Classifying pollutant flush signals in stormwater using functional data analysis on TSS MV curves

Jensen, Ditte Marie Reinholdt; Sandoval, Santiago; Aubin, Jean-Baptiste; Bertrand-Krajewski, Jean-Luc; Xuyong, Li; Mikkelsen, Peter Steen; Vezzaro, Luca

Published in:
Water Research

Link to article, DOI:
10.1016/j.watres.2022.118394

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Water Research 217 (2022) 118394

Contents lists available at ScienceDirect

Water Research

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

Classifying pollutant flush signals in stormwater using functional data analysis on TSS MV curves

Ditte Marie Reinholdt Jensen [a,b,c,*], Santiago Sandoval [d,e], Jean-Baptiste Aubin [d], Jean-Luc Bertrand-Krajewski [d], Li Xuyong [b], Peter Steen Mikkelsen [a], Luca Vezzaro [a]

[a] Department of Environmental and Resource Engineering, Technical University of Denmark (DTU), Bygningstorvet, Bygning 115, 2800 Kongens Lyngby, Denmark
[b] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences (RCEES), Chinese Academy of Sciences (CAS), 18 Shuangqing Road, Beijing 100085, China
[c] Sino-Danish Center for Education and Research (SDC), Aarhus, Denmark and University of Chinese Academy of Sciences (UCAS), China
[d] University of Lyon, INSA Lyon, DEEP, EA 7429, F-69621 Villeurbanne cedex, France
[e] University of Applied Sciences and Arts of Western Switzerland (HES-SO), HEIA-Fr, ITEC, Boulevard de Pérolles 80, 1700 Fribourg, Switzerland

ARTICLE INFO

Keywords:
Pollutant flush
Separate sewer systems
MV curves
Clustering
Urban stormwater management

ABSTRACT

Pollution levels in stormwater vary significantly during rain events, with pollutant flushes carrying a major fraction of an event pollutant load in a short period. Understanding these flushes is thus essential for stormwater management. However, current studies mainly focus on describing the first flush or are limited by predetermined flush categories. This study provides a new perspective on the topic by applying data-driven approaches to categorise Mass Volume (MV) curves for TSS into distinct classes of flush tailored to specific monitoring location. Functional Data Analysis (FDA) was used to investigate the dynamics of MV curves in two large data sets, consisting of 343 measured events and 915 modelled events, respectively. Potential links between classes of MV curves and combinations of rain characteristics were explored through a priori clustering. This yielded correct class assignments for 23-63% of the events using different combinations of MV curve clustering and rainfall characteristics. This suggests that while global rainfall characteristics influence flush, they are not sufficient as sole explanatory variables of different flush phenomena, and additional explanatory variables are needed to assign MV curves into classes with a predictive power that is suitable for e.g. design of stormwater control measures. Our results highlight the great potential of the FDA methodology as a new approach for classifying, describing, and understanding pollutant flush signals in stormwater.

1. Introduction

Pollutants carried by urban runoff in separate sewer systems pose an environmental threat to the natural receiving water bodies (Brudler et al., 2019; Eriksson et al., 2011; Göbel et al., 2007; Koziel et al., 2019; Walsh et al., 2005; Zgheib et al., 2011) as well as to human health, if e.g runoff is used as source of potable water (Kus et al., 2010; Ma et al., 2019). To enable management actions aiming to reduce and mitigate this threat it is thus critical to gain knowledge of expected pollutant levels and of their spatial and temporal variations.

Pollutants in stormwater runoff stem from multiple release processes in the catchment: dry and wet atmospheric deposition, surface weathering, leaching from building materials, and release from anthropogenic sources, such as the use of specific chemicals (pesticides, spills, etc.) or vehicular activity (Deletic and Orr, 2005; Eriksson et al., 2005; Göbel et al., 2007; Müller et al., 2020). During rain events, these pollutants are washed off and transported away by rainfall and by the generated runoff. The resulting runoff quality thus depends on the factors driving the pollutant accumulation, the mobilisation and the transport mechanisms, resulting in large inter and intra-events variations in both concentrations and loads (Bertrand-Krajewski et al., 1998; Lee et al., 2002; Qin et al., 2016).

Historically, there has been a vast interest in describing pollutant flushes during a rain event, with great emphasis often being put on the *first flush*, i.e. pollutant flushes carrying out the majority of the event load during the first part of the event. The occurrence of first flush have

* Corresponding author.
E-mail addresses: dije@env.dtu.dk (D.M. Reinholdt Jensen), santiago.sandoval@hefr.ch (S. Sandoval), jean-baptiste.aubin@insa-lyon.fr (J.-B. Aubin), xyli@rcees.ac.cn (L. Xuyong), psmi@env.dtu.dk (P.S. Mikkelsen), luve@env.dtu.dk (L. Vezzaro).

https://doi.org/10.1016/j.watres.2022.118394
Received 24 November 2021; Received in revised form 16 March 2022; Accepted 2 April 2022
Available online 4 April 2022

been studied by a great number of studies, proposing different definition schemes, and developing different classification to describe this phenomenon (Jensen, 2022).

The increasing amount of available data on stormwater runoff quality has enabled the application of data-driven and statistical methods for exploring these datasets and gaining knowledge on pollutant flushes. For example, Deletic (1998) applied linear multi-regression analysis to search for correlations between the flush in the first 20% of the runoff and combinations of rainfall and runoff characteristics. Both Bach et al. (2010) and Sun et al. (2015) used the Wilcoxon Rank Sum test to group pieces of runoff events with similar concentration levels to determine the catchment-specific volume carrying the first flush. Li et al. (2012) applied redundancy analysis to quantify the influence of rain and catchment characteristics on the variation in the strength of the first flush. Li et al. (2015) used Pearson correlation analysis to connect Event Mean Concentrations (EMCs) and the strength of flush in the first 40% of the runoff to catchment and rain characteristics. Finally, Perera et al. (2019) applied a random forest, tree-based method in order to identify nonlinear relationships between rain and catchment characteristics and first flush.

However, a common denominator in studies of flush dynamics (including the data-driven studies mentioned above) is their sole focus on describing the first flush or their limitation to predetermined flush categories (Jensen, 2022). This exclusive focus might neglect other site-specific (and perhaps more relevant) tendencies and patterns, disregarding the span of variation in inter- and intra-event pollutant distributions that deviate from the general behaviour defined in literature. Indeed, there is a gap in the use of more exploratory approaches, with a broader scope than (and not limited to) the first flush. Such approaches can support stormwater planners in developing robust practices for managing pollutants in urban runoff by considering all possible flush patterns in a site and their impacts on the planned stormwater control measure.

Functional data analysis (FDA) is a strong tool for revealing dynamics of underlying processes in repeated observations (Wang et al., 2016), and it has the added benefit of not relying on approximations to a specific curve shape. As such, FDA has been employed for different purposes within the field of water management, e.g. for analysing spatial and temporal variabilities of rainfall data in order to identify basis functions of rainfall pattern at different stations or seasons (Suhaila and Yusop, 2017), for detecting anomalies in flow within water supply networks (Millán-Roures et al., 2018) or classifying streamflow hydrographs (Ternynck et al., 2016). These examples show how FDA was applied for classifying functional data, determining main trends or detecting outliers. To the best of our knowledge, this methodology has never been applied for analysing stormwater pollutant flush dynamics - possibly because it requires the availability of rather large data sets.

This study aims to provide a better understanding of pollutant flushes in separate sewer systems by using unsupervised analysis tools based on FDA to (i) explore and categorise distinct Mass Volume (MV) curve shapes and (ii) describe their frequency of occurrence. Furthermore, we (iii) assess the potential for linking the resulting flush classes to combinations of rain variables. The study is undertaken using large data sets of events from both monitored and modelled case studies.

2. Materials and methods

2.1. Theory of MV curves and flush drivers

A widely applied approach to describe pollutant flushes during an event is based on MV curves, which have been extensively investigated in scientific literature (e.g. Bertrand-Krajewski et al., 1998; Chow and Yusop, 2014; Deletic and Maksimovic, 1998; Kong et al., 2021; Métadier and Bertrand-Krajewski, 2012; Perera et al., 2021). MV curves present the relative cumulative mass (M) as a function of the relative cumulative volume (V), enabling the analysis of flush signals (see Fig. 1). The first

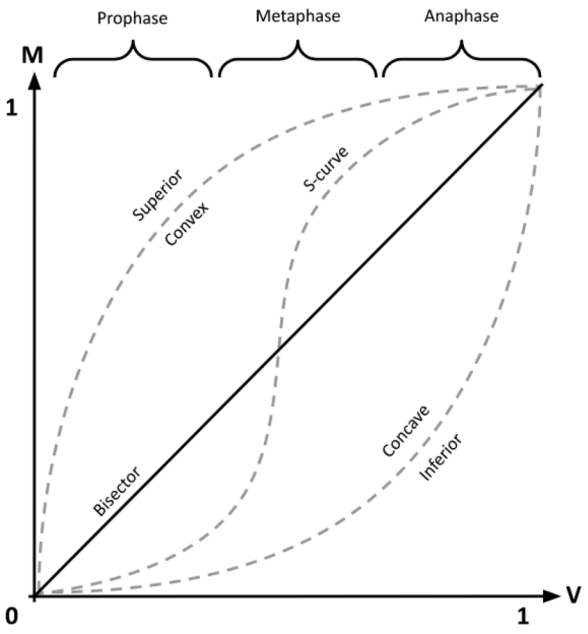


Fig. 1. Theoretical representation of MV curves and introduction of terminology. Inspired by Qin et al. (2016) and Métadier and Bertrand-Krajewski (2012).

flush takes place if the majority of the mass is carried in the prophase of the runoff volume, and its strength is described by the convexity of the (superior) MV curve. No flush occurs if the MV curve follows the bisector, while a second flush (or middle flush) occurs when the MV curve follows an s-shape, with the majority of the mass discharged in the metaphase. If the MV curve is concave (inferior), it indicates that the majority of the mass is discharged in the anaphase, resulting in third flush (or late flush). The MV curve approach thus enables the comparison of flush signals across events with different mass, volume, and duration. However, several flushes may occur during a single event, resulting in a large variety of possible MV curve shapes. MV curves are often fit to polynomial approximations in order to conduct data analysis (e.g. Métadier and Bertrand-Krajewski, 2012). This may be a good approximation for superior and inferior curves, but it is a poor estimation for s-curves or reverse s-curves.

While the event normalisation allows a comparison of dynamics across events that may otherwise be hard to compare, other relevant event characteristics (magnitude, duration, return period, etc.) are lost. Therefore, if relevant for e.g. design purpose, events should be preselected based on these characteristics before the MV curve normalisation and analysis.

Although several studies investigating the drivers of flush signals are based on MV curves, it is often hard to compare their results due to the use of different flush definitions, focus on different pollutants, and different local conditions. Indeed, Deletic (1998) noticed that the influential rainfall and runoff characteristics might differ for catchments of very similar characteristics. Perera et al. (2019) found that rain-related variables were ranked highest in terms of influence on first flush and suggested that their combination might yield better results. While the occurrence of first flush has been found to be affected by the type of pollutant (particulate or dissolved), the system (combined or separate), the catchment characteristics (land use, slope, size, etc.), the sampling technique, and the flush definition (Jensen, 2022), rain characteristics are among the most typically explored drivers - in part as they are more frequently accessible. One explanation could be that rain characteristics are assumed to be the major driver of inter-event variability, while other drivers are often assumed to be constant or to have little influence at the temporal scale.

A great disagreement is found in literature on the influence of the Antecedent Dry Weather Period (ADWP), also called Antecedent Dry

Weather Days (ADWD) or Antecedent Dry Days (ADD) (e.g. Chow and Yusop, 2014; Deletic and Maksimovic, 1998; Li et al., 2012; Schriewer et al., 2008). This is notable, as this variable constitutes the backbone of build-up modules in several accumulation-washoff models.

Several studies stressed the influence of rain intensity (e.g. Chow and Yusop, 2014; Deletic and Maksimovic, 1998; Li et al., 2012). However, Athanasiadis et al. (2010) found that intensity did not affect first flush effects of Cu in roof runoff, and Schriewer et al. (2008) found that Zn in roof runoff is correlated with lower rain intensities. Furthermore, Kong et al. (2021) found that first flush effect depended on the timing on the maximum intensity, and similarly, mixed interpretations can be found for the influence of total rain depth of an event and its duration (Athanasiadis et al., 2010; Chow and Yusop, 2014; Kang et al., 2008; Li et al., 2015; Perera et al., 2019).

As studies on influential flush factors often point in different directions, it is likely that the governing explanatory variables are case-specific. This stresses the need for approaches capable of evaluating flush signals systematically, rather than general definitions that may not apply under local conditions.

2.2. Case studies

In order to capture several different flush signals it is necessary to utilise an extensive data set from a specific location. Therefore, the analysis was carried out on a large data set of monitored Total Suspended Solids (TSS) data from Chassieu (France). As large data sets of TSS data are seldom available (with Chassieu being one of the few examples worldwide), these were supplemented by a set of modelled data from Albertslund (Denmark), obtained using a state-of-the-art accumulation-washoff model. While inherently uncertain due to the model simplification of reality, the modelled data have the additional benefit of being less affected by measurements noise and they are thereby assumed easier to analyse. For both data sets TSS concentration time series are used as indicator for pollutant levels.

2.2.1. Chassieu data set (monitored)

The Chassieu catchment covers an industrial area in the outskirts of Lyon (France) drained by a separate sewer system. Continuous monitoring of rain (6-minute time steps), flow and turbidity (2-minute time steps) was carried out from 2004 to 2011. The data set is described in Métadier and Bertrand-Krajewski (2012) and Sun et al. (2015), while the conversion from turbidity to TSS is reported in Métadier and Bertrand-Krajewski (2011). The data set includes 716 events, but not all of these contain complete measurement series (see Table 1). For further details, we refer to the previous works conducted on this data set (e.g. Métadier and Bertrand-Krajewski, 2011; 2012; Sandoval et al., 2018; Sun et al., 2015).

2.2.2. Albertslund data set (modelled)

The Albertslund catchment covers a residential/industrial area in the outskirts of Copenhagen (Denmark) drained by a separate sewer system. Continuous monitoring of rain, flow and turbidity was carried out in 2010-2011. These measurements were used to calibrate a dynamic stormwater pollution model, based on the state-of-the-art accumulation-washoff process (Vezzaro et al., 2015). The model parameters were estimated by using the pseudo-bayesian, uncertainty-based methodology described in Vezzaro et al. (2013). To propagate model uncertainties, a total of 1,000 parameter sets were sampled from the behavioural parameters identified in (Vezzaro et al., 2015). These were used to simulate a 10-year period (using rainfall data covering the period 1994-2004), resulting in a total of 1,094 discharge events. The characteristics of each event (volume, load, etc.) were calculated on the median of the 1,000 simulations (see Table 1).

Table 1
Main characteristics of the two analysed case studies. For more details please refer to Métadier and Bertrand-Krajewski (2012) and Vezzaro et al. (2015).

	A_{imp}	Events*	Rain $depth_{avr}$	Rain dur_{avr}	Rain int_{avr}	TSS SMC**
Chassieu	133 ha	343	8.49 mm	6.45 h	1.77 mm/h	119 g/m^3
Albertslund	94.7 ha	915	5.33 mm	5.37 h	1.26 mm/h	63 g/m^3

*The presented values are after step 1a in the methodology has been completed.
**SMC = Site Mean Concentration.

2.3. Data analysis

This study utilises a data-driven and exploratory approach based on techniques from Functional Data Analysis (FDA), which allow for the identification of multiple categories and types of flush signals without relying on predetermined definition schemes. An overview of the performed steps for both case studies is given in Fig. 2 and further elaborated in the following subsections:

1. Data preparation: a validation of the events was carried out, selecting those to be analysed, preparing MV curves, and extracting rain characteristics for the further steps.
2. MV curve analysis: this step explored the dynamics of MV curves and divided them into classes based on their shape.
3. Rain variables analysis: non-essential variables were eliminated before the combined analysis.
4. Combined analysis: this step explored the direct relationship between MV classes and rain variables and indirect relationships between MV classes and combinations of rain variables.

2.3.1. Data preparation

Events in the monitored data set (Chassieu) that contained missing values of runoff flow or TSS concentration for more than 10 continuous minutes were discarded. Missing values for less than 10 continuous minutes were linearly interpolated. The modelled data set (Albertslund) included small rainfall events that did not generate runoff. Therefore, events with a rain duration of shorter than 10 minutes were eliminated, as generally these events had very low rain depth.

For each event in the data set, MV curves were generated using TSS and flow data. To ensure that all MV curves had equal discretisation (a prerequisite for the used FDA algorithms), they were rescaled using interpolation to ensure a discretisation of 101 steps per curve (corresponding to a value of 0%, 1%,..., 100% of the duration of the runoff event).

An overview of the investigated rain characteristics is provided in Table 2, which includes both variables found to be significant by previous studies and additional new variables. For the Chassieu data set,

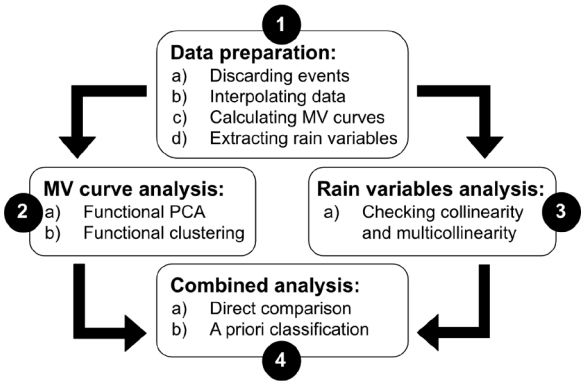


Fig. 2. Overview of applied methodologies and flow of approach.

Table 2
Overview of the investigated rain characteristic variables and their abbreviations. Variables marked in bold are those that passed through step 3 of the methodology (based on the qualitative assessment of plots found in Supplementary Material).

Acronym	Description	Unit
DUR	Total duration of rainfall	min
AVR_INT	Average rainfall intensity	μm/s
MAX_INT	The maximum 6-minute intensity in the rainfall event	μm/s
MAX_STEP	The number of the time step where the MAX_INT begins	number
MAX_STEP_REL	MAX_STEP divided by total number of time steps in the rainfall event	fraction
MAX_INT_50FIRST	Maximum intensity (6-minute steps) found in the first 50% of the total rain depth	μm/s
MAX_INT_50LAST	Maximum intensity (6-minute steps) found in the last 50% of the total rain depth	μm/s
MAX_INT_50REL	MAX_INT_50FIRST divided by MAX_INT_50LAST	fraction
DEPTH	Total depth of the rainfall	mm
NO_RAIN	The amount of minutes in the event without any rain divided by the total duration	fraction
ADWP	Antecedent dry weather period	hours
AR2	Antecedent rainfall 2 days from the beginning of the event	mm
AR5	Antecedent rainfall 5 days from the beginning of the event	mm
AR7	Antecedent rainfall 7 days from the beginning of the event	mm
AR14	Antecedent rainfall 14 days from the beginning of the event	mm
Max_int_last	Maximum 6-minute intensity in the last, preceding rain event	μm/s
Month	The month the event started in (for seasonality)	no.

rain data was only available for the events in the database (i.e. there were no data for the periods between the recorded events), so the ADWP was estimated based on the 716 available events. Also, for both the Chassieu and the Albertslund data sets, there was no information about rain preceding the first event in the database, so ADWP and values of antecedent rain in the last 2-14 days (AR2-14) were set to zero.

2.3.2. MV curve analysis

A functional principal components analysis (fPCA) was run to explore the shape and dynamics of the MV curves, followed by functional clustering to divide the MV curves into classes. fPCA is the most common first step of any FDA, as it reduces the dimensionality of the longitudinal data and creates a foundation for many other FDA methods (Chen et al., 2017), in addition to offering insight in itself. It is based on an eigenvalue analysis of the underlying covariance operators of the observed process. fPCA was used to generate eigenfunctions for the functional data until a high percentage of the variance (99.9%) was described. Each eigenfunction has a corresponding list of eigenvalues with length matching the number of events in the data set. This resembles the Principal Component (PC) scores for that particular PC index. Both the fPCA and the functional clustering were carried out using the 'fdapace' package (Carroll et al., 2020), which is an open-source R-package for FDA and empirical dynamics (an equivalent 'PACE' package is also available for Matlab: http://www.stat.ucdavis.edu/PACE/).

The functional clustering was carried out to identify natural MV profile classes in an unsupervised manner through strict partitioning clustering, meaning that each object belongs to exactly one cluster. The fdapace package offers two options for functional clustering: the Expectation Maximization Cluster (EMCluster) algorithm and the k-Centers Functional Clustering (kCFC) algorithm. EMCluster uses the fPC scores previously calculated through fdapace (Carroll et al., 2020) and creates multivariate Gaussian distributions for each cluster (Chen and Maitra, 2015). kCFC looks at both the mean and modes of the variation differentials between clusters by predicting the cluster membership through an iterative covariance updating scheme. Here, local linear smoothing is applied, covariance for each curve is calculated separately and aggregated to estimate the smoothed covariance, followed by eigenanalysis and calculation of fPC scores using conditional expectation (Chiou and Li, 2007). Both algorithms were tested to find the best way to divide the data set into a natural number of clusters. A comparison was made through visual inspection of the resulting clusters and by looking at the Sum of Squared Errors (SSE), i.e. the summed and squared mean distance from each fPC score (across all dimensions needed to reach 99.9% explained variance) to the centre of its associated cluster.

2.3.3. Rain variables analysis

In order to prepare the rain variables for the PCA and cluster analysis, they were first checked for collinearity, i.e. whether any of the variables were linear predictors of each other. While PCA can in itself remove collinearity, PC scores will be affected by the redundant features included in the analysis through collinear variables. Clustering collinear variables can indeed place additional weight on a feature because it is represented in multiple variables. Although removing collinear variables is usually considered not to affect k-means clustering, it has the extra benefit of speeding up calculations and achieving a more parsimonious model.

The collinearity was evaluated by looking at correlation plots of all the rain variables. Multivariate collinearity was evaluated by running a linear regression (with the fPC scores as response variables and rain variables as explanatory variables), and evaluating the Variance Inflation Factor (VIF) of the resulting regression.

2.3.4. Combined analysis

Firstly, a direct comparison of the MV curves and rain variables was conducted by looking at box plots of rain variable values for each cluster from the functional clustering analysis. Furthermore, a Kruskal-Wallis test was carried out to ascertain if significant difference had been obtained between the clusters for each rain variable (Kruskal and Wallis, 1952). It is a nonparametric, global test across all classes at once, consistent with the data-driven approach to cluster-analysis followed in this paper. The test uses a null hypothesis that the median (for each rain variable) is comparable across groupings (MV classes), which is rejected if any of the classes stand out with a significantly different value. If the test shows that some of the MV classes provide values of rain characteristics significantly different from the rest, this indicates a potential for using rain variables as explanatory factors of the MV classes.

Secondly, a clustering scheme was set up to combine the rain variables, investigating if their joint variation can be linked to the MV categories. The analysis first looked at using a single rain variable, and then an increasing number of variables were combined. Before clustering, the variables were standardised and ran through a PCA to reduce dimensionality and increase the efficiency of the analysis. Clustering was carried out using a standard k-means approach. The MV categories were used as an a priori classification for the clustering of the rain variables, and a contingency table (as shown in Table 3) was produced for each combination of rain variables. To evaluate the performance, we used the two indicators applied by Chiou and Li (2007) for comparison of two partitionings (cRate and aRand):

Table 3
Contingency table for comparing partition X of R clusters with partition Y of C clusters. Modified from Hubert and Arabie (1985).

	Y_1	Y_2	⋯	Y_C	Sums
X_1	n_{11}	n_{12}	⋯	n_{1C}	a_1
X_2	n_{21}	n_{22}	⋯	n_{2C}	a_2
⋮	⋮	⋮	⋮	⋮	⋮
X_R	n_{R1}	n_{R2}	⋮	n_{RC}	a_R
Sums	b_1	b_2	⋮	b_C	n

- **cRate**: the maximal possible ratio of correctly classified objects (bounded by 0 and 1), calculated as the sum of the diagonal in the contingency matrix (see Table 3) divided by the total number of events, as the diagonal represents events placed in the same cluster by both partitions.
- **aRand**: a corrected version of the Rand index (Rand, 1971) proposed by Hubert and Arabie (1985), which describes the agreement between two partitions of objects (events in this study) by looking at the number of paired objects that are in the same or different groups in both partitions. It is calculated as in Eq. 1 (with nomenclature referring to the contingency table in Table 3). Here, a comparison is made between clusters of partitions X and Y, with a_i as the total amount of events placed in the ith cluster in X, b_j as the total amount of events placed in the jth cluster in Y, n_{ij} as the amount of events from cluster i in the X partition placed in cluster j in the Y partition, and n as the total number of events. aRand has an upper bound of 1 and an expected value of 0, although negative values may occur if the degree of similarity between the two partitions is less than what would be expected in the case of two random partitionings.

$$\text{aRand} = \frac{\text{Index} - \text{Expected index}}{\text{Maximum index} - \text{Expected index}} = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}} \quad (1)$$

3. Results and discussion

3.1. Data preparation

The event discarding step led to a reduction in the number of events in each data set: from 716 to 343 events for the monitored data set (Chassieu) and from 1094 to 915 events for the modelled data (Albertslund). A summary of the validated events is given in Table 1.

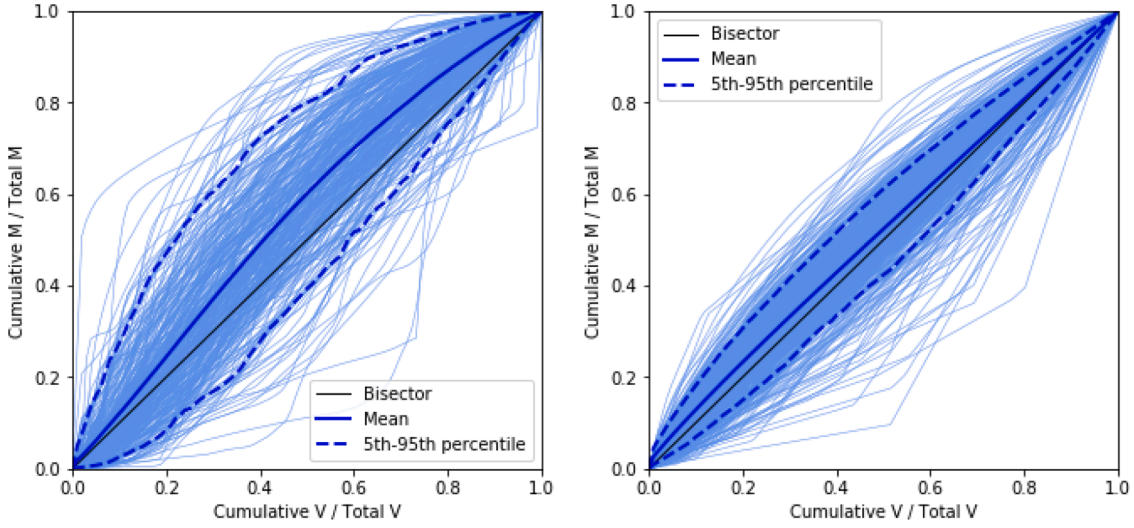
The MV curves for the Chassieu and Albertslund sites are shown in Fig. 3. Both plots show a wide span of MV curves on both sides of the bisector rather than any single, general trend in curve profile. Interestingly, the accumulation-washoff model does not produce evident first flush signals (see Fig. 3b), suggesting that differences in input rainfall dynamics between events are a major driver behind inter- and intra-event variation in MV curves.

As expected, the MV curves show a greater variability in the measured Chassieu data set compared to the modelled Albertslund data. The simplification and structural limitations of the model structure (accumulation-washoff) clearly lumped the variety of processes taking place in the natural system. Although accumulation-washoff processes are often used to support the theory behind first flush hypothesis, the simulation results display that the dynamic behaviour of rainfall plays a major role in the pollutant mobilisation in the catchment and transport in the drainage system, and thereby in the presence of flush signals. The mean of the MV curves shows a slight tendency of a middle flush in Chassieu, while the mean in Albertslund is closer to the bisector.

The extraction of rain characteristics from the data set resulted in seventeen different rain variables (see Table 2).

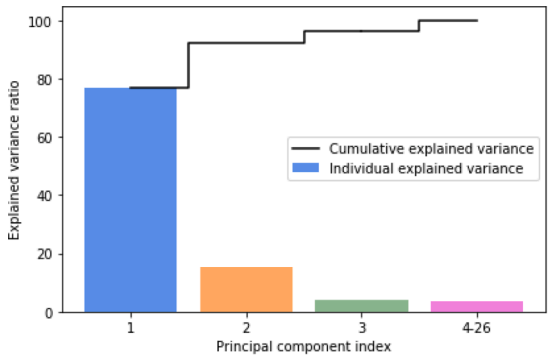
3.2. MV curve analysis

More than 96% of the Chassieu variability and more than 99% of the Albertslund variability is summarised in the first 3 principal components of the fPCA (see Fig. 4a and 4 b). The large amount of information included in such few components shows a high level of structure in the data set (i.e. the curves share many similarities). The first three eigenfunctions for the two case studies are shown in Fig. 4c-4 e. Each eigenfunction represents a component of the basis on which MV curves are rebuilt, while the eigenvalue gives information on the importance of the associated eigenfunction in this reconstruction. Interestingly, an additional inflection point appears for each new eigenfunction for both data sets, and this trend continues beyond the first three eigenfunctions

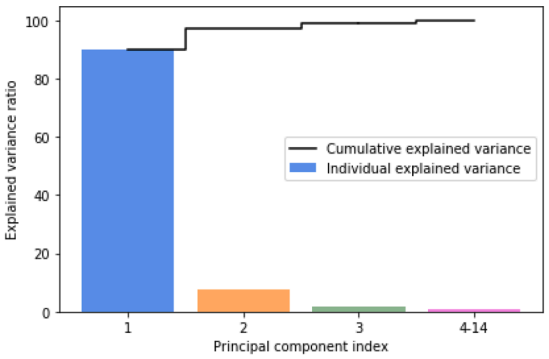


(a) Chassieu (b) Albertslund

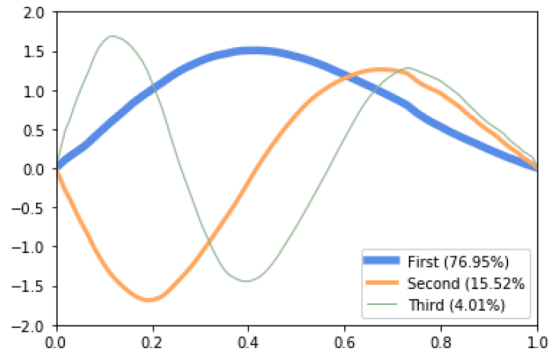
Fig. 3. Dimensionless MV curves for the two studied catchments: Chassieu (monitored) and Albertslund (modelled).



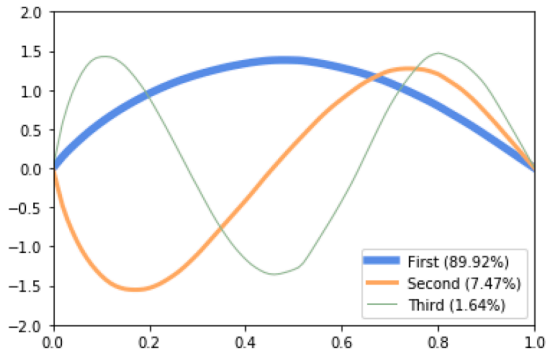
(a) Scree plot for the Chassieu data set.



(b) Scree plot for the Albertslund data set.



(c) Plot of eigenfunctions for the Chassieu data set.



(d) Plot of eigenfunctions for the Albertslund data set.

Fig. 4. Scree plots (top) and plots of eigenfunctions (bottom) for Chassieu (left) and Albertslund (right). The fraction of explained variance is shown in parenthesis.

shown in Fig. 4c. This shows that the majority of the MV curves (represented by the first eigenfunction) have only one inflection point, followed by curves with two inflection points, etc. Using approximations with a single inflection point (such as the polynomial approximation x^b, e.g. used by Métadier and Bertrand-Krajewski (2012)), would thus only describe the fraction of the MV curves corresponding to the variance explained by first eigenfunction (the first bar of the scree plots), i.e. 76.95% for Chassieu and 89.92% for Albertslund.

The eigenfunction plots (Fig. 4c-4d) show how the considered functions (the MV curves) differ from the mean. Similarly to the analysed MV curves, all the eigenfunctions have an initial value of 0 and an end value of 1. This demonstrates how the eigenfunctions can be used to highlight similarities across the data set, although this particular similarity is apparent without the aid of eigenfunctions. The plots show that the functions for Chassieu are less symmetrical than for the Albertslund case, with the first eigenfunction skewing left and the first modes of the following eigenfunctions having larger amplitudes than the following modes. This shift comes from the displacement of the mean from the bisector, as seen in Fig. 3a. There seems to be a larger mass fraction in the metaphase and a smaller in the anaphase, but the mean does not show an s-shape form crossing the bisector. If the mean MV curve moves from a convex to a s-shape, it will result in a displacement of the eigenfunctions, culminating with the introduction of a second inflection point already in the first eigenfunction. In terms of relevance to stormwater planners, the eigenfunctions and the scree plot provide an estimate of i) the fraction of events with none or one consistent flush effect (the first bar and first eigenfunction), and ii) the fraction of events that exhibit multiple flush effects and cross the bisector (the rest). The first information is relevant to time the capture of pollutant flushes by Stormwater Control Measures (SCMs), and the latter provides a quantification of the events with multiple flushes, indicating the fraction of events where traditional pollution management approaches, focusing solely on first flush (Jensen, 2022), are likely to fail or to provide a lower efficacy than originally planned.

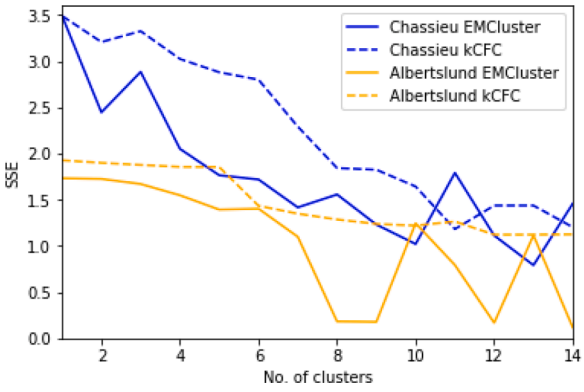


Fig. 5. Plot of Summed Squared Error (SSE) as a function of number of clusters using EMCluster or kCFC algorithm for both case studies.

Regarding the functional clustering, Fig. 5 shows the development in SSE as a function of the number of clusters for both case studies using both EMCluster and kCFC. The SSE was calculated using a number of fPCA components explaining more than 99.99% of the variance (26 iterations for Chassieu and 14 for Albertslund). The SSE was calculated on the fPC scores from the fPCA. Since these scores are used directly for clustering by the EMCluster algorithm, it is no surprise that the EMCluster approach performs better (lower SSE) than the kCFC algorithm. The latter method, in fact, calculates its own fPC scores as part of the clustering process. Both methods obtained a lower SSE in Albertslund compared to Chassieu, explained by the smaller variability shown by the modelled MV curves.

Both the EMCluster and the kCFC algorithms recognised principal curve profiles. However, the EMCluster seemed to create clusters with less overlap, and it also ran considerably faster. The analysis of the MV curves classified according to cluster memberships reveals that the local

peaks in the SSE-plot arise when a cluster includes events from both sides of the bisector (see supplementary material for plots of the classified MV curves). This is sufficient if one is only interested in the degree of flush and not in its type. By applying the elbow criterion to the SSE-plot, the optimal number of cluster was 10 for the Chassieu data and 8 for Albertslund. However, the use of such a great number of classes of MV curves might not be recommended for practical applications, i.e. in a planning situation where decision makers need to assess a limited number of events or response options, and it might be difficult to maintain an overview or differentiating responses if too many classes are used.

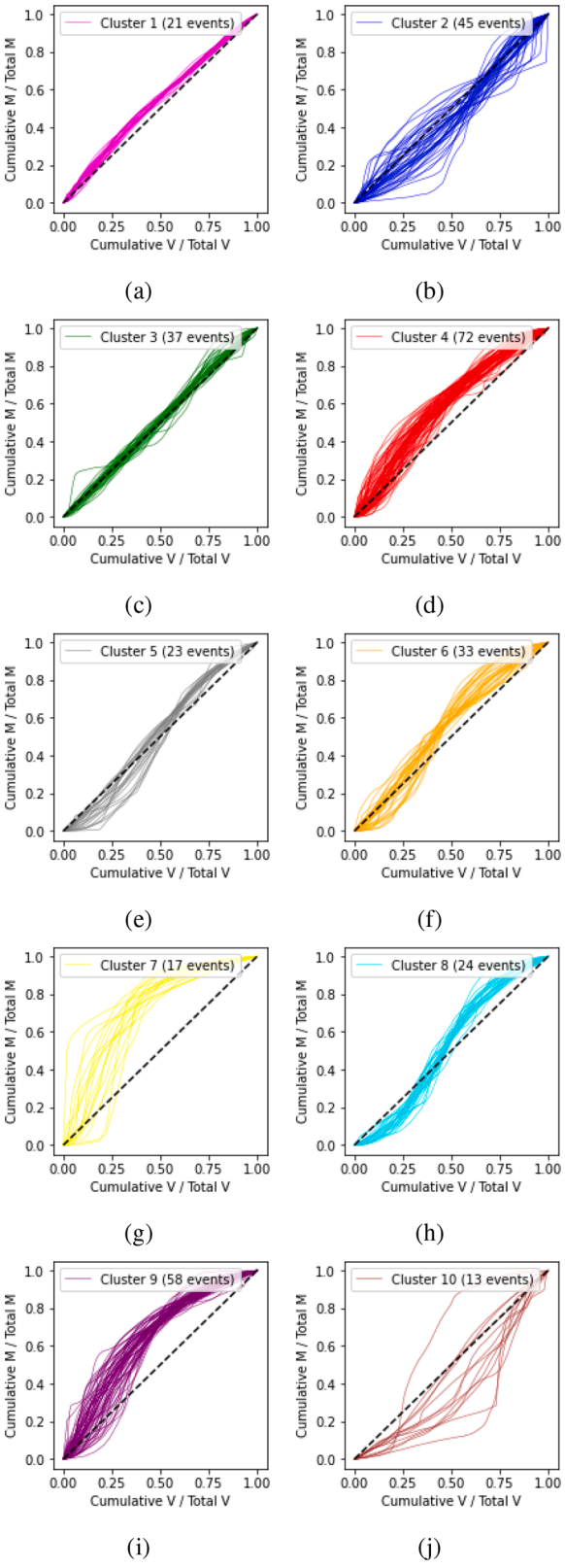


Fig. 6. Chassieu MV curves divided into 10 clusters using EMCluster. The dashed line represents the bisector.

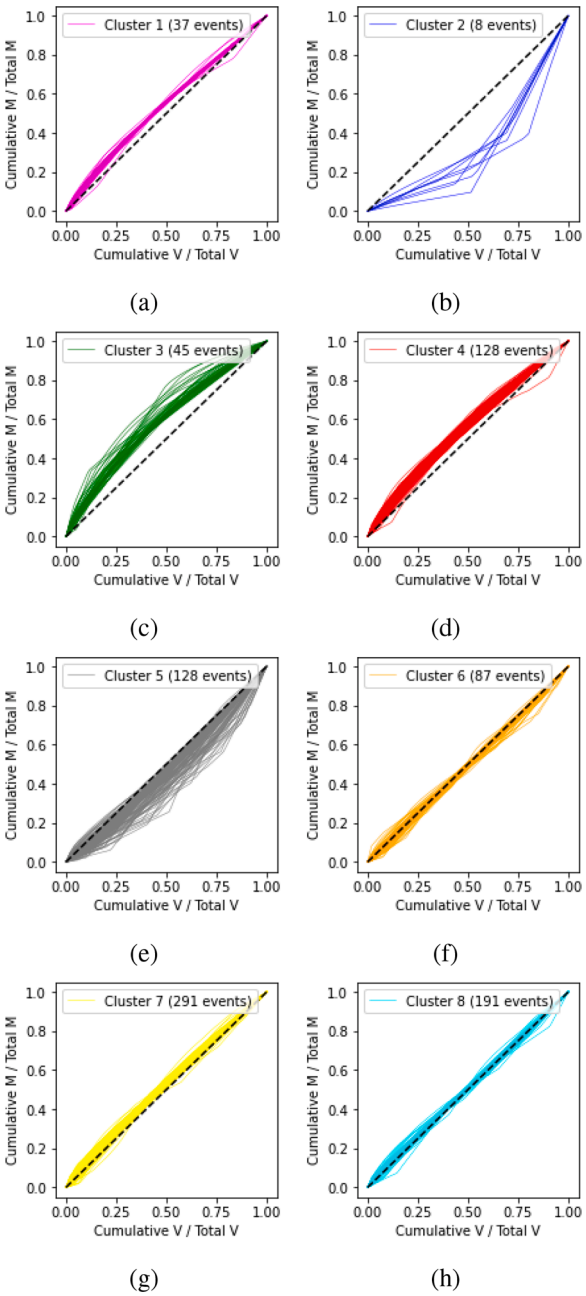


Fig. 7. Albertslund MV curves divided into 8 clusters using EMCluster. The dashed line represents the bisector.

Therefore, it was chosen to perform the analysis both with the optimal number of clusters (10 for Chassieu and 8 for Albertslund, respectively) and with a reduced number of clusters (4 for Chassieu and 3 for Albertslund, respectively), assumed to be more manageable by decision makers in a planning situation. These clusters represent an advancement compared to the use of predefined flush categories, which might not reflect the characteristics of the catchment under study. Figs. 6–9 display the MV curves grouped for the different number of clusters. When a smaller number of clusters was used, a great number of

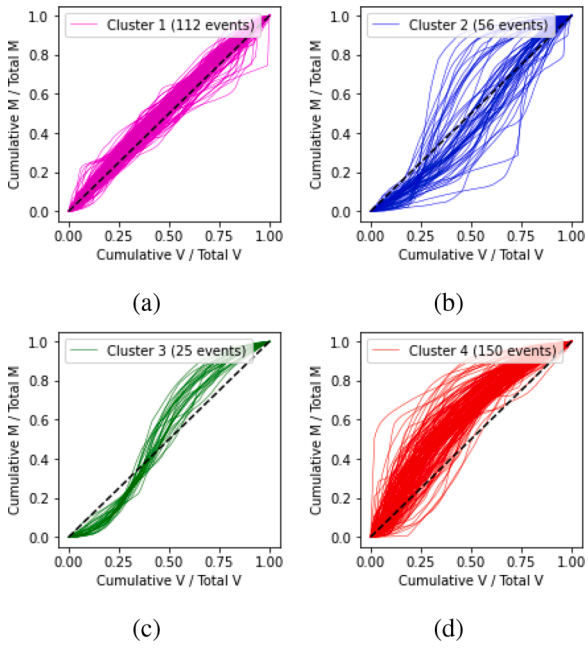


Fig. 8. Chassieu MV curves divided into 4 clusters using EMCluster. The dashed line represents the bisector.

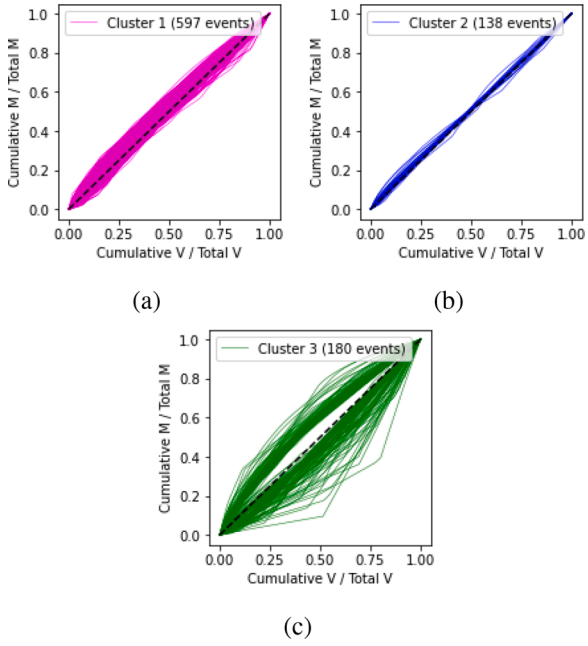


Fig. 9. Albertslund MV curves divided into 3 clusters using EMCluster. The dashed line represents the bisector.

classes tend to cross the bisector, showing how these clusters are suited for describing the magnitude of flush but not its type. A greater number of clusters resulted in a clearer, more distinct classification of MV curves, where patters in terms of flush type (s-shape, inferior or superior) and the magnitude of flush (amplitude of concavity or convexity) can be identified.

Overall, the fPCA and the functional clustering approach provided useful insights into the data sets. The widespread binary tendency to sort pollutant patterns into either fist flush or not first flush means that we lose the opportunity for tailoring the appropriate management response to the actual degree of flush. As the FDA approach is not limited by such predetermined classes, it is a much more feasible approach to categorise flushes. Also, it analyses the MV curves directly - i.e. without making polynomial approximations to extract curve shape features - and it is thus better suited for representing the actual flush patterns. Furthermore, the fPCA methodology proved to contribute insights into the governing MV curve shapes, e.g. in terms of number of inflection points, which is important to evaluate the implications of multiple flushes. The clustering can be coupled in the future with the identification of generalised curves for different flush patterns, which can be used in planning situation as "flush scenarios", limited to the number of cluster that are deemed as manageable by decision makers (e.g. 4 clusters for the Chassieu case).

Few studies have addressed the uncertainty related to pollutant flushes. Some examples can be seen in the work by Sharifi et al. (2011) and Perera et al. (2021), but these both used an exponential approximation and focused solely on the first flush. Considering the benefits of the FDA methodology applied in this study, it would be interesting to also apply FDA to describe flush uncertainties.

3.3. Rain variables analysis

The qualitative analysis of inter-correlation plots between the variables led to the elimination of six variables showing a high degree of collinearity with other variables (see supplementary material). The 11 remaining variables are highlighted in bold Table 2.

For the estimation of multicollinearity between the variables, all VIF values for both the Chassieu and the Albertslund data set proved to be less than 5 (see supplementary material), which is not considered problematic (Craney and Surles, 2002), and did not give rise to further variable elimination. Therefore, multicollinearity should not affect the results when combining several rain variables to match the MV clusters (methodology step 4).

3.4. Combined analysis

Regarding the direct comparison between rain variables and MV classes, box plots for Chassieu (4 and 10 clusters) and for Albertslund (3 and 8 clusters) are enclosed in the supplementary material. Table 4 summarises the p-values from the Kruskal-Wallis test for each rain variable. As the 44 tests were carried out simultaneously, the Bonferroni correction was applied to ensure adequate significance before rejecting the null hypothesis, which is why P-values < 0.001 ($p/n = 0.05/44$) are highlighted in the Table.

Regarding the a priori classification, Table 5-6 shows the best performing combination (measured by cRate) for each number of variables included in the clustering. It should be noted that the highest aRand values were not always achieved by the same combination of variables that yielded the highest cRate.

Table 4
P-values from the Kruskal-Wallis test. Values that discard the null hypothesis that the median for each variable is comparable across the MV classes (< 0.001) are marked in bold.

Rain characteristic	Chassieu		Albertslund	
	4 clusters	10 clusters	3 clusters	8 clusters
DUR	0.58	0.01	**1.16E-49**	**2.09E-08**
MAX_INT	0.01	0.06	**4.47E-17**	0.54
MAX_STEP_REL	0.52	0.40	**2.41E-4**	0.01
MAX_INT_50REL	**1.12E-04**	0.01	**6.97E-17**	0.12
DEPTH	**5.92E-06**	**6.86E-04**	**4.00E-37**	**1.38E-04**
NO_RAIN	**3.77E-04**	0.01	**3.18E-16**	**6.94E-04**
ADWP	**6.84E-06**	3.53E-03	0.33	0.04
AR2	1.36E-03	0.08	0.59	0.12
AR7	0.01	0.25	0.56	0.21
Max_int_last	0.32	0.32	0.59	0.35
Month	0.11	0.54	0.21	0.64

Table 5
Performance of the best performing combined rain variable clustering (ranked according to cRate) for each number of variables included in the Chassieu case.

		Variable names	cRate	aRand
Chassieu 4 clusters	1	AR2	0.45	0.02
	2	MAX_STEP_REL and AR2	0.45	0.02
	3	MAX_STEP_REL, MAX_INT_50REL and AR2	0.45	0.02
	4	MAX_STEP_REL, MAX_INT_50REL, NO_RAIN and AR2	0.45	0.02
	5	MAX_STEP_REL, MAX_INT_50REL, NO_RAIN, Month and AR2	0.45	0.02
	6	MAX_STEP_REL, MAX_INT_50REL, NO_RAIN, Month, Max_int_last and AR2	0.42	0.01
	7	MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Month, AR2 and AR7	0.40	0.00
	8	MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Month, Max_int_last, AR2 and AR7	0.40	0.02
	9	MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.39	-0.01
	10	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.39	-0.02
	11	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.35	-0.01
Chassieu 10 clusters	1	AR2	0.23	0.02
	2	MAX_INT_50REL and AR2	0.23	0.02
	3	MAX_INT_50REL, NO_RAIN and AR2	0.23	0.02
	4	MAX_INT, MAX_STEP_REL, NO_RAIN and AR2	0.22	-0.00
	5	MAX_STEP_REL, MAX_INT_50REL, DEPTH, Month and AR2	0.22	0.00
	6	MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Month, AR2	0.22	0.00
	7	MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Month, Max_int_last and AR2	0.21	-0.01
	8	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Month, Max_int_last and AR2	0.20	-0.01
	9	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month and Max_int_last	0.19	-0.01
	10	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last and AR2	0.19	-0.01
	11	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.18	-0.01

Table 6
Performance of the best performing combined rain variable clustering (ranked according to cRate) for each number of variables included in the Albertslund case.

		Variable names	cRate	aRand
Albertslund 3 clusters	1	DEPTH	0.63	0.04
	2	MAX_STEP_REL and DEPTH	0.63	0.04
	3	MAX_INT, DEPTH and Max_int_last	0.63	0.04
	4	MAX_INT, MAX_STEP_REL, DEPTH and Max_int_last	0.63	0.04
	5	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH and Max_int_last	0.63	0.04
	6	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN and Max_int_last	0.63	0.04
	7	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN and ADWP	0.60	0.10
	8	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP and Month	0.60	0.10
	9	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month and Max_int_last	0.60	0.10
	10	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last and AR2	0.60	0.10
	11	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.60	0.10
Albertslund 8 clusters	1	DEPTH	0.30	0.10
	2	MAX_INT and DEPTH	0.31	0.09
	3	DEPTH, NO_RAIN and Max_int_last	0.31	0.09
	4	DEPTH, NO_RAIN, Month and AR2	0.32	0.05
	5	MAX_STEP_REL, DEPTH, NO_RAIN, Month and AR2	0.32	0.05
	6	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN and Max_int_last	0.31	0.08
	7	MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, Max_int_last and AR2	0.30	0.06
	8	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP and Month	0.29	0.13
	9	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month and Max_int_last	0.29	0.13
	10	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last and AR2	0.29	0.13
	11	DUR, MAX_INT, MAX_STEP_REL, MAX_INT_50REL, DEPTH, NO_RAIN, ADWP, Month, Max_int_last, AR2 and AR7	0.29	0.13

3.4.1. Chassieu

The box plots of rain variable distributions across the MV classes (see supplementary material) showed a visible difference for the NO_RAIN variable in both the 4 and the 10 MV cluster scenario. In the latter scenario, cluster 8 stands out: the MV curves from this cluster (Figure) are in fact mostly s-shaped curves. It is worth highlighting this result, as this group of MV curves cannot be approximated using the traditional X^b formula, and it seems to have a clear predictor variable. The Kruskal-Wallis test (summarised in Table 4) also had low p-values for the NO_RAIN variable (only for 4 clusters), as well as for several of the other rain variables. This indicates that the applied MV classes manage to divide the rain variables into groups so that at least one group is significantly different from the rest.

Considering the a priori classification (Table 5), then the maximum rate of correct FDA class assignment of MV curves based on rain characteristics was 45% for the 4 MV classes scenario and 23% for the 10 MV classes scenario. It is expected that it will be harder to achieve a high rate of correct assignment when more MV classes are involved. In both cases, there was no increase in performance when including more rain variables in combination. The calculations that relied on only a single variable resulted in the following ranking of the rain (ordered by cRate):

- 4 MV classes: AR2 (0.45), Max_int_last (0.41), DEPTH (0.40), ADWP (0.39), AR7 (0.38), MAX_STEP_REL (0.37), DUR (0.35), MAX_INT (0.35), MAX_INT_50REL (0.33), Month (0.31) and NO_RAIN (0.30)
- 10 MV classes: AR2 (0.23), MAX_INT (0.20), AR7 (0.20), Max_int_last (0.19), MAX_STEP_REL (0.18), ADWP (0.18), DEPTH (0.18), NO_RAIN (0.17), DUR (0.17), MAX_INT_50REL (0.17) and Month (0.17)

All the best performing combinations from both MV class scenarios, except one, included the variable AR2, indicating that the very recent antecedent rainfall is influential to the MV curve shape. This supports

that the MV curve shape is susceptible to the applied event definition, as different event definitions might have included or excluded more of the rain close to the event. The aRand values generally indicated that the model performance was close to that of random assignment (~0).

3.4.2. Albertslund

The box plots of rain variables distributions across the MV classes (see supplementary material) showed a visible difference for MAX_STEP_REL, Max_int_50REL, NO_RAIN and DUR in both the 3 and the 8 MV cluster scenario. For NO_RAIN, the cluster that stands out the most in both scenarios includes the MV curves that are closest to the bisector (see Figures and), and which seem to have had a continuous rain pattern without breaks (low values of NO_RAIN). For MAX_STEP_REL and Max_int_50REL, the cluster that stands out the most in the 8 MV cluster scenario is the one including extremely concave curves (see Figure), because the maximum intensity of the rain has taken place late in the event, resulting in a high MAX_STEP_REL and a low Max_int_50REL. In the 3 cluster scenario, these curves have been mixed with the extremely convex curves (see), causing a larger span in the MAX_STEP_REL and Max_int_50REL values. Regarding DUR, both the very concave and convex curves (see - and) had long durations. There is a possible bias, as events with long duration will have more monitoring points (or data extraction points in the case of modelled data), meaning that they are more likely to display higher amplitudes, while events with shorter duration will have fewer monitoring points and therefore concentration variations might be missed.

The Kruskal-Wallis test (summarised in Table 4) revealed several cases of low p-value, including MAX_STEP_REL (only for 3 clusters), Max_int_50REL (only for 3 clusters), NO_RAIN and DUR.

Regarding the a priori classification (Table 6), there seems to be a slight trade-off between the increase in cRate and aRand. The maximum rate of correct assignment was 63% for the 3 MV classes scenario and 32% for the 8 classes scenario, and in general the aRand values suggest a performance that is marginally better than random assignment. The calculations that relied on only a single variable resulted in the following ranking of the rain (ordered by cRate):

- 3 MV classes: ADWP (0.52), NO_RAIN (0.50), DEPTH (0.48), DUR (0.47), AR2 (0.46), MAX_INT (0.44), Max_int_last (0.43), MAX_STEP_REL (0.33), MAX_INT_50REL (0.33), AR7 (0.30) and Month (0.27)
- 8 MV classes: DEPTH (0.30), NO_RAIN (0.30), DUR (0.30), MAX_INT (0.27), ADWP (0.25), AR2 (0.23), Max_int_last (0.23), MAX_INT_50REL (0.23), MAX_STEP_REL (0.21), AR7 (0.19) and Month (0.18).

The variable DEPTH is included in all the best performing combinations of variables, closely followed by NO_RAIN. ADWP was expected to be influential, as the pollutant build-up in the model directly depends on this variable. However, it is not included in the best performing combinations until 7-8 variables are combined. Instead, variables related to the washoff process, such as the MAX_INT, proved more influential.

3.4.3. Joint reflections

Overall, better MV class prediction was achieved with the modelled data set compared to the monitored, a direct consequence of the simplification of reality behind the used model. Other than measurement uncertainty, the long period covered by the Chassieu data might be affected by changes in the pollutant sources across the catchment (which, considering the industrial characteristics of the Chassieu catchment, cannot be excluded). Conversely, given the shorter monitoring period, pollutant sources were considered constant in the model of the Albertslund catchment. Both data sets are, however, influenced by the dynamics and non-linearities of the drainage systems, which was not taken into account for either data set when using only rain variables.

The Kruskal-Wallis test illustrated how the rain variables had easier discrimination of clusters for the cases with fewer MV classes, and similarly, the degree of correct cluster assignment (cRate) approximately doubled in the cases of fewer MV classes. However, aRand values did not improve for fewer MV classes and in fact decreased for the Albertslund data set, which underlines how fewer classes also mean higher chance of correct assignment by a random grouping.

When looking at both data sets, the gain in cRate from including multiple rain variables in combination as predictors seems almost negligible, as approximately the same accuracy is achieved from using only a single rain characteristic. While it was possible to recognise connections between the classes of MV curves and the individual rain variables, a satisfactory MV curve class assignment could not be reached purely based on rain characteristics for either of the data sets. This study included rain characteristics that were identified as relevant by other flush studies, which represent conditions linked to both each individual event, preceding events, and seasonality, and which are expressed in both relative (e.g. NO_RAIN) and absolute terms (e.g. DEPTH). It should be underlined that most of the variables listed in Table 2 are “lumped” variables, i.e. they considerably simplify the representation of rainfall events, in particular they neglect the dynamics within each event, i.e. the variation of the rainfall intensity at short time step. Pollutographs (and the resulting MV curves), depend on variations in the hyetograph (in addition to other causes different from the rainfall events). Consequently, it is not surprising that simple global indicators, even combined or aggregated, are not fully relevant to predict MV curves, and it seems unlikely that any additional rain characteristics could significantly improve the MV class assignment. Ultimately, it must be concluded that rain characteristics seem to influence the flush signal, but that clusters of rain variables alone are not sufficient as explanatory variables for the classes of MV curves created by the FDA approach.

3.5. Future research areas

The results of this study represent only a first step into the application of new methods to describe the behaviour of stormwater pollution during rain event. As such, new opportunities and research areas are now outlined:

- Multi-catchment comparison. The Chassieu data set represent so far one of the largest available TSS data sets, but it can be expected that new monitoring campaigns will provide new data that will increase the applicability of FDA based approaches, and will allow to investigate the presence of site-specific and/or more general patterns;
- Identification of new explanatory variables, capable of describing and understanding the drivers behind the inter- and intra-event variability, strengthening the predictive power of water quality models. This would require the analysis of additional data sets from other study areas to identify variables linked to the catchment spatial characteristics (e.g. land usage, climatic regime);
- Formalised description of clusters, i.e. describing general curves for each cluster, corresponding to different flush behaviours. These can eventually be used to assess the performance of pollution control strategies under different ”flush scenarios”. This approach resembles the one increasingly applied in stormwater quantity management, where SCMs are evaluated under different rain domains (3 Points Approach - (Fratini et al., 2012));
- Use of FDA to quantify the uncertainties related to pollutant level variations. Indeed, FDA was successfully able to extract information on typical MV curve shapes and their frequency of occurrence. A better quantification and communication of the uncertainties connected to flush and pollutant concentrations will enable stormwater managers to assess the risk of exceedance of water quality standards caused by these pollutant flux variations;
- Development of pollution control strategies, based on the identified drivers, which can help stormwater managers to minimise pollution

load and/or maximise their interception (e.g. source control actions targeting identified drivers);
- Development of model-based assessment of dynamic control actions aiming at maximising the pollutant load intercepted and treated by SCMs (building e.g. upon the approach presented in Ly et al. (2019).

4. Conclusion

The analysis and categorisation of flush effects in urban runoff from separate sewer systems showed how unsupervised methods (Functional Data Analysis (FDA), including functional Principal Components Analysis (fPCA) and functional clustering using the EMCluster and kCFC algorithms) were able to identify different flush patterns emerging from the analysed data. Since this novel approach is not limited by any predetermined flush definition scheme, it is thus more suitable for identifying specific characteristics from a specific monitored site.

Specifically, the FDA approach showed its capacity to analyse TSS MV curves from the Chassieu (measured data) and Albertslund (modelled data), describing and classifying different types of MV curve shapes, including those crossing the bisector (which cannot be described by the polynomial approximations used in earlier studies).

A Kruskal-Wallis test showed that the MV classed defined by FDA divided several of the recorded rain variables into groups with significantly different values. When using a different number of clusters, the a priori clustering of rain characteristics obtained a correct class assignment of only 23% for Chassieu and 32% for Albertslund with an optimal number of classes (10 and 8, respectively) and 45% and 63% with a number of classes that is judged as feasible by decision makers and stormwater planners (4 and 3 respectively).

The analysis of the modelled data corroborated the analysis of the measured data, showing how even in data characterised by a lower variability than measured data (consequence of the model simplification of the processes taking place in the real system, along with its sole dependency on the rainfall input) the FDA approach did not identify a clear influence of the rainfall variables. All these results suggest that additional variables (not related to rainfall characteristics) should be investigated as explanatory variables for flush phenomena.

The application of FDA to MV curves opens for a variety of different research areas, building upon the proposed curve clustering. On the long term, the investigated methodology will provide decision makers in urban stormwater management with vital knowledge to understand the current system and plan robust management pollution control strategies.

CRediT authorship contribution statement

Ditte Marie Reinholdt Jensen: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Santiago Sandoval:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Jean-Baptiste Aubin:** Methodology, Investigation, Writing – review & editing. **Jean-Luc Bertrand-Krajewski:** Data curation, Writing – review & editing. **Li Xuyong:** Conceptualization, Writing – review & editing. **Peter Steen Mikkelsen:** Conceptualization, Investigation, Writing – review & editing. **Luca Vezzaro:** Conceptualization, Methodology, Data curation, Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge the Sino-Danish Center for Education and Research (SDC) for the financial support provided to the first author to carry out PhD research studies. Furthermore, we are grateful to Niels Aske Lundtorp Olsen from DTU Compute for his insights on functional data analysis and Lena Mutzner from DTU Environment for sparring in regards to scoping and presenting the work.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118394

References

Athanasiadis, K., Horn, H., Helmreich, B., 2010. A field study on the first flush effect of copper roof runoff. Corrosion Science 52 (1), 21–29. https://doi.org/10.1016/j.corsci.2009.08.048.

Bach, P.M., McCarthy, D.T., Deletic, A., 2010. Redefining the stormwater first flush phenomenon. Water Research 44 (8), 2487–2498. https://doi.org/10.1016/j.watres.2010.01.022.

Bertrand-Krajewski, J.L., Chebbo, G., Saget, A., 1998. Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. Water Research 32 (8), 2341–2356. https://doi.org/10.1016/S0043-1354(97)00420-X.

Brudler, S., Rygaard, M., Arnbjerg-Nielsen, K., Hauschild, M.Z., Ammitsœ, C., Vezzaro, L., 2019. Pollution levels of stormwater discharges and resulting environmental impacts. Science of the Total Environment 663, 754–763. https://doi.org/10.1016/j.scitotenv.2019.01.388.

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Mueller, H.-G., Wang, Jane-Ling, 2020. fdapace: Functional Data Analysis and Empirical Dynamics. R package version 0.5.5. https://cran.r-project.org/package=fdapace.

Chen, K., Zhang, X., Petersen, A., Müller, H.G., 2017. Quantifying Infinite-Dimensional Data: Functional Data Analysis in Action. Statistics in Biosciences 9 (2), 582–604. https://doi.org/10.1007/s12561-015-9137-5.

Chen, W.-C., Maitra, R., 2015. EMCluster: EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution. R Package. http://cran.r-project.org/package=EMCluster.

Chiou, J.-M., Li, P.-L., 2007. Functional Clustering and Identifying Substructures of Longitudinal Data. Royal Statistical Society 69 (4), 679–699.

Chow, M.F., Yusop, Z., 2014. Sizing first flush pollutant loading of stormwater runoff in tropical urban catchments. Environmental Earth Sciences 72 (10), 4047–4058. https://doi.org/10.1007/s12665-014-3294-6.

Craney, T.A., Surles, J.G., 2002. Model-dependent variance inflation factor cutoff values. Quality Engineering 14 (3), 391–403. https://doi.org/10.1081/QEN-120001878.

Deletic, A., 1998. The first flush load of urban surface runoff. Water Research 32 (8), 2462–2470. https://doi.org/10.1016/S0043-1354(97)00470-3.

Deletic, A., Orr, D.W., 2005. Pollution Buildup on Road Surfaces. Journal of Environmental Engineering-Asce 131 (1), 49–59. https://doi.org/10.1061/(ASCE)0733-9372(2005)131:1(49).

Deletic, A.B., Maksimovic, v.T., 1998. Evaluation of water quality factors in storm runoff from paved areas. Journal of Environmental Engineering (September), 869–879.

Eriksson, E., Baun, A., Mikkelsen, P.S., Ledin, A., 2005. Chemical hazard identification and assessment tool for evaluation of stormwater priority pollutants. Water Science and Technology 51 (2), 47–55. https://doi.org/10.1371/journal.pone.0084006.

Eriksson, E., Ledin, A., Baun, A., Lützhøft, H.-C.H., Mikkelsen, P.S., 2011. Stormwater priority pollutants versus surface water quality criteria. ATV Jord og Grundvand (november), 33–44.

Fratini, C.F., Geldof, G.D., Kluck, J., Mikkelsen, P.S., 2012. Three Points Approach (3PA) for urban flood risk management: A tool to support climate change adaptation through transdisciplinarity and multifunctionality. Urban Water Journal 9 (5), 317–331. https://doi.org/10.1080/1573062X.2012.668913.http://www.tandfonline.com/doi/abs/10.1080/1573062X.2012.668913

Göbel, P., Dierkes, C., Coldewey, W.G., 2007. Storm water runoff concentration matrix for urban areas. Journal of Contaminant Hydrology 91 (1-2), 26–42. https://doi.org/10.1016/j.jconhyd.2006.08.008.

Hubert, L., Arabie, P., 1985. Comparing partitions. Journal of Classification 2 (1), 193–218. https://doi.org/10.1007/BF01908075.

Jensen DMR. Challenging current practices for management of pollution in separate stormwater discharges. Kgs. Lyngby: DTU Environment, 2022. 222 p. https://orbit.dtu.dk/en/publications/challenging-current-practices-for-management-of-pollution-in-sepa.

Kang, J.H., Kayhanian, M., Stenstrom, M.K., 2008. Predicting the existence of stormwater first flush from the time of concentration. Water Research 42 (1-2), 220–228. https://doi.org/10.1016/j.watres.2007.07.001.

Kong, Z., Shao, Z., Shen, Y., Zhang, X., Chen, M., Yuan, Y., Li, G., Wei, Y., Hu, X., Huang, Y., He, Q., Chai, H., 2021. Comprehensive evaluation of stormwater pollutants characteristics, purification process and environmental impact after low impact development practices. Journal of Cleaner Production 278, 123509. https://doi.org/10.1016/j.jclepro.2020.123509.

Koziel, L., Juhl, M., Egemose, S., 2019. Effects on biodiversity, physical conditions and sediment in streams receiving stormwater discharge treated and delayed in wet ponds. Limnologica 75 (January), 11–18. https://doi.org/10.1016/j.limno.2019.01.001.

Kruskal, W.H., Wallis, W.A., 1952. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47 (260), 583–621. https://doi.org/10.1080/01621459.1952.10483441.

Kus, B., Kandasamy, J., Vigneswaran, S., Shon, H.K., 2010. Analysis of first flush to improve the water quality in rainwater tanks. Water Science and Technology 61 (2), 421–428. https://doi.org/10.2166/wst.2010.823.

Lee, J.H., Bang, K.W., Ketchum, J.H., Choe, J.S., Yu, M.J., 2002. First flush analysis of urban storm runoff. Science of the Total Environment 293 (1-3), 163–175. https://doi.org/10.1016/S0048-9697(02)00006-2.

Li, D., Wan, J., Ma, Y., Wang, Y., Huang, M., Chen, Y., 2015. Stormwater runoff pollutant loading distributions and their correlation with rainfall and catchment characteristics in a rapidly industrialized city. PLoS ONE 10 (3), 1–17. https://doi.org/10.1371/journal.pone.0118776.

Li, W., Shen, Z., Tian, T., Liu, R., Qiu, J., 2012. Temporal variation of heavy metal pollution in urban stormwater runoff. Frontiers of Environmental Science and Engineering in China 6 (5), 692–700. https://doi.org/10.1007/s11783-012-0444-5.

Ly, D.K., Maruéjouls, T., Binet, G., Bertrand-Krajewski, J.L., 2019. Application of stormwater massvolume curve prediction for water quality-based real-time control in sewer systems. Urban Water Journal 16 (1), 11–20. https://doi.org/10.1080/1573062X.2019.1611885.

Ma, Y., Deilami, K., Egodawatta, P., Liu, A., McGree, J., Goonetilleke, A., 2019. Creating a hierarchy of hazard control for urban stormwater management. Environmental Pollution 255, 113217. https://doi.org/10.1016/j.envpol.2019.113217.

Métadier, M., Bertrand-Krajewski, J.L., 2011. From mess to mass: A methodology for calculating storm event pollutant loads with their uncertainties, from continuous raw data time series. Water Science and Technology 63 (3), 369–376. https://doi.org/10.2166/wst.2011.230.

Métadier, M., 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. https://doi.org/10.1016/j.watres.2011.12.030.

Millán-Roures, L., Epifanio, I., Martínez, V., 2018. Detection of anomalies in water networks by functional data analysis. Mathematical Problems in Engineering 2018. https://doi.org/10.1155/2018/5129735.

Müller, A., Österlund, H., Marsalek, J., Viklander, M., 2020. The pollution conveyed by urban runoff: A review of sources. Science of the Total Environment 709. https://doi.org/10.1016/j.scitotenv.2019.136125.

Perera, T., McGree, J., Egodawatta, P., Jinadasa, K.B.S.N., Goonetilleke, A., 2019. Taxonomy of influential factors for predicting pollutant first flush in urban stormwater runoff. Water Research 166. https://doi.org/10.1016/j.watres.2019.115075.

Perera, T., McGree, J., Egodawatta, P., Jinadasa, K.B.S.N., Goonetilleke, A., 2021. New conceptualisation of first flush phenomena in urban catchments. Journal of Environmental Management 281 (December 2020). https://doi.org/10.1016/j.jenvman.2020.111820.

Qin, H.p., He, K.m., Fu, G., 2016. Modeling middle and final flush effects of urban runoff pollution in an urbanizing catchment. Journal of Hydrology 534 (June), 638–647. https://doi.org/10.1016/j.jhydrol.2016.01.038.

Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66 (336), 846–850. https://doi.org/10.1080/01621459.1971.10482356.

Sandoval, S., Vezzaro, L., Bertrand-Krajewski, J.L., 2018. Revisiting conceptual stormwater quality models by reconstructing virtual state variables. Water Science and Technology 78 (3), 655–663. https://doi.org/10.2166/wst.2018.337.

Schriewer, A., Horn, H., Helmreich, B., 2008. Time focused measurements of roof runoff quality. Corrosion Science 50 (2), 384–391. https://doi.org/10.1016/j.corsci.2007.08.011.

Sharifi, S., Massoudieh, A., Kayhanian, M., 2011. A Stochastic Stormwater Quality Volume-Sizing Method with First Flush Emphasis. Water Environment Research 83 (11), 2025–2035. https://doi.org/10.2175/106143011x12989211.

Suhaila, J., Yusop, Z., 2017. Spatial and temporal variabilities of rainfall data using functional data analysis. Theoretical and Applied Climatology 129 (1-2), 229–242. https://doi.org/10.1007/s00704-016-1778-x.

Sun, S., Barraud, S., Castebrunet, H., Aubin, J.B., Marmonier, P., 2015. Long-term stormwater quantity and quality analysis using continuous measurements in a French urban catchment. Water Research 85, 432–442. https://doi.org/10.1016/j.watres.2015.08.054.

Ternynck, C., Alaya, M.A.B., Chebana, F., Dabo-Niang, S., Ouarda, T.B.M.J., 2016. Streamflow hydrograph classification using functional data analysis. Journal of Hydrometeorology 17 (1), 327–344.https://www.jstor.org/stable/24915571

Vezzaro, L., Mikkelsen, P.S., Deletic, A., McCarthy, D.T., 2013. Urban drainage models simplifying uncertainty analysis for practitioners. Water Science and Technology 68 (10), 2136–2143.

Vezzaro, L., Sharma, A.K., Ledin, A., Mikkelsen, P.S., 2015. Evaluation of stormwater micropollutant source control and end-of-pipe control strategies using an uncertainty-calibrated integrated dynamic simulation model. Journal of Environmental Management 151, 56–64. https://doi.org/10.1016/j.jenvman.2014.12.013.

Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M., Morgan, R.I. P., 2005. The urban stream syndrome: current knowledge and the search for a cure. J. N. Am. Benthol. Soc. 24 (3), 706–723.

Wang, J.L., Chiou, J.M., Müller, H.G., 2016. Functional Data Analysis. Annual Review of Statistics and Its Application 3, 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624.

Zgheib, S., Moilleron, R., Saad, M., Chebbo, G., 2011. Partition of pollution between dissolved and particulate phases: What about emerging substances in urban stormwater catchments? Water Research 45 (2), 913–925. https://doi.org/10.1016/j.watres.2010.09.032.