Water Science & Technology

Water Science & Technology Vol 87 No 9, 2292 doi: 10.2166/wst.2023.125

Pattern-based assessment of the influence of rainfall characteristics on urban stormwater quality

Anupam Chowdhury[a,*], Prasanna Egodawatta[b] and James McGree[b]
[a] Department of Civil Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[b] Science and Engineering Faculty, Queensland University of Technology (QUT), GPO Box 2434, Brisbane, Queensland 4001, Australia
*Corresponding author. E-mail: anupam@ce.ruet.ac.bd

ABSTRACT

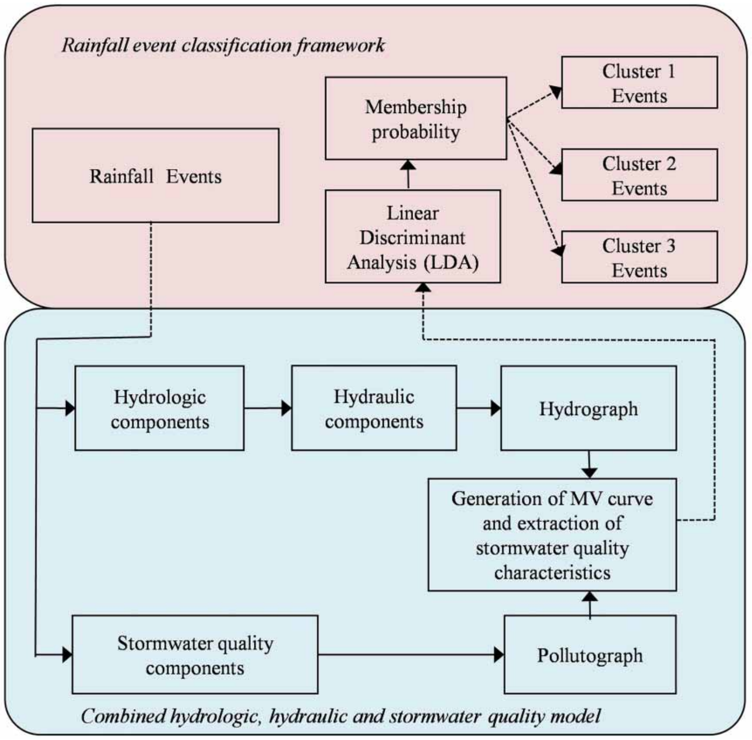
Urbanisation increases pollutant generation within catchments and their transport to receiving waters. Changes to rainfall patterns, particularly in the age of climate change, make pollution mitigation a challenging task. Understanding how rainfall characteristics could influence the changes to stormwater pollutant runoff is important for designing effective mitigation strategies. This study employed a pattern-based assessment of relationships between rainfall characteristics and stormwater quality in urban catchments to develop this understanding. The research outcomes showed that rainfall events could be distinctly clustered based on intensity and duration, and each cluster of events would produce different stormwater quality responses. The high-intensity bursts occurring in the latter part of long-duration events were found to produce uniform and low concentrations of suspended solids. One the contrary, high intensity bursts occurring in the initial part of short-duration events triggered the first-flush effect, thus producing high concentrations of suspended solids. Furthermore, the first-flush effect was likely to present when the high intensity bursts occurred in the mid portion of rainfall events and produced variable concentrations of suspended solids. It was also found that the average rainfall intensity plays a key role in mobilising and transporting pollutants accumulated on urban surfaces.

Key words: first flush, rainfall characteristics, rainfall pattern, stormwater quality, urban water pollution

HIGHLIGHTS

- Rainfall events were clustered based on stormwater quality responses.
- Different portions of rainfall produce different stormwater quality responses.
- High-intensity bursts in short-duration rainfall events trigger first flush.
- High-intensity bursts in the mid portion of rainfall are likely to trigger first flush.
- The average rainfall intensity plays a key role in mobilising and transporting pollutants.

GRAPHICAL ABSTRACT



1. INTRODUCTION

As urbanisation transforms natural lands into built environments, urban receiving waters are becoming degraded due to the discharge of pollutant loads through stormwater runoff (Wijesiri *et al.* 2018; Seifollahi-Aghmiuni *et al.* 2022). This makes improving urban liveability a challenging issue as the quality of water resources is critical to protecting both human and ecosystem health (Qi *et al.* 2020; Wang *et al.* 2021). In fact, this has been exacerbated due to the impacts of climate change which significantly and continuously alters typical patterns of rainfall (Mujere & Moyce 2018; Han & Bu 2023; Xiong *et al.* 2023).

For example, lengthy dry periods between rainfall events and high-intensity, short-duration rainfall events have been predicted as a result of climate change (Batalha *et al.* 2018; Stein *et al.* 2021). Longer dry periods could create an opportunity for the continuous accumulation of pollutants on urban surfaces, which in turn influence the pollutant loads in stormwater runoff. On the other hand, high rainfall intensity could exert a greater impact (due to the kinetic energy of raindrops) on pollutants attached to surfaces such as roads, and in turn, easily mobilise. Most importantly, high-intensity rainfall events that have shorter duration can potentially increase the impact of 'the first flush' that typically occurs at the initial part of a rainfall event. First flush results in shock-loads of pollutants in stormwater runoff that eventually are transported into receiving waters (Wijesiri *et al.* 2020a, 2020b).

As Chaudhary *et al.* (2022) noted, peak rainfall intensity and its timing with respect to the start of a rainfall event trigger the first flush. Therefore, most stormwater treatment designs that have relatively small capacities target capturing highly concentrated initial portions of runoff volume from a larger number of frequent events. However, variable responses of stormwater quality to rainfall characteristics, such as rainfall intensity, depth, antecedent dry days (ADDs) and duration (Yan *et al.* 2023), make it difficult to acquire high performance in pollutant removal (Zhang *et al.* 2019). Furthermore, considering the overall characteristics of a rainfall event for stormwater treatment design practices, disregarding its temporal pattern overshadows the true influence of rainfall characteristics on stormwater quality. This would not benefit designing effective pollution mitigation measures (Wijesiri *et al.* 2020a; Yan *et al.* 2023). An in-depth, pattern-based analysis is necessary to comprehend the impact of rainfall characteristics on stormwater quality.

However, past attempts to link rainfall events and stormwater quality have been challenging due to the lack of reliable data on stormwater quantity and quality. While reliable and appropriate data are scarce, there is no appropriate modelling tool available to simulate rainfall scenarios accurately and generate stormwater quality and quantity data. In fact, most of the commonly used hydrological modelling tools are not capable of accurately simulating complex hydrological phenomena because (1) complex models require data that are not available in common databases; (2) simplistic models generate less accurate estimations; (3) lack of data for model calibration; (4) models having correlating parameters result in problematic calibration outcomes and (5) models lack in-built automatic calibration procedures (MikeUrban 2019; King *et al.* 2021; Decsi *et al.* 2022). Furthermore, the approach of model calibration for parameter estimation can have a critical effect on the accuracy of final predictions. In this instance, having an automatic calibration procedure is an added advantage for most of the hydrological modelling tools. This is because manual calibration or 'trial and error' process can fail as the number of model parameters increases and for large datasets (Tian *et al.* 2019). This highlights the importance of using a specialised stormwater quality modelling tool with robust replication methods and automatic calibration procedures (Chowdhury & Egodawatta 2022).

The investigation discussed in this paper focused on identifying the key features of rainfall temporal patterns and predicting their influences on urban stormwater quality. This was achieved by generating reliable stormwater quantity and quality data using a computer-based modelling tool, thus identifying clusters of rainfall events based on rainfall characteristics and their corresponding stormwater quality responses. The new insights into how stormwater quality varies depending on the changes in rainfall characteristics will contribute to generating reliable information about catchment stormwater quality and accurately formulating stormwater quality modelling tools and thereby contributing to designing effective stormwater pollution mitigation measures.

2. MATERIALS AND METHODS

2.1. Study sites

The study was undertaken based on field data relating to the Gold Coast region, Australia. Three small urban catchments, namely, Alextown, Gumbeel and Birdlife Park, located within the larger Highland Park catchment, were selected. The catchment selection was based on the catchment area, impervious fraction, urban form and the availability of historical rainfall data. The details of the distinct characteristics of the catchments are presented in Figure 1.

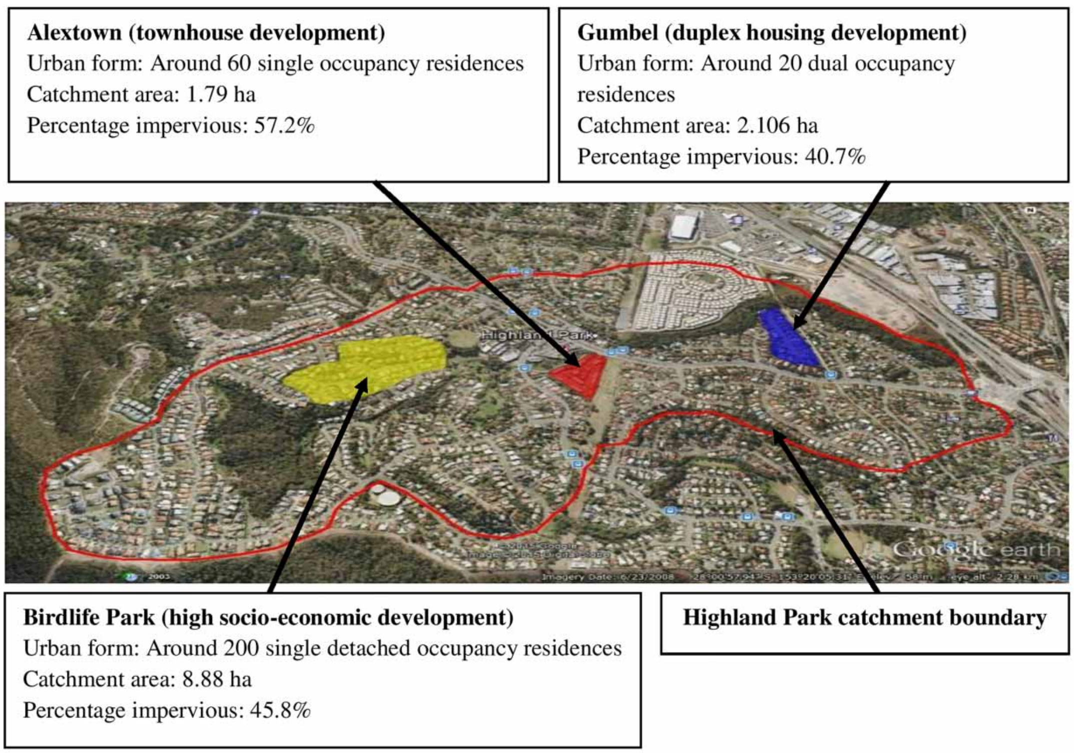


Figure 1 | Locations of study catchments and their characteristics (map data: Google, Digital Globe).

The selection of a rain gauge station and the collection of historical rainfall events were necessary to generate the required datasets for model simulations. Hence, three pluviograph stations (Hinze Dam, Coombabah WTP and Gold Coast Seaway) located in the vicinity of the study areas were selected from the Bureau of Meteorology (BOM) website for initial analysis. The locations of the rain gauge stations around the study areas are shown in Fig. S1. The station number, location coordinates and data ranges for three rain gauge stations are shown in Table S1. Among these three stations, the Hinze dam station is the closest to the study catchments compared to the other two stations. Hence, data from the Hinze Dam station was selected for the modelling exercise, which ranges from January 2000 to August 2014. This is because the spatial and temporal variability of rainfall increases with the increase in spatial distance from the point of measure (Emmanuel *et al.* 2012).

2.2. Selection of rainfall and water quality characteristics

To assess the impact of rainfall characteristics on stormwater quality, the selection of suitable rainfall characteristics is critically important. As pointed out by Egodawatta *et al.* (2007), the inclusion of insignificantly correlating characteristics in the analysis can overshadow the actual relationships, while ignoring important characteristics can lead to poorly performing relationships. As such, average intensity (AI), ADD and total duration (TDU) were selected as they have been identified as the major factors influencing stormwater quality (Mahbub *et al.* 2010; Zhang *et al.* 2015; Gong *et al.* 2016; Liu *et al.* 2016).

On the other hand, stormwater quality varies from the start to the end of a runoff event. This is primarily due to the difference in wash-off characteristics induced by the temporal variations of the rainfall event. Therefore, first, it was necessary to understand this variability. Thus, pollutant wash-off load at different fractions of runoff volume was determined using an *MV* curve. An *MV* curve is a graphical illustration where *M* denotes the pollutant load as a percentage with respect to the total and *V* denotes the runoff volume as a percentage with respect to the total. The *MV* curve is widely used to explain the phenomenon of first flush (Kang *et al.* 2008; Bach *et al.* 2010; Alias 2013).

To identify the subtle changes to stormwater quality due to variable rainfall characteristics, 10 water quality parameters (P01, P12, P23, P34, P45, P56, P67, P78, P89 and P910) were generated from the *MV* curve for this study as shown in Fig. S2 in the Supplementary Information. Stormwater quality parameters were selected based on the hypothesis that rainfall characteristics in a given segment of a rainfall event influence the pollutant wash-off behaviour, creating variability in stormwater quality characteristics. In this regard, the parameter P_{xy} was defined such that it denotes the percentage increment of pollutant load with respect to the increase of runoff volume from '*x*' to '*y*' percentage. For example, parameter P12 indicates the amount of total suspended solid (TSS) load increment percentage corresponding to the increment of cumulative percentage runoff volume from 10 to 20%.

2.3. Selection of rainfall events and generation of the data matrix

It was necessary to identify a sufficiently large number of rainfall events for which the data of rainfall depth are available, preferably at the 1-min intervals. This is to ensure the reliability of modelling outcomes. Therefore, historical rainfall records from the Hinze Dam station for a 10-year period (2003–2012) were selected for this study. These rainfall records were pre-processed and 251 events were selected by avoiding small and dependent events that would not cause significant wash-off. This was done by considering the start of an event when intensity exceeds 5 mm/h, and the total depth is more than 3 mm and the overall average intensity is more than 5 mm/h. Accordingly, the average intensity, the antecedent dry period and the duration of these events were obtained. Stormwater quality and quantity for these selected 251 rainfall events were generated by simulating them using a combined hydrological, hydraulic and stormwater quality (CHHSWQ) model developed for selected study catchments. The model was developed using commonly adopted techniques for modelling hydrological, hydraulic and pollutant processes and an automatic calibration framework based on Approximate Bayesian Computation (ABC). Further details of the model development, calibration and validation processes can be found in Chowdhury & Egodawatta (2022).

The strategy adopted in this study is to assess the calibration ranges of the developed CHHSWQ model and use the model to evaluate the characteristics of historical 251 rainfall events. As illustrated in Fig. S3 in the Supplementary Information, it was evident that the characteristics of 52 observed rainfall events used for CHHSWQ model calibration are closely comparable and scattered in the envelope created by the 251 historical rainfall events. This suggested that the calibrated CHHSWQ model is suitable for simulating the rainfall characteristics of 251 historical rainfall events.

Direct model outputs of 251 historical event simulations were pre-processed on an individual event basis to generate a relevant data matrix for this study. In this regard, the *MV* curve, as discussed in Section 2.2 for each event, was plotted and the selected 10 stormwater quality parameters were extracted from each event. Accordingly, a large data matrix of (251×13) was formed for each study catchment by combining the datasets of rainfall (AI, TDU and ADD) and stormwater quality characteristics (P01, P12, P23, P34, P45, P56, P67, P78, P89, P910). The generated data matrix will further enable to identify the unique pattern of rainfall events based on their stormwater quality characteristics. In this regards, cluster analysis is one of the techniques used for identifying similar structural characteristics of a dataset by separating the data into groups or clusters. It is a data exploration technique and is commonly used for those datasets where prior information on cluster number and membership is unknown.

3. RESULTS AND DISCUSSIONS

3.1. Clustering of rainfall events based on stormwater quality

To identify similar rainfall events in relation to their stormwater quality response, cluster analysis was first performed on stormwater quality data using the water quality variables obtained for each event. *K*-means clustering was used as the primary technique of clustering. The details of the *K*-means clustering technique can be found in Jain (2010) and Liu & Liu (2016).

The determination of the optimum cluster number (*K*-value) is an important part of *K*-means clustering procedure. The strategic selection of *K*-value leads to efficient partitioning of a given dataset (Jain 2010). In this regard, two methods, namely, the Elbow method and the Bayesian Information Criterion (BIC), were applied to define the *K*-value prior to the application of *K*-means clustering. The analytical outcomes of the Elbow method and the BIC are given in Fig. S4 in the Supplementary Information. It was evident that the outcomes of the Elbow methods were consistent with cluster 3, which was also justified by the outcomes of BIC. This suggests that three clusters are needed for the analysis of the stormwater quality dataset.

K-means clustering was then performed for three water quality data matrices of Alextown, Gumbeel, and Birdlife Park catchments, each consisting of 251 events and 10 parameters of stormwater quality. Due to having many variables and dimensions of stormwater quality data, direct visualisation of the clustered events could be complex. To overcome this, principal component analysis (PCA) was applied to decrease the dimensions of the generated data matrix.

The PCA is a data retrenchment process commonly used to decrease a large number of variables to a lesser number of self-reliant new variables known as principal components (PCs). In PCA, the PCs should be selected in such a way that the selected PCs can explain most of the variance of the whole dataset. Further details of the PCA can be found in Bro & Smilde (2014). To identify the number of significant PCs, Eigen values corresponding to each component were plotted as a scree-diagram for the three catchments, as shown in Fig. S5 in the Supplementary Information. Accordingly, the first two PCs have an Eigen value greater than 2 and have explained 64.8, 57.6 and 55.2% of the total data variance for Alextown, Gumbeel and Birdlife Park catchments, respectively. Hence, the first two components were considered adequate for the analysis of all three catchments.

The selected first two PCs of PCA outcomes are plotted in Figure 2 for the visualisation of the clustered objects. From Figure 2, it is evident that the three clusters (marked by black, red and green circles) show three distinct groups of stormwater quality characteristics. This confirmed that the identified three clusters represent three unique stormwater quality characteristic patterns.

3.2. Classification of rainfall events based on stormwater quality

Given that there is a strong relationship between stormwater quality responses (pollutograph) and characteristics of rainfall events, it can be hypothesised that the clustering of events into three groups, based on stormwater quality, is affected by rainfall characteristics. This hypothesis was tested by applying linear discriminant analysis (LDA) to the dataset; the procedure is described in the Supplementary Information.

There was a potential for events to be wrongly classified compared to the clarification received in *K*-means clustering. This means that the events previously classified into three groups based on stormwater quality characteristics (by *K*-means clustering) may not show the same cluster memberships when classified according to rainfall characteristics. To study this, the percentage of correctly classified events was calculated and presented in Table S2 in the Supplementary Information. Accordingly, a total of 251 rainfall events, 139 events, 128 events and 133 events were correctly classified into three groups for Alextown, Gumbeel and Birdlife Park catchments, respectively. The overall percentage of correctly classified events for Alextown was found to be 55.4%, whereas the individual clusters 1, 2 and 3 had 50, 60.3 and 55.8%, respectively. Similarly, the overall percentage of correctly classified events for Gumbeel and Birdlife Park was found to be 51 and 52.99%, where the

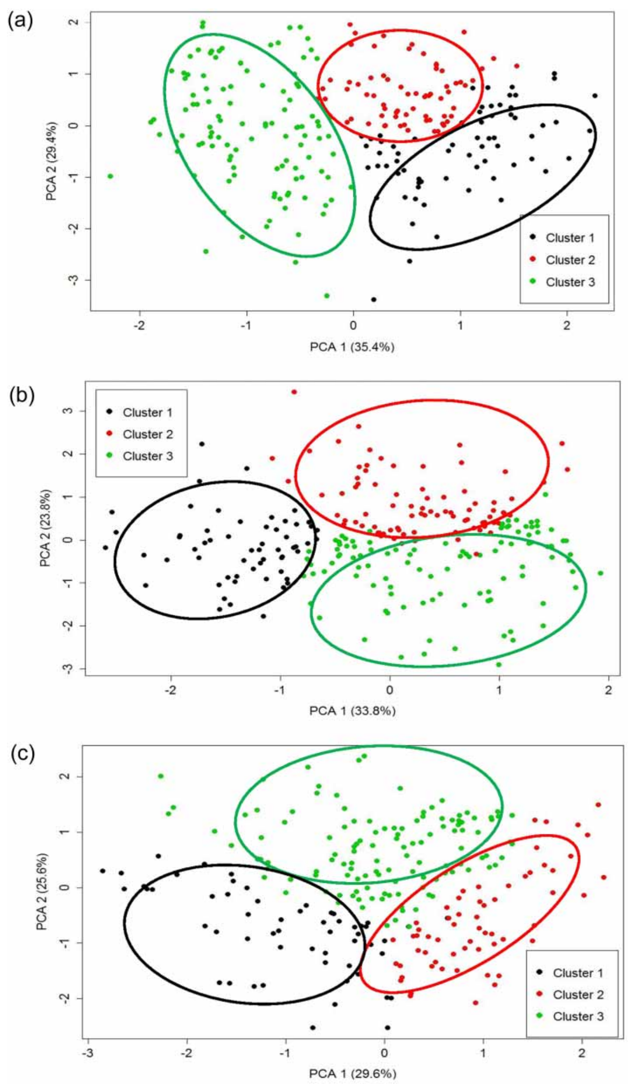


Figure 2 | PCA plot of three clusters: (a) Alextown; (b) Gumbeel; (c) Birdlife Park.

individual clusters 1, 2 and 3 had 33.33, 53.01, 56.41% and 66.07, 27.87, 58.96%, respectively. This suggests that the above hypothesis is partially true. Over 50% correct classification of events shows that there is a significant impact of rainfall characteristics on stormwater quality. However, it is not the only dominant influence for certain rainfall events. Other influences, such as catchment characteristics, may overshadow the influence of rainfall characteristics for certain rainfall events. The details of the 10 water quality parameter (P01, P12, P23, P34, P45, P56, P67, P78, P89 and P910) data for three clustered events for study catchments are presented in Tables S3–S5 in the Supplementary Information.

Each rainfall event classified into one of the pre-defined groups can be assessed using the membership probability plot as shown in Figures S6–S8 in the Supplementary Information. As such, each event represents membership probability corresponding to three cluster groups. According to the classification rule, rainfall events are classified to the cluster group that represents the highest membership probability. In order to understand the distinction in rainfall characteristics among three cluster groups, univariate analysis was undertaken and details are presented in the following section.

3.2.1. Univariate analysis of rainfall clusters

The rainfall characteristics such as TDU, average intensity and ADDs for three rainfall clusters are presented in Box–Whisker plots as illustrated in Figure 3. Accordingly, cluster 1 contains low-intensity long durational events with low ADDs. Cluster 2

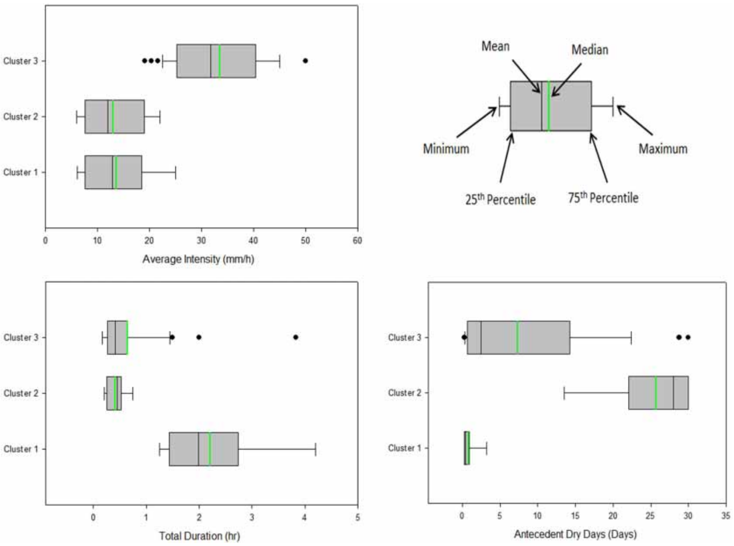


Figure 3 | Rainfall characteristics of three class patterns.

contains slow intensity and has short durational events with high ADDs. Cluster 3 contains high-intensity events with variable durations and ADDs.

Cluster 3 events show relatively high average intensity compared to cluster 1 and 2 events. The average intensities for cluster 1 and 2 events are in similar ranges from 6 to 25 and 6 to 22 mm/h, while the range of cluster 3 events is 20–50 mm/h. This suggests the potential for cluster 3 events to generate a high wash-off load compared to cluster 1 and 2 events. In the case of total duration, cluster 1 events are of comparatively longer duration (1.2–4.3 h) than cluster 2 (0.25–0.75 h) and cluster 3 (0.25–3.8 h) events. In contrast, ADDs for cluster 1 events were found to be less than 5 days, which is comparatively lower than cluster 2 and 3 events. Cluster 2 events show high ADD greater than 30 days, whereas ADDs for cluster 3 events vary from 2 to 32 days.

The above discussion suggests that the identified three clusters have distinct rainfall characteristics. Hence, it could be concluded that the classification of events into three clusters that depend on rainfall characteristics is accurate. However, in-depth knowledge between the rainfall clusters and stormwater quality characteristics is essential for understanding how first flush could influence the pollutant loads in stormwater runoff during different rainfall events and thereby designing effective stormwater treatment systems.

3.2.2. Multivariate analysis of rainfall clusters

Understanding the reasons for generating unique stormwater quality responses for the identified three clusters are important. This required a detailed analysis of the identified relationships with respect to existing knowledge bases relating to pollutant wash-off and transport behaviours. Developing such understanding requires a detailed assessment of both stormwater quality-related variables and rainfall characteristic variables in the same platform. This was done by using PCA with three matrices of 139, 128 and 133 events for Alextown, Gumbeel and Birdlife Park catchments, respectively, with each having 13 variables P01, P12, P23, P34, P45, P56, P67, P78, P89, P910, average intensity, total duration and ADDs. The output PCA bi-plots for the three catchments are shown in Figure 4. According to the Eigen value criterion, the first two PCs explain 55.5, 50.1 and 50.1% of data variance for Alextown, Gumbeel and Birdlife Park, respectively, which were considered acceptable for the assessment. The scree plot for PCA is shown in Figures S9–S11 in the Supplementary Information.

As evident from Figure 4, the three clusters show clear partitioning and are associated with three distinct rainfall and stormwater quality variable groups, even though catchment characteristics are different for Alextown, Gumbeel and Birdlife Park. This suggests that stormwater quality can be represented by three characteristic patterns of rainfall events. The positioning of the three clusters in the PC1–PC2 plane and with respect to the rest of the variables varies for the three catchments. In addition, rainfall events classified into three clusters for one catchment are not exactly similar to the other clustered

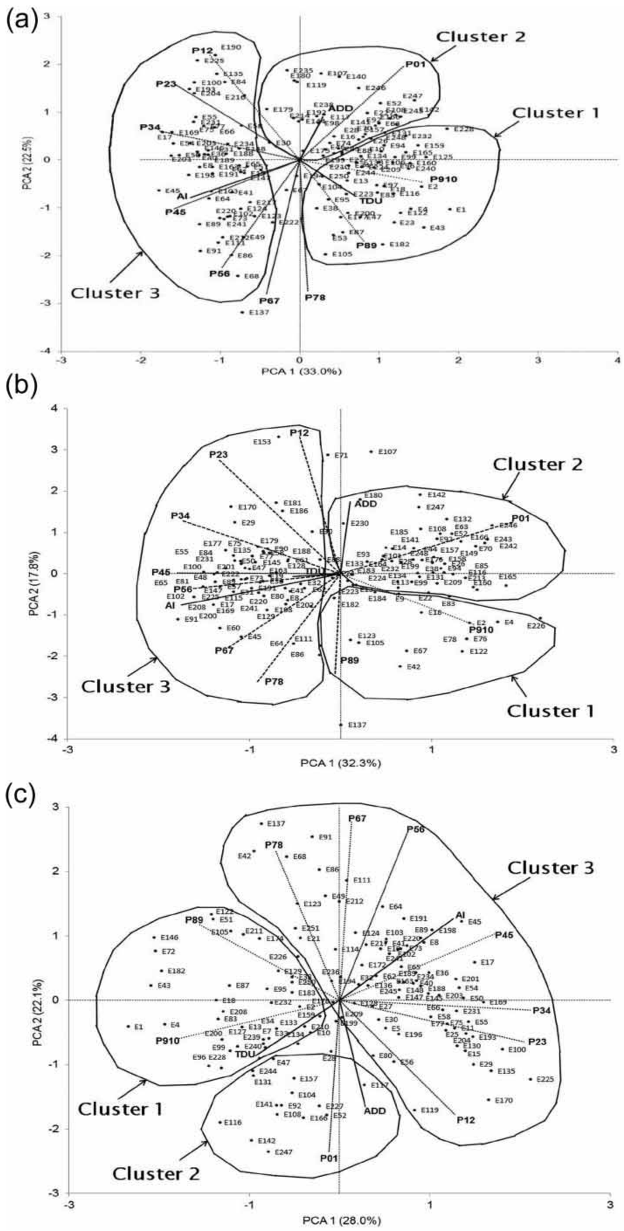


Figure 4 | PCA of correctly classified rainfall events: (a) Alextown – 139 events; (b) Gumbeel – 128 events; (c) Birdlife Park – 133 events.

events for the three catchments. This indicates that stormwater quality characteristics can vary from catchment to catchment, even though the same rainfall events are used for the three study catchments. This confirms that differences in catchment characteristics can also influence stormwater quality characteristics. To support understanding, Figure 5 shows the rainfall characteristics that dictate stormwater quality responses, temporal patterns of rainfall events and pollutographs (marked is a black line) for the events selected from the three clusters.

As evident in Figure 4, water quality variables P89 and P910 have projected towards cluster 1 events for all three catchments. This suggests that cluster 1 events produce relatively high pollutant loads towards the end of a runoff event and are correlated with the total duration vector. However, cluster 1 events have a relatively long duration (Figure 3). Rainfall temporal patterns of cluster 1 events show that high rainfall intensity bursts are located towards the end of the rainfall events (Figure 5(a)).

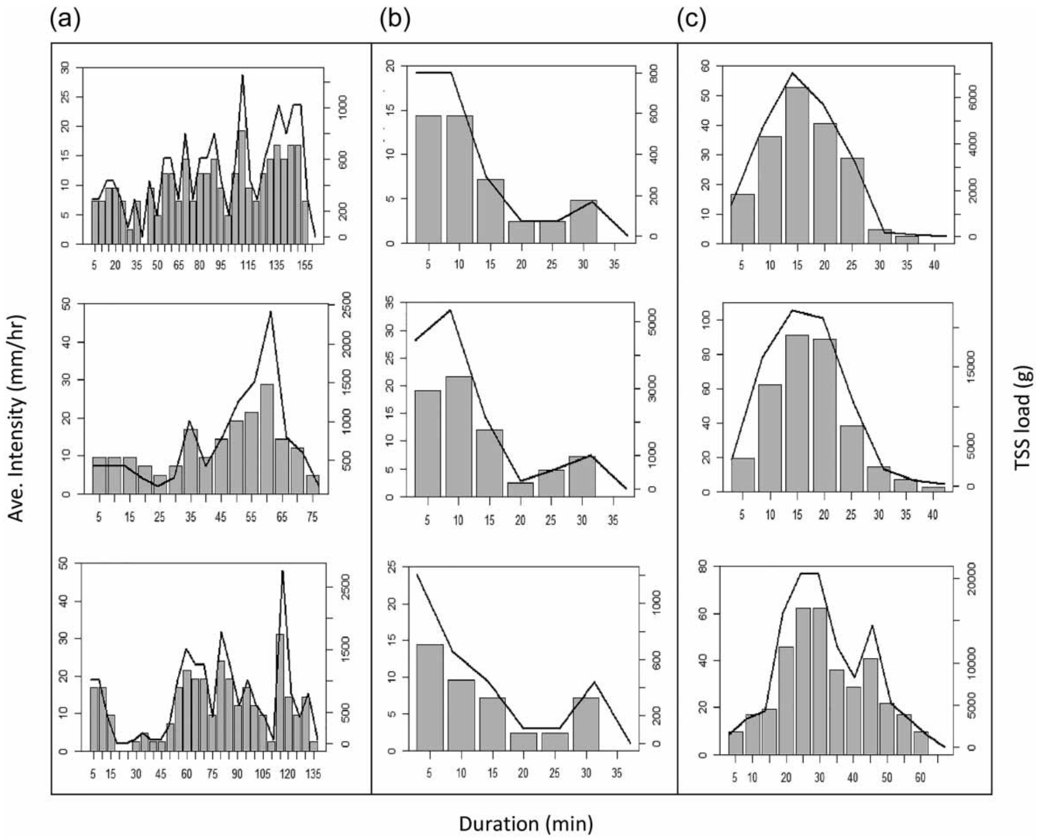


Figure 5 | Examples of rainfall temporal patterns and pollutographs: (a) cluster 1; (b) cluster 2; and (c) cluster 3.

Further, cluster 1 events can produce large volumes of stormwater due to having a long duration and low ADD (Figure 3). This may result in significant runoff and pollutant contributions from pervious surfaces towards the latter part of rainfall events. With respect to the common principles of stormwater treatment, which target frequent (small) events having first flush for treatment, it is difficult to achieve effective treatment of cluster 1 rainfall events (Herngren *et al.* 2005). Hence, the large volume of the latter part of runoff can often bypass treatment systems due to limited storage capacity. The presence of a large volume in stormwater runoff causes a dilution effect that further reduces the overall concentration of pollutants. This finding can be further confirmed by the Box–Whisker plots of event mean concentrations (EMCs) of TSS for three clusters at the study catchments, as shown in Figure 6. Accordingly, cluster 1 events produce relatively lower EMCs compared to cluster 2 and 3 events for the three catchments.

As evident from Figure 4, the P01 variable is projected towards the cluster 2 events in all three catchments. This suggests that cluster 2 events generate highly concentrated pollutant load corresponding to the first 10% of runoff volume and it varies strongly with ADDs. This is a category that produces strong first flush. This is also supported by the higher EMCs in cluster 2 events for the three study catchments (Figure 6). As seen in Figure 5(b), cluster 2 events demonstrate relatively high intensity during initial bursts of rainfall with low duration, even though they show negative scores for average intensity and total duration (Figure 4).

This demonstrates that these events are short duration and have high concentration of wash-off at the initial fraction of runoff due to relatively high intensity. However, as noted by Egodawatta *et al.* (2009), the intensities in the initial part of rainfall may not warrant complete wash-off. As they noted, intensities up to 20 mm/h only can wash-off 20% of the total build-up, even if they continue for a longer duration. This suggests the dominance of an easily washable fraction of pollutants in cluster 2 runoff events. Such pollutants can accumulate progressively with the increasing ADDs and are easily mobilised due to runoff created during the initial part of the rainfall events. Such events can be treated in common treatment systems due

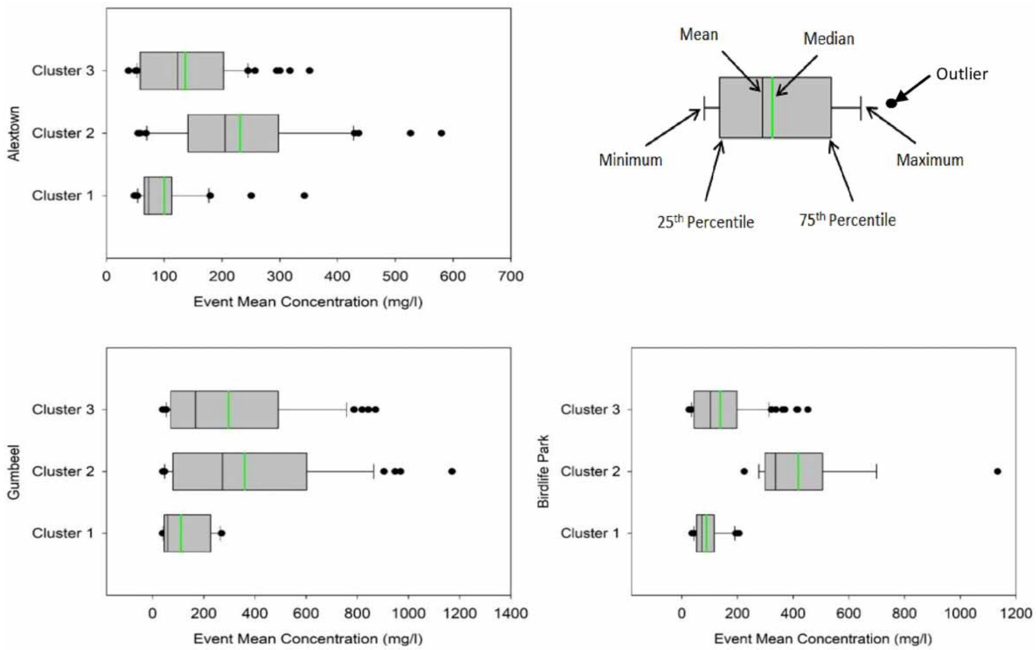


Figure 6 | Comparison of EMC of TSS for three clusters at the study catchments.

to their low runoff volumes and high first-flush effect. However, these events would not constitute target events for stormwater treatment design due to their comparatively low overall pollutant loads (Egodawatta *et al.* 2009; Liu *et al.* 2012).

Moreover, as evident in Figure 4, P12, P23, P34, P45, P56, P67 and P78 variables are projected towards the cluster 3 events for the three study catchments. This suggests that cluster 3 events are relatively widespread in terms of their stormwater quality responses and produce relatively high wash-off loads, corresponding to 10–80% runoff volume. Cluster 3 objects are also correlated to the average intensity vector. This suggests these events are of relatively high intensity. As evident in Figure 5(c), the rainfall temporal pattern shows a bell-shape variation with relatively high-intensity bursts placing in the mid portion of rainfall events. Furthermore, events show negative to positive scores associated with the total duration vector, suggesting that the events can have variable rainfall durations. As evident in Figure 4, events with durations comparable to cluster 2 events (events in positive PC2) produce high loads of wash-off pollutants within the first 40% of runoff and can be categorised as first-flush events. Such events can be considered as target events for stormwater treatment. Cluster 3 events with durations comparable to cluster 1 events (events in negative PC2) produce high loads of pollutants in 40–80% of runoff volumes, which cannot constitute the first-flush event.

These characteristics can be confirmed by the fact that median EMCs for cluster 3 events were found between those of cluster 1 and 2 events (Figure 6). Additionally, ADD of cluster 3 events varies between the lower ADD of cluster 1 to higher ADD cluster 2 events (Figure 3). Irrespective of rainfall duration, all cluster 3 events create rainfall intensity influenced wash-off. As Egodawatta *et al.* (2009) noted, these events have rainfall intensities that can mobilise 20–50% of pollutants deposited on road surfaces.

In summary, cluster 1 events (low average intensity and long duration) would account for wash-off loads corresponding to 80–100% runoff volume; cluster 2 events (low average intensity and short duration) would account for wash-off loads corresponding to the first 10% of runoff volume; and cluster 3 events (high average intensity short duration) would account for wash-off loads corresponding to 10–80% runoff volume. The threshold limits identified for three rainfall clusters in this study are slightly different from the results noted by Liu *et al.* (2012). The difference can be due to variations in variables used for analysis. Additionally, rainfall event selection criteria and rainfall classification method can influence the stormwater quality responses.

4. CONCLUSIONS

Pollutant load corresponding to each 10% stormwater runoff volume was analysed for a large number of rainfall events to understand the changes to stormwater quality in response to different rainfall characteristics. As such, the study outcomes primarily showed that rainfall events clustered based on intensity and duration can produce distinctly different stormwater quality responses. The classification techniques developed in this study would enable to classify the new event in one of the three clusters that would further enable the selection of rainfall events accounting for the stormwater quality responses in order to design stormwater pollution mitigation strategies. The rainfall events that have relatively high intensity in the latter part and comparatively long duration were found to produce a higher volume of runoff with relatively uniform and low EMCs of TSS. Those events with relatively high intensity in the initial part and comparatively short duration would produce small runoff volume and relatively high EMCs of TSS due to effects of first flush. Furthermore, rainfall events with relatively high intensity in the placed in the mid portion of rainfall events and comparatively short duration could produce variable EMCs of TSS with first-flush effect. The study also identified the average rainfall intensity as the most important independent rainfall characteristics other than ADDs in the wash-off of pollutants, and such rainfall events with high average intensity are capable of mobilising and transporting higher loads of pollutants into stormwater runoff.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions from Bureau of Meteorology (BOM) for providing historical rainfall data for this study.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alias, N. B. 2013 *First Flush Behaviour in Urban Residential Catchments*. *Ph.D.*, Queensland University of Technology, Brisbane, Queensland, Australia.

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.

Batalha, M. S., Barbosa, M. C., Faybishenko, B. & Van Genuchten, M. T. 2018 Effect of temporal averaging of meteorological data on predictions of groundwater recharge. *Journal of Hydrology and Hydromechanics* **66** (2), 143–152. https://doi.org/10.1515/johh-2017-0051.

Bro, R. & Smilde, A. K. 2014 Principal component analysis. *Analytical Methods* **6** (9), 2812–2831. https://doi.org/10.1039/C3AY41907J.

Chaudhary, S., Chua, L. H. & Kansal, A. 2022 Event mean concentration and first flush from residential catchments in different climate zones. *Water Research* **219**, 118594. https://doi.org/10.1016/j.watres.2022.118594.

Chowdhury, A. & Egodawatta, P. 2022 Automatic model calibration of combined hydrologic, hydraulic and stormwater quality models using approximate Bayesian computation. *Water Science & Technology* **86** (2), 321–332. doi: 10.2166/wst.2022.207.

Decsi, B., Ács, T., Jolánkai, Z., Kardos, M. K., Koncsos, L., Vári, Á. & Kozma, Z. 2022 From simple to complex – comparing four modelling tools for quantifying hydrologic ecosystem services. *Ecological Indicators* **141**, 109143. https://doi.org/10.1016/j.ecolind.2022.109143.

Egodawatta, P., Thomas, E. & Goonetilleke, A. 2007 Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. *Water Research* **41** (13), 3025–3031.

Egodawatta, P., Thomas, E. & Goonetilleke, A. 2009 Understanding the physical processes of pollutant build-up and wash-off on roof surfaces. *Science of the Total Environment* **407** (6), 1834–1841. https://doi.org/10.1016/j.scitotenv.2008.12.027.

Emmanuel, I., Andrieu, H., Leblois, E. & Flahaut, B. 2012 Temporal and spatial variability of rainfall at the urban hydrological scale. *Journal of Hydrology* **430**, 162–172.

Gong, Y., Liang, X., Li, X., Li, J., Fang, X. & Song, R. 2016 Influence of rainfall characteristics on total suspended solids in urban runoff: a case study in Beijing, China. *Water* **8** (7), 278. https://doi.org/10.3390/w8070278.

Han, Y. & Bu, H. 2023 The impact of climate change on the water quality of Baiyangdian Lake (China) in the past 30 years (1991–2020). *Science of the Total Environment* **870**, 161957. https://doi.org/10.1016/j.scitotenv.2023.161957.

Herngren, L., Goonetilleke, A. & Ayoko, G. A. 2005 Understanding heavy metal and suspended solids relationships in urban stormwater using simulated rainfall. *Journal of Environmental Management* **76**, 149–158.

Jain, A. K. 2010 Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31** (8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.

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), 220–228. https://doi.org/10.1016/j.watres.2007.07.001.

King, R., Miara, A., Glaws, A. & Macknick, J. 2021 *Improving Short Term Predictability of Hydrologic Models with Deep Learning*. Report Number: AI4ESP-1068. Office of Scientific and Technical Information, United States. https://doi.org/10.2172/1769722.

Liu, Y. & Liu, L. 2016 *Rainfall Feature Extraction Using Cluster Analysis and its Application on Displacement Prediction for a Cleavage – Parallel 2 Landslide in the Three Gorges Reservoir Area 3*.

Liu, A., Goonetilleke, A. & Egodawatta, P. 2012 Taxonomy for rainfall events based on pollutant wash-off potential in urban areas. *Ecological Engineering* **47**, 110–114. https://doi.org/10.1016/j.ecoleng.2012.06.008.

Liu, A., Guan, Y., Egodawatta, P. & Goonetilleke, A. 2016 Selecting rainfall events for effective water sensitive urban design: a case study in Gold Coast City, Australia. *Ecological Engineering* **92**, 67–72.

Mahbub, P., Ayoko, G. A., Goonetilleke, A., Egodawatta, P. & Kokot, S. 2010 Impacts of traffic and rainfall characteristics on heavy metals build-up and wash-off from urban roads. *Environmental Science & Technology* **44** (23), 8904–8910.

MikeUrban 2019 *Mike Urban Collection System – User Guide*. Danish Hydraulic Institute. Available from: https://manuals.mikepoweredbydhi.help/2019/Cities/CollectionSystem.pdf.

Mujere, N. & Moyce, W. 2018 Climate change impacts on surface water quality. In: *Hydrology and Water Resource Management: Breakthroughs in Research and Practice*, pp. 97–115. IGI Global, Hershey, PA, USA. doi: 10.4018/978-1-5225-3427-3.ch004.

Qi, Y., Chan, F. K. S., Thorne, C., O'Donnell, E., Quagliolo, C., Comino, E., Pezzoli, A., Li, L., Griffiths, J. & Sang, Y. 2020 Addressing challenges of urban water management in Chinese sponge cities via nature-based solutions. *Water* **12** (10), 2788. https://doi.org/10.3390/w12102788.

Seifollahi-Aghmiuni, S., Kalantari, Z., Egidi, G., Gaburova, L. & Salvati, L. 2022 Urbanisation-driven land degradation and socioeconomic challenges in peri-urban areas: insights from Southern Europe. *AMBIO A Journal of the Human Environment* **51** (6), 1446–1458. https://doi.org/10.1007/s13280-022-01701-7.

Stein, L., Clark, M. P., Knoben, W. J., Pianosi, F. & Woods, R. A. 2021 How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. *Water Resources Research* **57** (4), 1–21. https://doi.org/10.1029/2020WR028300.

Tian, F., Hu, H., Sun, Y., Li, H. & Lu, H. 2019 Searching for an optimized single-objective function matching multiple objectives with automatic calibration of hydrological models. *Chinese Geographical Science* **29**, 934–948. https://doi.org/10.1007/s11769-019-1068-5.

Wang, J., Xue, F., Jing, R., Lu, Q., Huang, Y., Sun, X. & Zhu, W. J. S. 2021 Regenerating sponge city to sponge watershed through an innovative framework for urban water resilience. *Sustainability* **13** (10), 5358. https://doi.org/10.3390/su13105358.

Wijesiri, B., Liu, A., Gunawardana, C., Hong, N., Zhu, P., Guan, Y. & Goonetilleke, A. 2018 Influence of urbanisation characteristics on the variability of particle-bound heavy metals build-up: a comparative study between China and Australia. *Environmental Pollution* **242**, 1067–1077. https://doi.org/10.1016/j.envpol.2018.07.123.

Wijesiri, B., Bandala, E., Liu, A. & Goonetilleke, A. J. S. 2020a A framework for stormwater quality modelling under the effects of climate change to enhance reuse. *Sustainability* **12** (24), 10463. https://doi.org/10.3390/su122410463.

Wijesiri, B., Liu, A. & Goonetilleke, A. 2020b Impact of global warming on urban stormwater quality: from the perspective of an alternative water resource. *Journal of Cleaner Production* **262**, 121330. https://doi.org/10.1016/j.jclepro.2020.121330.

Xiong, J., Zheng, Y., Zhang, J., Quan, F., Lu, H. & Zeng, H. 2023 Impact of climate change on coastal water quality and its interaction with pollution prevention efforts. *Journal of Environmental Management* **325**, 116557. https://doi.org/10.1016/j.jenvman.2022.116557.

Yan, H., Zhu, D. Z., Loewen, M. R., Zhang, W., Liang, S., Ahmed, S., Duin, B. V., Mahmood, K. & Zhao, S. 2023 Impact of rainfall characteristics on urban stormwater quality using data mining framework. *Science of the Total Environment* **862**, 160689. https://doi.org/10.1016/j.scitotenv.2022.160689.

Zhang, H., Wang, Z., Zhang, Y., Ding, M. & Li, L. 2015 Identification of traffic-related metals and the effects of different environments on their enrichment in roadside soils along the Qinghai–Tibet highway. *Science of the Total Environment* **521**, 160–172.

Zhang, K., Manuelpillai, D., Raut, B., Deletic, A. & Bach, P. M. 2019 Evaluating the reliability of stormwater treatment systems under various future climate conditions. *Journal of Hydrology* **568**, 57–66. https://doi.org/10.1016/j.jhydrol.2018.10.056.

First received 19 January 2023; accepted in revised form 12 April 2023. Available online 24 April 2023