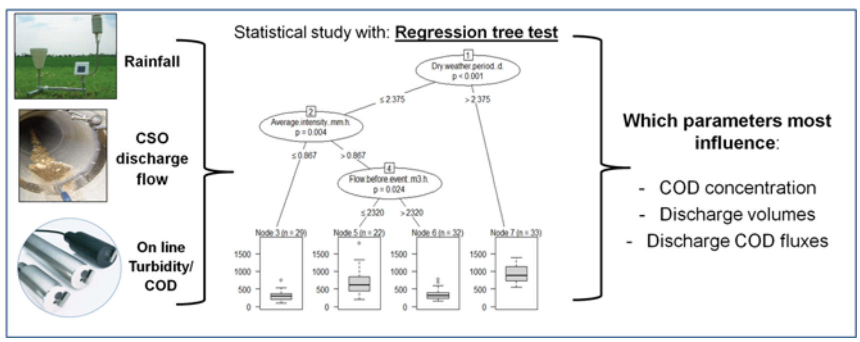
T. Bersinger [a], G. Bareille [a], T. Pigot [b], N. Bru [c], I. Le Hécho [a,*]

[a] CNRS/Univ Pau & Pays Adour, Institut des Sciences Analytiques et de Physico-Chimie Pour l'Environnement et Les Materiaux, UMR5254, 64000 Pau, France
[b] CNRS/Univ Pau & Pays Adour, Institut des Sciences Analytiques et de Physico-Chimie Pour l'Environnement et Les Materiaux, UMR5254, 64600 Anglet, France
[c] CNRS/Univ Pau & Pays Adour, Laboratoire de Mathematiques et de leurs Applications de Pau, UMR5142, 64600 Anglet, France

HIGHLIGHTS

- Which parameters influence COD concentration, discharge flow and flux in sewer
- Conditional regression tree test is used for the first time in wastewater field.
- Three main parameters can explain maximum COD concentration.
- COD discharge flux is mainly driven by discharge volume and then by concentration.
- Continuous measurements and statistical analysis: a useful tool for sewer management

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 November 2017
Received in revised form 20 December 2017
Accepted 20 December 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
Regression trees
Statistical study
Combined sewer system
Continuous monitoring
Turbidity
Pollutant

ABSTRACT

A good knowledge of the dynamic of pollutant concentration and flux in a combined sewer network is necessary when considering solutions to limit the pollutants discharged by combined sewer overflow (CSO) into receiving water during wet weather. Identification of the parameters that influence pollutant concentration and flux is important. Nevertheless, few studies have obtained satisfactory results for the identification of these parameters using statistical tools. Thus, this work uses a large database of rain events (116 over one year) obtained via continuous measurement of rainfall, discharge flow and chemical oxygen demand (COD) estimated using online turbidity for the identification of these parameters. We carried out a statistical study of the parameters influencing the maximum COD concentration, the discharge flow and the discharge COD flux. In this study a new test was used that has never been used in this field: the conditional regression tree test. We have demonstrated that the antecedent dry weather period, the rain event average intensity and the flow before the event are the three main factors influencing the maximum COD concentration during a rainfall event. Regarding the discharge flow, it is mainly influenced by the overall rainfall height but not by the maximum rainfall intensity. Finally, COD discharge flux is influenced by the discharge volume and the maximum COD concentration. Regression trees seem much more appropriate than common tests like PCA and PLS for this type of study as they take into account the thresholds and cumulative effects of various parameters as a function of the target variable. These results could help to improve sewer and CSO management in order to decrease the discharge of pollutants into receiving waters.

* Corresponding author.
E-mail addresses: thomas.bersinger@univ-pau.fr (T. Bersinger), gilles.bareille@univ-pau.fr (G. Bareille), thierry.pigot@univ-pau.fr (T. Pigot), noelle.bru@univ-pau.fr (N. Bru), isabelle.lehecho@univ-pau.fr (I. Le Hécho).

https://doi.org/10.1016/j.scitotenv.2017.12.239
0048-9697/

1. Introduction

The impact of Combined Sewer Overflow (CSO) on receiving waters has been demonstrated many times in the past decade (Even et al., 2007; Passerat et al., 2011; Weyrauch et al., 2010).

The European Water Framework Directive aims to achieve "good ecological and chemical status" for all of Europe's aquatic environments (WFD, 2000). In this context, a better management of CSO and Wastewater Treatment Plants (WWTP) has become a priority. This challenge requires a better understanding of the volume, the concentration or the pollutant flux in the sewage system and at CSO. Different studies have been previously conducted on these subjects.

Some studies have focused on CSO discharge prediction, number of discharge or discharge duration (Mailhot et al., 2015; Montserrat et al., 2015; Thorndahl and Willems, 2008; Yu et al., 2013) but they did not give results about concentration or pollutant flux. Different statistical tools are used such as decision trees, cluster analysis or regression models.

Other authors have focused more on concentration, volume and pollutant flux dynamic (Hannouche, 2012; Lacour et al., 2009b; Langeveld et al., 2005; Metadier and Bertrand-Krajewski, 2012). Most of these studies used continuous turbidity measurements to obtain databases large enough to conduct reliable statistical studies. Recent studies have demonstrated the usefulness of measuring turbidity as a substitute for conventional analyses (Hannouche et al., 2011; Lacour et al., 2009a; Metadier and Bertrand-Krajewski, 2012). The good correlation between turbidity and COD means turbidity measurements can be used to monitor the COD (Bersinger et al., 2015; Lacour et al., 2009b; Lawler et al., 2006; Metadier and Bertrand-Krajewski, 2012). Thereby, using online monitoring means that the dynamic of the studied systems can be obtained with a better temporal resolution than with discrete analyses at laboratory scale, also online monitoring allows reliable statistical analyses to be performed. These large databases using online turbidity have been statistically treated with PCA, correlation analyses and regression models in most cases. Results obtained in these studies show some trends to explain the flow and pollutant dynamic: antecedent dry weather period, rainfall height and intensity seem to be the most important parameters. Nevertheless, no reliable relationship can be statistically demonstrated and the conclusions drawn highlighted the complexity to predict or model sewer system dynamics.

The study by Sandoval et al. (2013) is the only one that achieves the objective of statistical determination of parameters influencing volume, COD concentration (UV–visible spectrometer data) and flux. Canonical Correlations (CC) and Partial Least Squares (PLS) were used in this study and the authors showed that maximum and average intensity are the parameters the most correlated to CSO volume and flux while rainfall height and duration are correlated to pollutant concentration. Nevertheless, the study is based on a small number of events (22).

Even though a large variety of parameters and statistical tests can be used, it is delicate to highlight common parameters linking all of these studies to explain discharge flow and flux at CSO. Prediction of pollutant behavior and volume related to specific rainfall events would allow a better management system to be put in place to decrease the pollutant load discharged into receiving waters.

In this context, the objective of this study is to statistically determine the parameters that most influence: (i) the maximum COD concentration, (ii) the wastewater volume discharge and (iii) the COD flux discharge during a rainfall event using an innovative statistical tool, the regression tree test. This test is able to find nonlinear relationships and to detect some threshold effects, which is not possible with tests used in previous studies. Furthermore, to our knowledge, this test has never been used in this field of research. A large database of 116 rainfall events that occurred over one year was analyzed. For all these events, characteristics of rainfall, flow, pollutant concentration and flux were recorded. The information obtained allows a better understanding of the combined sewer system behavior during various weather conditions.

2. Materials and method

2.1. Study site and online equipment

The combined sewer system of the urban area of the city of Pau (south-west of France) covers about 50 km^2 with 150,000 inhabitants. The site is mainly composed of residential areas with several commercial and industrial areas. The average level of imperviousness is 80%, ranging from 60% for residential to 90% for highly urbanized areas (town center). More than three-quarters of the population are served by a combined sewer system (CSS); the other 25% are connected to a separate sewer. The CSS is connected to an activated sludge WWTP (Wastewater Treatment Plant) with a capacity of 190,000 Population Equivalent. At the entrance of the WWTP during a representative dry weather period, the flow is comprised between 800 m^3/h (night) and 2200 m^3/h (day) with an average of 1400 m^3/h. The COD dry weather concentration values vary from 100 mg O_2/L (night) to 800 mg O_2/L (maximum during the day) and the average value is 400 mg O_2/L.

This study was focused on the main CSO of the urban area (60% of the total discharge volume) that is located at the entrance of the WWTP. At this point, discharge and treated flow where both measured each hour with an acoustic flow meter (Doppler principles Mainstream, Hydreka). A turbidity sensor (NF EN ISO 7027, 2000) (Solitax, Hach Lange) was installed just upstream of the CSO allowing continuous measurement of the turbidity of the wastewater going in the CSO and in the WWTP. Every 5 min, the mean turbidity was calculated by integration of 15 s of data. One rain gauge (tipping-bucket rain gauge, Pulsonic) was installed in the center of the watershed to monitor rainfall characteristics, the measurement was taken every hour. The provided data correspond to the cumulative rainfall height and cumulative flow during 1 h. Considering the important length of the sewer network (around 800 km), the average water transfer duration is important: 2 h. So the data recording interval of 1 h fits the reaction time of the studied sewer. Moreover most of the studied variables (target and explanatory) are cumulative ones. More information about the study site and sensors used are available in Bersinger et al. (2015).

2.2. Evaluation of COD concentration and flux

To determine the relationships between COD and turbidity, 108 wastewater samples were collected and tested during 10 different sampling sessions. Linear positive correlations between turbidity/COD were found to be consistent. The corresponding linear model is given below:

$$COD_i = T_i \times 2.85 + 44.96 \ (r^2 = 0.91) \tag{1}$$

With COD_i concentration in mg O_2/L at time i and T_i turbidity (FNU) recorded at time i. The establishment of these correlations is detailed in Bersinger et al. (2015). The correlation coefficient (r^2) over 0.9 shows, in accordance with the statistical test of residuals, that this relationship is acceptable for online evaluation of COD concentration from turbidity data. An uncertainty of 15% was calculated on the COD concentration $u(C_{CODi})$ taking into account uncertainty on turbidity measurement, uncertainty due to the correlation function and using the law of propagation of uncertainties (LPU).

To calculate the hourly COD flux (Fh_{COD}) (kg O_2/h), average turbidity is calculated every hour and the associated uncertainty $u(Fh_{COD})$ was calculated with Eq. (3) according to Metadier and Bertrand-Krajewski (2011). Fh_{COD} was calculated from the continuous pollutant time series measured at each time step i between the beginning of the hour (bh) and the end of the hour (eh):

$$Fh_{COD} = \Delta t \cdot \sum_{i=bh}^{eh} Qi \cdot C_{CODi} \tag{2}$$

For C_{CODi} the concentration of COD was estimated from the turbidity at the time step i. Qi is the discharge flow at time step i and Δt the data acquisition time step.