The evaluation of rainfall influence on combined sewer overflows characteristics: the Berlin case study

S. Sandoval, A. Torres, E. Pawlowsky-Reusing, M. Riechel and N. Caradot

ABSTRACT

The present study aims to explore the relationship between rainfall variables and water quality/quantity characteristics of combined sewer overflows (CSOs), by the use of multivariate statistical methods and online measurements at a principal CSO outlet in Berlin (Germany). Canonical correlation results showed that the maximum and average rainfall intensities are the most influential variables to describe CSO water quantity and pollutant loads whereas the duration of the rainfall event and the rain depth seem to be the most influential variables to describe CSO pollutant concentrations. The analysis of partial least squares (PLS) regression models confirms the findings of the canonical correlation and highlights three main influences of rainfall on CSO characteristics: (i) CSO water quantity characteristics are mainly influenced by the maximal rainfall intensities, (ii) CSO pollutant concentrations were found to be mostly associated with duration of the rainfall and (iii) pollutant loads seemed to be principally influenced by dry weather duration before the rainfall event. The prediction quality of PLS models is rather low ($R^2 < 0.6$) but results can be useful to explore qualitatively the influence of rainfall on CSO characteristics.

Key words | canonical correlation, CSO, online monitoring, partial least squares regression, UV–Vis spectrometry

S. Sandoval
A. Torres
Grupo de Investigación Ciencia e Ingeniería del Agua y el Ambiente,
Facultad de Ingeniería,
Pontificia Universidad Javeriana, Edificio J.G. Maldonado, S.J., Carrera 7 No. 40-62,
Bogotá,
Colombia

E. Pawlowsky-Reusing
Berliner Wasserbetriebe,
Netz- und Anlagenbau, Neue Jüdenstrasse 1, 10864 Berlin,
Germany

M. Riechel
N. Caradot (corresponding author)
Kompetenzzentrum Wasser Berlin gGmbH,
Cicerostraße 24, 10709 Berlin,
Germany
E-mail: *nicolas.caradot@kompetenz-wasser.de*

INTRODUCTION

The understanding of combined sewer overflow (CSO) pollutant dynamics requires detailed analyses of the variability and complexity of processes and interactions involved. The mentioned variability in wastewater and stormwater has already been addressed by multiple authors (e.g. Barraud *et al.* 2002; Veldkamp *et al.* 2002; Grüning & Orth 2002; Gruber *et al.* 2005; Hochedlinger *et al.* 2006; Bertrand-Krajewski *et al.* 2007; Aumond & Joannis 2008; Lacour *et al.* 2009; Schilperoort 2011; Métadier & Bertrand-Krajewski 2012). To better understand the processes and interactions related to this variability, different statistical methods can be applied (e.g. Grum & Aalderink 1999; Goodwin *et al.* 2003; Olyphant *et al.* 2003; Neuman *et al.* 2004; Even *et al.* 2007).

In the city of Berlin, a continuous integrated monitoring, using state-of-the-art online sensors, was started in 2010. It combined (i) continuous measurements of water quality and flow rates of CSO at one main CSO outlet and (ii) continuous measurements of water quality parameters at four sites within the urban stretch of the receiving river. Locally, the collection of data was also aimed at (i) characterizing CSO emissions, (ii) assessing the local dynamics and intensity of CSO impacts on the river and (iii) calibrating sewer and river water quality models as part of a planning tool for future CSO management in Berlin (Riechel *et al.* 2012). By means of characteristics of 22 online monitored rainfall and CSO events (e.g. total masses/volumes, mean and maximum concentrations and flow), multivariate statistical methods were implemented in order to (i) establish associations between rainfall and CSO characteristics (canonical correlation) and (ii) develop rainfall–CSO characteristics models (partial least squares (PLS) regression). The selected statistical methods could make it possible to: (i) identify which rainfall variable seems to be the most important in terms of prediction of each CSO characteristic; and (ii) verify the predictability of the models. Furthermore, uncertainties in the rainfall and CSO characteristics were assessed based on the variability of the PLS regression using the Monte Carlo method.

doi: 10.2166/wst.2013.524

MATERIALS AND METHODS

Study site

The CSO monitoring station was installed in a major overflow sewer about 230 m downstream of the main overflow structures and 500 m upstream of the outlet to the river. The station is located in a densely populated residential area with very few industrial activities. The catchment has ~126,000 inhabitants, with a total impervious area of about 800 ha. The total area of the sub-catchment upstream of the outlet of the CSO monitoring is 100 ha and the impervious area of this sub-catchment is 74 ha. The sewer system is fully combined. Flow was measured directly in the overflow sewer, based on water level (air-ultrasound, Nivus company) and flow velocity (two ultrasound sensors, Nivus company). Water quality was measured on a bypass fed by a peristaltic pump, using (i) a UV–visible spectrometer (spectro::lyser, s::can company, Vienna) for the measurement of absorption water spectra which enable calculating equivalents of chemical oxygen demand (CODeq), dissolved COD (CODdeq) and total suspended solids (TSSeq) and (ii) a conductivity sensor (condu::lyser, s::can company, Vienna) for measurements of electrical conductivity (EC) and temperature. Rainfall measurements have been considered from four rain gauges (Figure 1) located around the CSO monitoring. Details about monitoring implementation can be found in Caradot (2012).

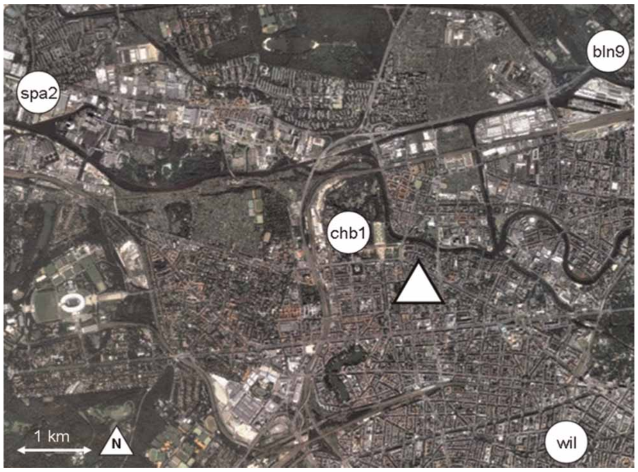

Figure 1 | Situation of the CSO monitoring station (white triangle) and of the four considered rain gauges: Berlin 9 (bln9), Spandau 2 (spa2), Charlottenburg 1 (chb1) and Wilmersdorf (wil).

CSO and rainfall characteristics

CSO and rainfall characteristics (see Table 1) as well as their uncertainties have been calculated for 22 events during the

Table 1 | CSO and rainfall characteristics for 22 events during the monitoring period

Rainfall characteristics						CSO characteristics					
Charac.	**Name**	**Unit**	**Mean**	**Min.**	**Max.**	**Charac.**	**Name**	**Unit**	**Mean**	**Min.**	**Max.**
dur	Duration	min	236	45	965	dur	Duration	min	55	9	163
imax_bln9	Max. intensity	mm/h	26.5	7.2	94.8	max_Q	Max. flow	m^3/s	1.6	0.6	3.5
imax_wil	Max. intensity	mm/h	27.9	1.2	115.2	V	Volume	m^3	3,452	304	12,375
imax_spa2	Max. intensity	mm/h	23.5	4.8	73.2	mean_Q	Mean flow	m^3/s	1.0	0.4	2.0
imax_chb1	Max. intensity	mm/h	22.4	3.6	103.2	M_COD	COD load	kg	1,186	68	4,003
h_chb1	Depth	mm	10.4	2.3	25.6	M_TSS	TSS load	kg	784	46	2,765
h_bln9	Depth	mm	12.5	2.6	31.8	M_CODd	CODd load	kg	209	12	775
h_wil	Depth	mm	12.5	1.8	29.1	mean_TSS	Mean TSS	mg/l	248	22	580
h_spa2	Depth	mm	12.7	1.5	34.9	mean_COD	Mean COD	mg/l	377	35	908
imoy_chb1	Average intensity	mm/h	4.4	0.6	13.5	mean_CODd	Mean CODd	mg/l	69	9	158
imoy_bln9	Average intensity	mm/h	4.6	1.1	11.8	mean_EC	Mean EC	μS/cm	262	25	549
imoy_wil	Average intensity	mm/h	5.0	0.2	14.6	C_vol _waste	Wastewater ratio (volume)	%	11	0	34
imoy_spa2	Average intensity	mm/h	4.5	0.9	13.8	C_COD _waste	Wastewater ratio (COD load)	%	16	0	46
DWdur	Dry weather duration	day	3.9	0.0	18.9						

monitoring period. Rainfall characteristics have been determined at the four considered rain gauges presented in Figure 1. Total measurements uncertainty has been estimated accounting for errors from sensor, lab chain, local calibration and field conditions (representativeness of site, see Métadier & Bertrand-Krajewski (2011)).

Over the 22 rainfall events, CSO volume ranges between 304 and 12,375 m³ and CSO COD load varies between 68 and 4,003 kg. Total COD load is 29 t ± 20%. COD mean concentration is 377 mg/l but it shows extreme variations from 35 to 908 mg/l. The average proportion of wastewater in the CSO (C_vol_waste) is 11% and the average contribution of wastewater to total CSO load (C_vol_waste) is 16%. This means that about 84% of the COD carried in the CSO originates from two other major sources: wash-off by stormwater runoff and resuspension of sewer sediments. For a full description of the CSO characteristics and calculation details, see Caradot *et al.* (2013).

Canonical correlation analysis

Relations between rainfall variables and CSO characteristics could be explored by means of multivariate statistical methods. Canonical correlation is a way of evaluating the linear relationship between two multidimensional data sets (Leurgans *et al.* 1993; González *et al.* 2008). The proposed method develops a number of functions that maximizes the correlation among canonical variates, which are orthogonal linear combinations of the original CSO and rainfall variables.

A couple of vectors a_1 and b_1 are found by maximizing the correlation function $corr(a_1^T X, b_1^T Y)$ between the two sets of variables X (input rainfall characteristics) and Y (output CSO characteristics), allowing the calculation of a first pair of canonical variates $CVx_1 = a_1^T X$ and $CVy_1 = b_1^T Y$. From this point, successive pairs of canonical variates CVx_i and CVy_i are computed with additional vectors a_i and b_i for explaining the residual correlation between the two sets of variables X and Y, i.e. the correlation between X and Y not accounted for by the first pair of canonical variates (Hair *et al.* 2005). Each additional pair of canonical variates CVx_i and CVy_i is orthogonal and independent of all other variates and the maximum number of canonical variates equals the number of variables in the smallest dataset, independent or dependent (Leurgans *et al.* 1993; González *et al.* 2008).

The canonical loadings (ρ) measure the linear correlations between the original variables and the canonical variates. These expressions can be computed as: $\rho = corr(CVx_i, X[:k])$ for a given rainfall characteristic k and the canonical variate CVx_i; and $\rho = corr(CVy_i, Y[:,j])$ for a given CSO characteristic j and the canonical variate $CVyi$. Therefore, the canonical loadings may lead to the identification of which rainfall variables are correlated with CSO characteristics, by analyzing the correlation of each individual variable over statistically significant (p-value <0.05) canonical variates (Hair *et al.* 2005).

PLS regression

PLS is a regression method widely used in chemistry, especially in chromatography and spectrometry applications (Tenenhaus 1998) where the number of variables (i.e. wavelengths) that characterize the spectra is very large compared to the number of observations, and where the prediction variables are highly correlated (Aji *et al.* 2003). This multicollinearity problem (and also the singularity ones) of the matrix composed by the independent variables is solved with PLS, by generalizing and fusing the principal component analysis and the multiple regression methods (Abdi 2003; Lee *et al.* 2011).

The PLS regression can integrate multiple dependent and independent variables (see Abdi 2003; Bertrand 2005). However, whenever more than one dependent variable is included (multivariate output), the PLS model loses accuracy, as well as predictability (Martínez Galera *et al.* 1997). Thus, multiple PLS models could be formulated by considering in all cases rainfall characteristics as the multivariate input, testing with different CSO characteristics as an individual output, adjusting the model parameters by a cross-validation procedure (e.g. a PLS model can be formulated with the following architecture: rainfall mean intensity, duration of the rainfall and total rainfall volume as three inputs and the TSS mean concentration as the unique output).

The PLS regressions can also bring up the identification of which variable of the multivariate independent variables set (rainfall characteristics) is the most representative over the specific output variable to be modeled (a given CSO characteristic), by the weights of the adjusted parameters of the regression (Rada Ariza & Torres 2011). Therefore, the PLS regression results to be obtained are: (i) the identification of the most important rainfall variable (related to the prediction of a CSO characteristic); and (ii) the predictability of the model (measured by the determination coefficient R^2). However, these results are sensitive to uncertainties, not only of the rainfall variables, but also of the CSO water quality/quantity characteristic to be predicted. Thus, the representativeness of these uncertainties in the results for (i) and (ii) (for any of the CSO characteristics) is assessed by the implementation of Monte Carlo simulations. The method consists of the repetitive generation of a set of random numbers that belong to the uncertainty interval of the different rainfall and CSO values (where each repetition will be a simulation). The results for (i) and (ii) can be estimated several times, according to the number of simulations, which can lead to identification of the number of times in which a specific rainfall characteristic was identified as the most important variable (probability of most important), as well as the variability and trend of the R^2 (adjustment of the PLS model), in terms of explaining a given CSO characteristic.

RESULTS AND DISCUSSION

Canonical correlation analysis

From the canonical correlation analysis, only two canonical variates (CV5 and CV6) appeared to be statistically significant. For both canonical variates, the correlation between CVx and CVy is close to one (e.g. $corr(a5^{\mathrm{T}}X, b5^{\mathrm{T}}Y) \sim 1$) so in the following CV refers to CVx = CVy. Table 2(a) shows the canonical loadings (ρ) obtained between CV5 and CV6 and the rainfall characteristics presented in Table 1 (independent variables). On the other hand, Table 2(b) shows the canonical loadings (ρ) obtained between canonical variates CV5 and CV6 and the CSO characteristics presented in Table 1 (response variables). From these tables, ρ values higher than 0.4–0.5 between CV5 and CV6 and both rainfall and CSO characteristics indicate possible influences of rainfall variables on CSO. Two complementary patterns can be identified:

(i) From the analysis of CV5, the maximum rainfall intensity and the average rainfall intensity seemed to have an influence on CSO quantity (volume, maximum and average flow rates) and CSO pollutant loads (TSS, COD and dissolved COD) (Table 2(a) and (b), in grey). Since the main CSO load contributions are the stormwater runoff and the resuspension of sewer sediments, the maximum rainfall intensity may influence the erosion and transport of particles in the wash-off.

(ii) From the analysis of CV6, the rain duration and the rain depth seemed to have an influence on CSO water quantity (duration and volume) and CSO water quality (mean concentration, EC and proportion of wastewater). The combination of rain duration and depth seems to be a good indicator to describe the CSO dilution and thus the contribution of wastewater to the CSO load.

These correlations results have been obtained using the maximum rainfall intensity calculated with a time step of 30

Table 2 | Canonical loadings (ρ) between (a) explanatory rainfall variables and (b) response variables with significant canonical variates CV5 and CV6

(a)	Correlations (canonical loadings ρ) between original *X* variables and significant canonical variates (CV)		(b)	Correlations (canonical loadings ρ) between original *Y* variables and significant canonical variates (CV)	
	CV5	CV6		CV5	CV6
dur	0.24	0.58	dur	0.00	0.64
imax_bln9	−0.43	0.35	max_Q	−0.54	0.39
imax_wil	−0.16	0.18	V	−0.40	0.72
imax_spa2	−0.45	0.49	mean_Q	−0.64	0.09
imax_chb1	−0.31	−0.34			
h_chb1	0.11	0.37	M_COD	−0.63	0.12
h_bln9	0.07	0.84	M_TSS	−0.51	0.23
h_wil	0.11	0.72	M_CODd	−0.72	0.13
h_spa2	−0.07	0.89			
imoy_chb1	−0.45	−0.45	mean_TSS	−0.24	−0.55
imoy_bln9	−0.52	−0.17	mean_COD	−0.41	−0.57
imoy_wil	−0.07	−0.25	mean_CODd	−0.47	−0.62
imoy_spa2	−0.66	0.06	mean_EC	−0.23	−0.71
DWdur	−0.21	0.22	C_vol_waste	−0.29	−0.71
			C_COD_waste	0.11	−0.60

minutes. This time step was estimated to be relevant considering the response time of the catchment. However, since the values of maximal rainfall intensity depend strongly on this time step, additional correlations have been performed with rainfall intensities related to shorter time steps of 5 and 15 minutes, obtaining similar results (results not shown here).

PLS regression

Several PLS models were proposed for each CSO characteristic as output variable, by using all the rainfall variables as explanatory variables (including rainfall maximum intensities for 5, 15 and 30 min). One thousand Monte Carlo simulations were undertaken for each CSO characteristic as the output of the model, taking into account the uncertainties of rainfall, pollutant concentrations and masses and flow rate measurements (a PLS model was constructed for each simulation and CSO characteristic, reaching a total of 13,000 calibrated models). Figure 2 shows an example of simulation results for COD mean concentration. In addition, an example of PLS architecture for a specific Monte Carlo simulation is also reported (Equation (1)).

$$\begin{aligned} \mathrm{mean_{COD}} &= 0.003*\mathrm{imax}_{\mathrm{chb1}_{30}} - 0.005*\mathrm{imax}_{\mathrm{bln9}_{30}} \\ &- 0.001*\mathrm{imax}_{\mathrm{wil}_{30}} - 0.500*\mathrm{dur} + 0.010*\mathrm{imax}_{\mathrm{chb1}_{15}} \\ &+ 0.006*\mathrm{imax}_{\mathrm{bln9}} - 0.004*\mathrm{imax}_{\mathrm{bln9}_{15}} - 0.013*\mathrm{imax}_{\mathrm{wil}} \\ &- 0.004*\mathrm{imax}_{\mathrm{wil}_{15}} - 0.000*\mathrm{imax}_{\mathrm{spa2}} - 0.003*\mathrm{imax}_{\mathrm{spa2}_{15}} \\ &- 0.009*\mathrm{imax}_{\mathrm{spa2}_{30}} - 0.007*\mathrm{h}_{\mathrm{chb1}} - 0.021*\mathrm{h}_{\mathrm{bln9}} - 0.016*\mathrm{h}_{\mathrm{wil}} \\ &- 0.026*\mathrm{h}_{\mathrm{spa2}} 0.006*\mathrm{imoy}_{\mathrm{chb1}} + 0.002*\mathrm{imoy}_{\mathrm{bln9}} + 0.003 \\ &*\mathrm{imoy}_{\mathrm{wil}} + 0.000*\mathrm{imoy}_{\mathrm{spa2}} + 0.011*\mathrm{DWdur} + 0.038*\mathrm{imax}_{\mathrm{chb1}} \\ &+ 435.608 \end{aligned} \tag{1}$$

For this case, the highest regression coefficient was $\beta_4 = 0.5$ (related to the rain duration). Therefore, for this specific Monte Carlo simulation and PLS model, the mentioned rainfall variable will be listed as the most important (see the results for all simulations as 'probability of most important variable' in Figure 3(b)). In order to summarize the results, Figure 3(a) shows the probability that a given rainfall variable is the most important in terms of predicting each CSO characteristic. The variation and trend of the adjustment for each PLS model was assessed by boxplots of the determination coefficient R^2 (Figure 3(b)).

Figure 3(a) suggests that the duration of the rainfall event (dur_rain), the maximum intensity registered at each rain gauge (imax) and the dry weather duration before the rain (DWdur) are the most important variables for reproducing the studied CSO characteristics, by the calibrated PLS models. Results underline three main influences of rainfall on CSO characteristics:

(i) CSO water quantity: the maximum rainfall intensities (imax) seemed to be the most important rainfall variable for predicting the duration of CSO (dur), the maximum and average flow rates during CSO (max_Q and mean_Q) and the CSO volume (V).

(ii) CSO water quality: the duration of the rainfall event (dur_rain) showed to be the most important rainfall variable in terms of predicting the mean electrical conductivity (mean_EC) and the mean pollutant concentrations (mean_COD, mean_TSS and mean_CODd). The duration of the rainfall event can be also a good estimator of the CSO dilution (C_vol_waste) and thus of the mean pollutant concentrations.

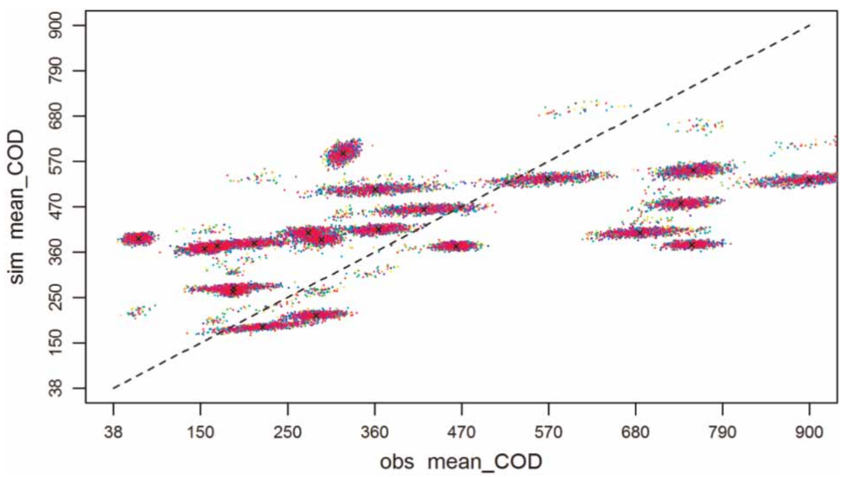

Figure 2 | Observed/simulated mean COD values for 1000 PLS simulations (unit mg/l); 1000 sets of random rainfall and CSO values within their uncertainty interval are generated.

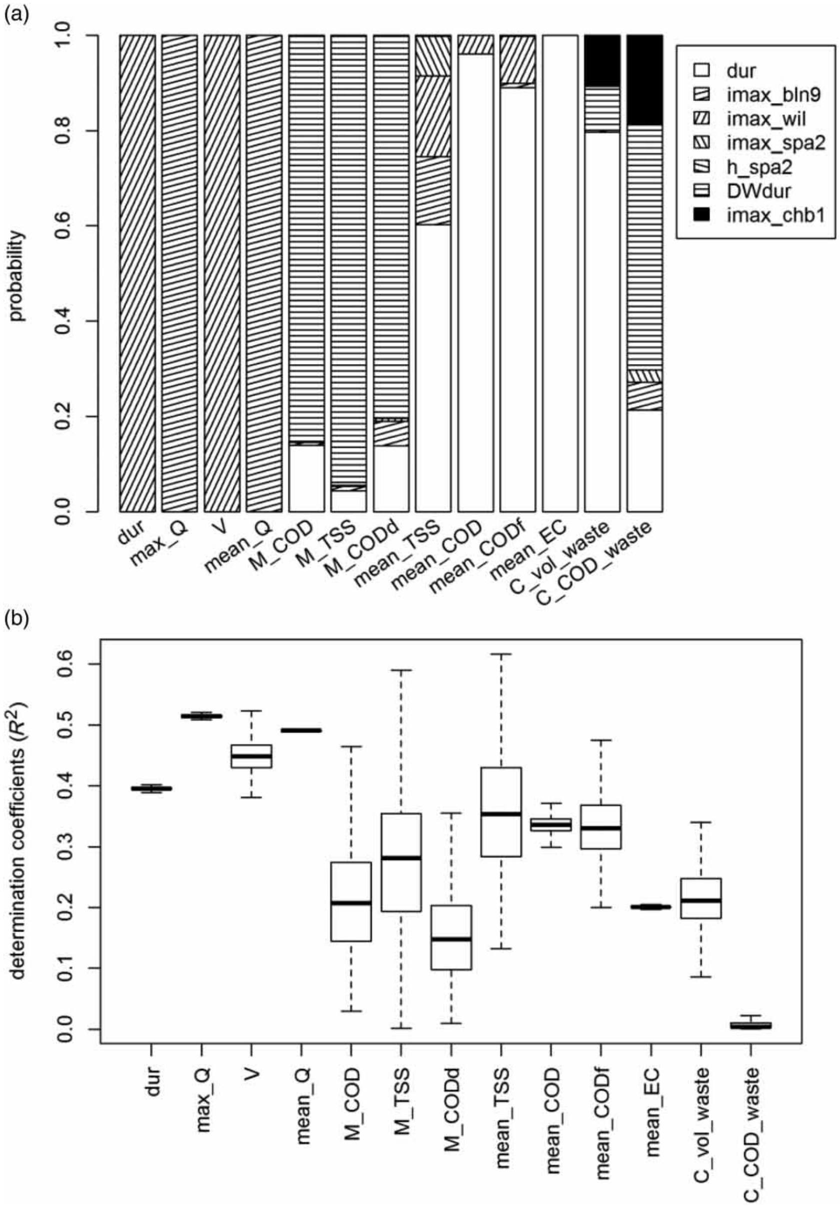

Figure 3 | (a) Probability of being the most important rainfall variable and (b) determination coefficients (R^2), for prediction of CSO characteristics (x-axis) by PLS models (variability assessed by Monte Carlo method).

(iii) CSO pollutants loads: the dry weather duration before the rainfall event (DWdur) showed to be a relevant rainfall variable for reproducing the total loads of pollutants (M_COD, M_TSS and M_CODd). The dry weather duration may influence the amount of suspended solids washed off by the rain runoff and the quantity of in-sewer sediments. The results confirm the high contribution of rain runoff and resuspension of sewer sediments to total CSO loads (see also Passerat *et al.* 2011; Caradot *et al.* 2013).

In addition, it can be observed that, for some CSO variables (duration of CSO, maximum and average flow rates, volume and mean conductivity), the PLS models were constructed with the same most important variable, for all the Monte Carlo simulations (Figure 3(a)). For the remaining CSO variables, the PLS models had more variability in terms of their architecture, leading to the conclusion that these CSO variables possibly contain more uncertainty in their measurements or their prediction is more sensitive to uncertainty in rainfall. This fact is underlined by the strong variabilities of R^2 for PLS models constructed for COD, TSS and dissolved COD loads as well as for TSS and dissolved COD mean concentrations (Figure 3(b)).

Regarding the quality of predictions, Figure 3(b) shows R^2 lower than 0.6, indicating low accuracies of the PLS models as well as low prediction properties, especially for COD, TSS and dissolved COD loads and the contribution of wastewater to CSO load (C_COD_waste). Even if results show low prediction quality, models can be useful for exploring qualitatively the influence of rainfall on CSO characteristics.

CONCLUSIONS

Preliminary results from the canonical correlation analysis indicate that the maximum and average rainfall intensities have an influence on CSO quantity (volume, maximum and average CSO flow rates) as well as on CSO pollutant loads (TSS, COD and dissolved COD) measured at a main CSO outlet in Berlin (Germany). On the other hand, the rain duration and the rain depth seemed to have an influence on CSO water quantity (duration and volume) and CSO water quality (mean concentration, EC and proportion of wastewater).

Results from PLS regressions combined with Monte Carlo simulations to consider uncertainties, partially confirm findings from the canonical correlation analysis. PLS highlights three main influences of rainfall on CSO characteristics: (i) CSO quantity is mainly influenced by the maximal intensity of the rainfall; (ii) CSO pollutant concentrations are mainly influenced by the duration of the rainfall; (iii) CSO pollutant loads are mainly influenced by the dry weather duration before the rainfall.

Regarding variables related to CSO pollutant concentrations and loads, the PLS models show high variabilities in terms of architecture and quality of prediction, revealing high measurement uncertainties. Furthermore, low determination coefficients (lower than 0.6) were obtained showing that the proposed PLS models might not be suitable for prediction purposes, although they can be useful for exploring the qualitative influence of rainfall on CSO characteristics.

The results underline the potential of statistical modelling of CSO quantity and quality characteristics for defining future data acquisition plans or real-time pollution control operation based on the rainfall characteristics mentioned above. With the aim of building models with a higher predictability, further investigations concerning the following aspects are therefore recommended: (i) implementing other linear and non-linear statistical models (e.g. artificial neural networks, support vector machines, kernel methods and fuzzy models); and (ii) testing models with additional CSO input variables (e.g. first flush indicators).

Incoming studies can apply the proposed methods for different urban catchments and water systems (e.g. urban rivers, wastewater treatment plants and sewer networks). On the other hand, since online river impact measurements are also available in Berlin, linkages between rainfall, CSO and resulting river impacts can be examined as well by future research.

ACKNOWLEDGMENTS

The presented work was carried out in the framework of the KWB research project MIA-CSO. The project is funded by Veolia Water and Berliner Wasserbetriebe. The authors would like to thank all colleagues from Berliner Wasserbetriebe, Berlin Senate Department for Urban Development and the Environment (SenStadtUm) and Veolia Water who supported the project.

REFERENCES

Abdi, H. 2003 Partial least squares (PLS) regression. In: *Encyclopedia of Social Sciences Research Methods* (M. Lweis-Beck, A. Bryman & T. Futing, eds). Sage, Thousand Oaks, CA, USA, pp. 792–795.

Aji, S., Tavolaro, S., Lantz, F. & Faraj, A. 2003 Apport du bootstrap à la régression PLS: application à la prédiction de la qualité des gazoles (Contribution of bootstrap techniques to PLS regression: application to the prediction model of gas-oil quality control). *Oil Gas Science and Technology* **58** (5), 599–608.

Aumond, M. & Joannis, C. 2008 Processing sewage turbidity and conductivity recorded in sewage for assessing sanitary water and infiltration/inflow discharges. *Proceedings of the 11th International Conference on Urban Drainage*, Edinburgh, Scotland, UK, 31 Aug to 5 Sept.

Barraud, S., Gibert, J., Winiarski, T. & Bertrand-Krajewski, J.-L. 2002 Implementation of a monitoring system to measure impact of stormwater runoff infiltration. *Water Science and Technology* **45** (3), 203–210.

Bertrand, D. 2005 Etalonnage multidimensionnel: application aux données spectrales (Multi-dimensional calibration: application to spectral data). *Les Techniques de l'Ingénieur*, article P264, mars 2005, 21 pp. + annexes.

Bertrand-Krajewski, J.-L., Barraud, S., Gibert, J., Malard, F., Winiarski, T. & Delolme, C. 2007 The OTHU case study: integrated monitoring of stormwater in Lyon, France (Chapter 23). In: *Data Requirements for Integrated Urban Water Management*. Urban Water Series – UNESCO IHP

(T. Fletcher & A. Deletic, eds). Taylor and Francis, London, UK, pp. 303–314.
Caradot, N. 2012 *Continuous Monitoring of Combined Sewer Overflows in the Sewer and the Receiving River: Return on Experience*. Report Kompetenzzentrum Wasser Berlin. Available at: http://www.kompetenz-wasser.de/fileadmin/user_upload/pdf/forschung/MIA-CSO/Bericht_Continous_Monitoring_of_Combined_Sewer_Overflows_in_the_Sewer_and_the_Receiving_River_Return_on_Experience.pdf.
Caradot, N., Matzinger, A., Riechel, M., Sonnenberg, H., Pawlowski-Reusing, E., Heinzmann, B., von Seggern, D. & Rouault, P. 2013 The use of continuous sewer and river monitoring data for CSO characterization and impact assessment. In: *Proceedings of Novatech 2013, 8th International Conference on Sustainable Techniques and Strategies in Urban Water Management*, Lyon, France. Available at: http://www.kompetenz-wasser.de/fileadmin/user_upload/pdf/forschung/MIA-CSO/MIA-CSO_D23_Monitoring_data_analysis.pdf.
Even, S., Mouchel, J. M., Servais, P., Flipo, N., Poulin, M., Blanc, S. & Paffoni, C. 2007 Modelling the impacts of combined sewer overflows on the river Seine water quality. *Science of the Total Environment* **375** (1), 140–151.
González, I., Déjean, S., Martin, P. G. P. & Baccini, A. 2008 CCA: an R package to extend canonical correlation analysis. *Journal of Statistical Software* **23** (12), 1–14.
Goodwin, T. H., Young, A. R., Holmes, M. G., Old, G. H., Hewitt, N., Leeks, G. J. & Smith, B. P. 2003 The temporal and spatial variability of sediment transport and yields within the Bradford Beck catchment, West Yorkshire. *Science of the Total Environment* **314**, 475–494.
Gruber, G., Winkler, S. & Pressl, A. 2005 Continuous monitoring in sewer networks an approach for quantification of pollution loads from CSOs into surface water bodies. *Water Science and Technology* **52** (12), 215–223.
Grum, M. & Hans Aalderink, R. 1999 Uncertainty in return period analysis of combined sewer overflow effects using embedded Monte Carlo simulations. *Water Science and Technology* **39** (4), 233–240.
Grüning, H. & Orth, M. 2002 Investigations of the dynamic behaviour of the composition of combined sewage using online analyzers. *Water Science and Technology* **45** (4–5), 77–83.
Hair, J. F., Black, B., Babin, B., Anderson, R. E. & Tatham, R. L. 2005 *Multivariate Data Analysis*. 6th edn, Prentice Hall, Upper Saddle River, NJ, USA.
Hochedlinger, M., Kainz, H. & Rauch, W. 2006 Assessment of CSO loads based on UV/VIS-spectroscopy by means of different regression methods. *Water Science and Technology* **54** (6–7), 239–246.
Lacour, C., Joannis, C. & Chebbo, G. 2009 Assessment of annual pollutant loads in combined sewers from continuous turbidity measurements: sensitivity to calibration data. *Water Research* **43** (8), 2179–2190.
Lee, H., Lee, Y., Cho, H., Im, K. & Kim, Y. S. 2011 Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decision Support Systems* **52** (1), 207–216.
Leurgans, S. E., Moyeed, R. A. & Silverman, B. W. 1993 Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society*. Series B **55**, 725–740.
Martínez Galera, M., Martínez Vidal, J. L., Garrido Frenich, A. & Gil García, M. D. 1997 Evaluation of multi-wavelength chromatograms for the quantification of mixtures of pesticides by high-performance liquid chromatography-diode array detection with multivariate calibration. *Journal of Chromatography, A* **778** (1–2), 139–149.
Métadier, M. & Bertrand-Krajewski, J.-L. 2011 Assessing dry weather flow contribution in TSS and COD storm event loads in combined sewer systems. *Water Science and Technology*, **63** (12), 2983–2991.
Métadier, M. & Bertrand-Krajewski, J.-L. 2012 Pollutographs, concentrations, loads and intra-event mass distributions of pollutants in urban wet weather discharges calculated from long term on line turbidity measurements. *Water Research* **46** (20), 6836–6856.
Neuman, M. B., Ort, C., Daebel, H. & Gujer, W. 2004 Legal criteria, variability and uncertainty in the design of CSO detention. *Proceedings of Novatech, International Conference on Sustainable Techniques and Strategies in Urban Water Management*, Lyon, France.
Olyphant, G. A., Thomas, J., Whitman, R. L. & Harper, D. 2003 Characterization and statistical modeling of bacterial (*Escherichia coli*) outflows from watersheds that discharge into southern Lake Michigan. (B. D. Melzian, V. Engle, M. McAlister, S. Sandhu & L. K. Eads, eds). In: *Coastal Monitoring through Partnerships*. Springer, Dordrecht, The Netherlands, pp. 289–300.
Passerat, J., Ouattara, N. K., Mouchel, J. M., Vincent, R. & Servais, P. 2011 Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River. *Water Research* **45** (2), 893–903.
Rada Ariza, A. & Torres, A. 2011 Preliminary assessment of the influence of Salitre basin rainfall on the pumping operation of Salitre wastewater treatment plant (Bogota). *Brasil. Evento: 12th International Conference on Urban Drainage*, Porto Alegre/Brazil, 11–16 September 2011.
Riechel, M., Matzinger, A., Sonnenberg, H., Caradot, N., Meier, I., Heinzmann, B. & Rouault, P. 2012 Validation and sensitivity of a coupled model tool for CSO impact assessment in Berlin, Germany. *6th International Congress on Environmental Modelling and Software iEMSs*, Leipzig.
Schilperoort, R. P. S. 2011 *Monitoring as a Tool for the Assessment of Wastewater Quality Dynamics*. PhD thesis, TU Delft, The Netherlands.
Tenenhaus, M. 1998 *La régression PLS, théorie et pratique*. Technip, Paris, France.
Veldkamp, R., Henckens, G., Langeveld, J. G. & Clemens, F. 2002 Field data on time and space scales of transport processes in sewer systems. *Proceedings of the 9th International Conference on Urban Drainage*, Portland, OR, USA, 8–13 September.

First received 28 June 2013; accepted in revised form 19 August 2013. Available online 25 October 2013