The development of a novel approach for assessment of the first flush in urban stormwater discharges

P. M. Bach, D. T. McCarthy and A. Deletic

ABSTRACT

The management of stormwater pollution has placed particular emphasis on the first flush phenomenon. However, definition and current methods of analyses of the phenomena contain serious limitations, the most important being their inability to capture a possible impact of the event size (total event volume) on the first flush. This paper presents the development of a novel approach in defining and assessing the first flush that should overcome these problems.

The phenomenon is present in a catchment if the decrease in pollution concentration with the absolute cumulative volume of runoff from the catchment is statistically significant. Using data from seven diverse catchments around Melbourne, Australia, changes in pollutant concentrations for Total Suspended Solids (TSS) and Total Nitrogen (TN) were calculated over the absolute cumulative runoff and aggregated from a collection of different storm events. Due to the discrete nature of the water quality data, each concentration was calculated as a flow-weighted average at 2 mm runoff volume increments. The aggregated concentrations recorded in each increment (termed as a 'slice' of runoff) were statistically compared to each other across the absolute cumulative runoff volume. A first flush is then defined as the volume at which concentrations reach the 'background concentration' (i.e. the statistically significant minimum). Initial results clearly highlight first flush and background concentrations in all but one catchment supporting the validity of this new approach. Future work will need to address factors, which will help assess the first flush's magnitude and volume. Sensitivity testing and correlation with catchment characteristics should also be undertaken.

Key words | total nitrogen, total suspended solids, water quality

P. M. Bach (corresponding author)
D. T. McCarthy
A. Deletic
Department of Civil Engineering,
Monash University,
Building 60,
Melbourne 3800, Victoria,
Australia
E-mail: *peterbach@gmail.com*;
david.mccarthy@eng.monash.edu.au;
Ana.Deletic@eng.monash.edu.au

INTRODUCTION

The existence of the first flush phenomenon has implications for the design of stormwater treatment systems in urban catchments around the world (see Kang *et al.* 2008). The occurrence of this phenomenon is characterised by a peak in pollutant concentration in the initial volume of stormwater runoff, which, if known, could lead to more efficient implementation of structural and non-structural water sensitive management applications. Much research has been devoted to the development of a model that can accurately predict pollutant loading and analyses have employed the use of dimensionless cumulative mass pollutant load versus cumulative volume curves coupled with multiple regression statistical analyses to assess the impact of an array of factors (Deletic & Maksimovic 1998; Lee *et al.* 2004). Using this model as a basis, Bertrand-Krajewski *et al.* (1998) proposed that a first flush was existent in a rainfall event if at least 80% of pollutant mass is transported in the first 30% of runoff volume. These values are arbitrary and variations of this definition have been proposed (Wanielista & Yousef 1993; Sansalone & Buchberger 1997). As such, a sound definition needs to be developed.

doi: 10.2166/wst.2010.209

Review of literature, however, shows that various flaws are present in the 'traditional' first flush assessment technique. Investigating individual rainfall events independently from each other showed that a first flush was only existent in a proportion of these events (Saget *et al.* 1996; Deletic 1998). Barco *et al.* (2008) employed this traditional method and showed that a shorter storm with less intense rainfall did not exhibit a strong first flush. The key flaws with the traditional method are outlined below.

(1) In a small event, which does not have enough energy to deplete the accumulated surface load, the pollutant concentrations within the event will be consistently high and as such, the traditional method will not detect a first flush.
(2) In an event with both high initial and end pollutant concentrations (which has been reported by many, e.g. Taylor 2006; McCarthy 2008), the traditional method would not detect a first flush.

This paper investigates the initial performance of a new technique for assessing the existence of a first flush using both Total Suspended Solids (TSS) and Total Nitrogen (TN) data from seven diverse catchments around Melbourne, Australia. This new method is different from the traditional method in a number of ways. Rather than normalised, dimensionless figures, the first flush is characterised in relation to absolute values of runoff discharged, which is more relevant to stormwater treatment practice. All events are analysed simultaneously and the distribution of changing pollutant concentrations throughout the event is assessed in increments of runoff depth. As opposed to an arbitrary definition, statistical tests are used to provide stronger support for the existence of a first flush.

MATERIALS AND METHODS

Data set

Catchment details

Doppler based flow meters (which recorded flows in one-minute timesteps) and standard tipping bucket rainfall gauges were installed at seven urban catchments in Melbourne, Australia. These catchments are described in Table 1 and all have separate storm and sanitary sewers. Narre Warren is the only catchment which has its sewage treated by on-site septic systems as well as cross connections with the stormwater system.

Sampling regime

Samples were collected throughout each event using flow-weighted intervals and were then analysed for TSS and TN concentrations. The 24-bottle autosampler was programmed to increase the flow-weightings between samples as more samples were taken throughout longer events. This was to ensure that samples could be collected throughout longer events despite the equipment's limitation of 24 bottles. Table 1 summarises the number of events monitored for TSS and TN at each site. Whilst all seven catchments contained data sets with TSS data, only four catchments had significant amounts of monitored TN data.

Table 1 | Characteristics of the seven catchments used in this study and number of events with available water quality data used in the analysis

Catchment	Primary land use	Catchment area, A (in ha)*	Effective impervious fraction, EIF†	No. of events with TSS data	No. of events with TN data
Burwood East	Mixed residential & commercial	186	0.24	33	
Mt. Waverley	Commercial	28	0.45	51	19
Monash Roof	Treated aluminium roof	0.05	0.90	31	9
Narre Warren	Rural residential	10	0.11	49	
Richmond	High density residential	89	0.31	45	21
Doncaster	Medium density residential	106	0.25	33	
Glen Waverley	Medium density residential	38	0.21	39	14

*Values based on Francey *et al.* (in press).
†EIF values obtained from data calibration in MUSIC (Model for Urban Stormwater Improvement Conceptualization).

The first flush assessment

Definition of first flush

The first flush is defined as the amount of runoff required to reduce discharged mass pollutant load from initial to background concentration levels. This definition was derived from Kang *et al.* (2006), who differentiated between short- and long-term pollutant sources and the depletion of the former during longer rainfall events.

Overview of methodology

This new technique analyses the changes in concentrations in increments of runoff throughout the event. The following steps were involved in the assessment:

(1) Selection of a suitable increment width of runoff (i.e. the 'slice size')
(2) Calculation of the *slice mean pollutant concentration* ($\bar{C}_{\text{slice}}$), that is, the mean pollutant concentration within the slice and *slice mean flow rate* ($\bar{Q}_{\text{slice}}$), that is, the mean flow rate within the slice throughout each event.
(3) Constructing box and whisker plots to characterise the distribution of site mean pollutant concentrations in each slice for the "*typical event*".
(4) Pooling groups of slices together to identify initial and background concentrations as well as the existence of a first flush and its magnitude.

Determining slice size

The irregular flow-weighting of the autosampler and discrete nature of the water quality data required this step to be performed. Interpolation was used to approximate concentrations at one-minute intervals, thereby transforming the data into a continuous set and allowing for flow-weighted mean concentrations to be calculated. To prevent the results from depending solely on approximated data, each slice in the analysis had to contain at least one water quality sample. Information on effective impervious area was used to convert flow into runoff depths [mm] instead of volumes [L]. Graphs of cumulative runoff depth vs. sample number were consequently created for all events and for each catchment. The suitable slicing increment was selected based on these graphs.

Characterising slices

For each event, data was divided into subsets based on the chosen increment of runoff volume (i.e. the slice size). Average flow-weighted concentrations and flow rates were then calculated for each slice based solely on the flow within this data subset. The following equation was used to determine the flow-weighted slice mean pollutant concentration ($\bar{C}_{\text{slice}}$) for a single slice:

Slice mean pollutant concentration:

$$\bar{C}_{\text{slice}} = \frac{\sum C_i Q_i \Delta t}{\sum Q_i \Delta t} \quad [\text{mg/L}]$$

where C_i is the pollutant concentration at a particular instant [mg/L], Q_i the flow rate at a particular instant [L/sec] and Δt the time between successive flow measurements [sec]. The slice mean flow rate ($\bar{Q}_{\text{slice}}$) for each slice was calculated as an arithmetic mean using the following equation:

Slice mean flow rate:

$$\bar{Q}_{\text{slice}} = \frac{\sum Q_i}{n_{\text{slice}}} \quad [\text{L/sec}]$$

where Q_i is the flow rate at a particular instant [L/sec] and n_{slice} the number of data points contained within the data subset (i.e. the slice).

The resulting output featured mean concentrations for each runoff volume increment in each event. For every catchment, all events were subsequently combined to produce a range of possible concentrations at each volume increment of runoff (with a decreasing number of contributing events for every consecutive slice due to varying event durations and the flow-weighted nature of the skewed sampling regime structure). The distribution of mean concentrations in each slice was subsequently assessed by constructing a box and whisker plot. These statistical descriptors were superimposed onto a scatter plot of slice pollutant concentration (from all events) vs. cumulative runoff depth. Due to the decreasing number of contributing events for slices at greater runoff depths, a minimum cut-off value of 5 events was used to determine the maximum runoff depth on the plot. This value was chosen as it was still deemed suitable for the reasonable calculation of descriptive statistical parameters.

This analysis sought to only determine whether a first flush existed for the most typically occurring event for each catchment and to characterise it by its magnitude (defined here as the magnitude of the difference between initial concentrations and background concentrations) and volume (defined here as the length or depth of runoff for which the first flush occurs). This "*typical event*" is referred to as the most occurring pollutant behaviour in the catchment and is statistically described by the median. As such, understanding what occurs most frequently at the catchment will be advantageous for best management practices.

Calculating first flush volumes

To provide more robust justification for conclusions made about the first flush, the Wilcoxon Rank Sum test was additionally used to assess the difference in medians between various slices. This test is nonparametric and regarded as suitable for realistically assessing a first flush as the result is not influenced by the presence of extreme outliers (which were present in the data, identifiable using the box and whisker plots).

The test was used to group similar slices together with the aim of identifying the magnitude and intensity of a first flush. If concentrations in two consecutive slices were statistically indifferent, their data points were pooled together and tested against the next consecutive slice. This was possible because the Wilcoxon Rank Sum test does not require equal number of data points in each sample. Testing continued, comparing an increasing group size with the single consecutive slice until a statistically different result was encountered. The whole process was repeated beginning with a new slice (where previous testing stopped) through to the last slice (at the 5-event cut-off mark described above).

Box and whisker plots of the identified groups were then created, thereby conceptualising the "*typical event*" as several individual groups of runoff, each having significantly different pollutant concentrations. The background concentration was identified as the median value in the last group. The cumulative runoff volume, which marked the start of this background concentration, was identified as the first flush volume. A final effort was made to further conceptualize the "*typical event*" by statistically testing all data points within the first flush volume against all data points within the background concentration volume and reporting the Wilcoxon Rank Sum probability value denoted as $p_{FF/BG}$.

RESULTS AND DISCUSSION

Slice size

After investigating the available water quality data for the seven catchments, it was decided that a slice size, suitable for the analysis for all catchments, was 2 mm. Plots of cumulative runoff depth versus sample number assisted in this selection process as well as the assessment of flow weighting accuracy.

General overview of the results

Figure 1 shows two graphs from the Narre Warren catchment. In the first part of the analysis, box and whisker plots were created for every 2 mm off runoff (an example of which is shown in the left plot). A scatter of the mean slice concentrations for each event is shown by the dots. Figure 1 (right) shows the latter part of the analysis, which used the Wilcoxon Rank Sum test to group slices of statistically indifferent concentrations together.

In Figure 1 (left), a visible decline in concentration can be observed, supporting the existence of a first flush. An increase in concentration at 18 mm either describes an end flush occurring in several storm events around this volume, or the influence of the possible aforementioned septic cross-connections in the catchment. The increase in concentration at 36 mm was determined from only five data points. Although this visually depicts an end flush, the observation has to be treated with caution as it may be indifferent from adjacent boxes due to the low number of data points used to derive the box & whisker plot. The plots in Figure 1 terminate with 5 contributing events used to calculate the

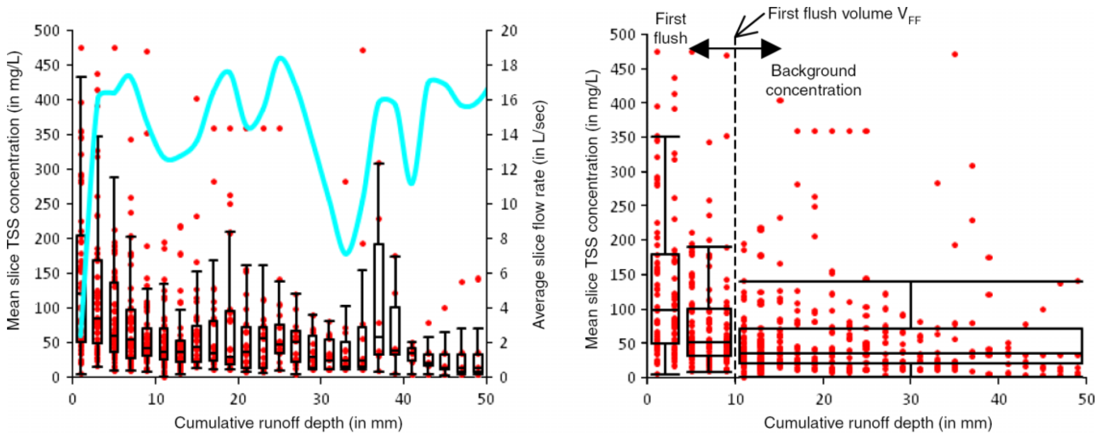

Figure 1 | Result of data slicing (left—the average flow rate within each slice is depicted by the continuous blue line) and grouping of slices (right) using 2 mm slices for TSS data from the Narre Warren catchment.

statistical parameters, which for Narre Warren occurred at 50 mm of runoff.

As indicated in Figure 1 (right), the background concentration and first flush regions are found by locating the point at which the background concentration is reached. The background concentration is shown by the last box in Figure 1 (right).

Further results are presented in Figure 2, which shows the final results for TSS and TN for four catchments (Mt. Waverley, Monash Roof, Richmond and Glen Waverley). Graphs of TSS analysis from Doncaster and Burwood East are not present here (due to space constrains), but the trends for these sites are similar to that of the Monash Roof, which features only two distinctly different boxes.

According to the definition of the first flush stated earlier, results show that all of the catchments exhibit the first flush behaviour for both TSS and TN except at Glen Waverley for TSS (Figure 2, bottom left). However, the characteristics of the first flush are different between catchments, as well as between pollutants.

First flush volumes and magnitudes

For all seven catchments, the first flush volume V_{FF} was determined from the graphs (i.e. Figures 1 and 2). These are shown in Table 2 along with the Wilcoxon Rank Sum probability value $p_{FF/BG}$ between the two regions for both pollutants TSS and TN and the median initial pollutant and background concentrations. For Glen Waverley, it was assumed that there was no first flush region (see Figure 2) and that the corresponding probability value was equal to unity.

Having determined values of V_{FF}, the magnitude and volume of the first flush in each catchment can be characterised conceptually. A high value of V_{FF} indicates a large volume of water transporting the first flush. A low $p_{FF/BG}$ indicates a highly significant difference between the two regions of pollutant concentrations and supports a stronger first flush.

Relating catchments with first flush characteristics

Results for TSS (Table 2) show that Narre Warren's first flush behaviour is similar to that of Richmond (a V_{FF} of 10 mm), despite the large difference in character. This is possibly due to Narre Warren's septic cross connections, which makes the water to be treated 'dirtier' than the older Richmond catchment and hence may explain the strong statistical rejection and higher background concentration. Mt. Waverley, a commercial catchment, shows the largest first flush volume (16 mm) and a probability three orders of magnitude smaller than that of Richmond (therefore, larger first flush magnitude). This may have occurred because the catchment itself is quite old and therefore considered quite 'dirty' (maybe caused by ageing infrastructure). The significantly different land uses may also contribute to this behaviour.

An explanation for why Glen Waverley exhibits no first flush behaviour according to the method could be because the catchment was established quite recently and is not as

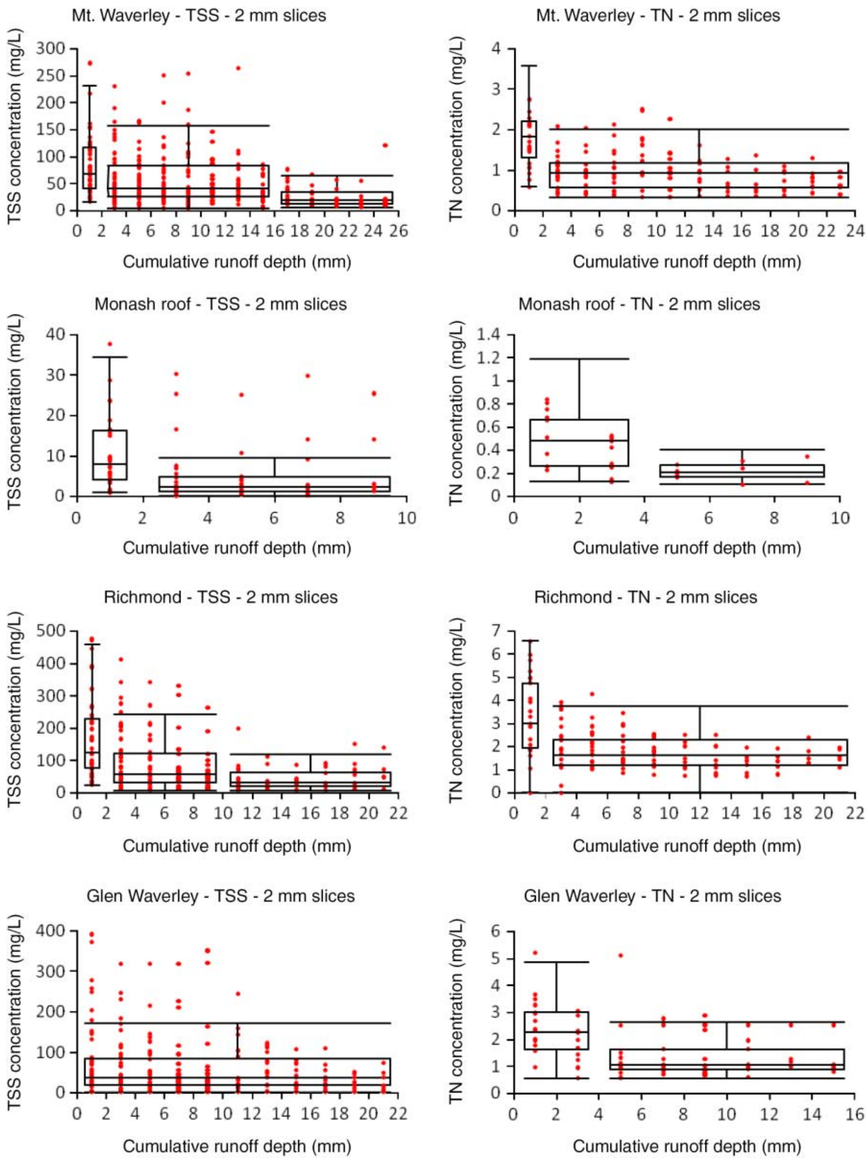

Figure 2 | Grouped slices box and whisker plots for TSS and TN in four catchments.

'dirty' as the other residential catchments (which may have older and poorly maintained infrastructure). Further work will need to be carried out to properly understand why a first flush was not observed in this catchment. Doncaster and Burwood East are residential districts and along with the roof all exhibit a TSS first flush volume of 2 mm. Furthermore, all first flush volumes in Table 2 are multiples of two (naturally because 2 mm is the smallest runoff increment that boxes were grouped into). It however raises the question as to whether the value is dependent on the chosen slice size and whether a smaller slice size would conclude the same results. This is certainly subject for future investigations.

For TN data, making correlations is more difficult due to a smaller number of catchments studied. Both Mt. Waverley and Richmond, two contrasting catchments in terms of land use, show a first flush of TN of similar volume but different magnitude. Glen Waverley on the other hand, having no first flush of TSS, shows a large volume V_{FF} for TN. The Monash roof also exhibits a first flush of equal volume to Glen Waverley, but with the smallest magnitude. It should be mentioned that the noticeably different $p_{FF/BG}$

value for the roof is possibly due more to the lack of events containing TN data than to the catchment itself and should be treated with caution. Nevertheless, conclusions made for TN data differ from those made for TSS suggesting that more factors than those considered may contribute to the first flush's complex behaviour.

Noticeably, all sites (except for the roof and the clean industrial catchment) have similar background concentrations for TSS (all approximately between 29 and 40 mg/L). The trend is not as noticeable for TN results. Nevertheless, comparing the two residential catchments suggests a similar trend. Background concentrations would normally exist due to pervious areas, where soil particles may dislodge during longer or more intense events. The different initial pollutant TSS concentrations for the seven catchments would suggest similar observations as the ones made earlier regarding 'old and dirty' and 'new and clean' catchments.

CONCLUSIONS

An initial assessment of a new model for assessing the existence of a first flush initially showed in a study of seven diverse catchments around Melbourne Australia that a first flush was present in six catchments. A new proposed definition of the phenomenon characterised it as the amount of runoff required to reduce the initial pollutant concentration throughout every event to the background concentration. Statistical analysis used to group similar data subsets (slices) together gave rise to various types of first flush behaviours throughout catchments for TSS characterized by volume (V_{FF}) and magnitude ($p_{FF/BG}$). Qualitative correlation of the first flush parameters with various catchment characteristics however highlighted the complexity of this phenomenon and provided a few aspects of improvement of this method, which deserves a more in-depth investigation.

This paper only focused on the initial trial of a new technique, which involved the use of 2 mm slices on TSS and TN data for seven catchments. Future work on assessing the performance of this model will involve the testing of sensitivity of results to slice sizes, the performance of this model against other pollutant types, the use of data

Table 2 | Summary of calculated first flush volumes, Wilcoxon Rank Sum probabilities $p_{FF/BG}$, initial pollutant and background site concentrations (C_i and C_{bg}) for all seven catchments

	Total suspended solids (TSS)				Total nitrogen (TN)			
Catchment	First flush volume V_{FF} [mm]	Wilcoxon rank sum probability $p_{FF/BG}$	Median C_i [mg/L]*	Median C_{bg} [mg/L]*	First flush volume V_{FF} [mm]	Wilcoxon rank sum probability $p_{FF/BG}$	Median C_i [mg/L]*	Median C_{bg} [mg/L]*
Burwood East	2	1.42×10^{-9}	82	29				
Mt. Waverley	16	4.53×10^{-10}	67	19	2	6.70×10^{-6}	1.8	0.9
Monash Roof	2	3.42×10^{-5}	4	1	4	5.30×10^{-3}	0.3	0.2
Narre Warren	10	1.23×10^{-13}	99	36				
Richmond	10	1.02×10^{-7}	123	30	2	4.48×10^{-5}	3.0	1.6
Doncaster	2	2.36×10^{-7}	66	33				
Glen Waverley	0	1.00	36	36	4	8.91×10^{-5}	2.3	1.0

*Value of initial pollutant concentration C_i and background site concentration C_{bg} are quoted for the "typical event" for the total number of events analysed for each site in Table 1.

involving continuous water quality measurements and correlation analyses to search for factors which can explain the first flush phenomenon. Comparisons will be made between this method and several traditional first flush methods. With a better understanding of the first flush through this model, improvements can be made to current treatment and stormwater harvesting practices towards a more efficient design.

REFERENCES

Barco, J., Papiri, S. & Stenstrome, M. K. 2008 First flush in a combined sewer system. *Chemosphere* **71**(5), 827–833.

Bertrand-Krajewski, J. L., Chebbo, G. & Saget, A. 1998 Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. *Water Res.* **32**(8), 2341–2356.

Deletic, A. 1998 The first flush load of urban surface runoff. *Water Res.* **32**(8), 2462–2470.

Deletic, A. B. & Maksimovic, C. T. 1998 Evaluation of water quality factors in storm runoff from paved areas. *J. Environ. Eng.* **124**(9), 869–879.

Francey, M., Fletcher, T. D., Deletic, A. & Duncan, H. (in press). New insights into the quality of urban stormwater of South Eastern Australia. *J. Environ. Eng. (ASCE)*.

Kang, J.-H., Kayhanian, M. & Stenstrom, M. K. 2006 Implications of a kinematic wave model for first flush treatment design. *Water Res.* **40**(20), 3820–3830.

Kang, J.-H., Kayhanian, M. & Stenstrom, M. K. 2008 Predicting the existence of stormwater first flush from the time of concentration. *Water Res.* **42**(1–2), 220–228.

Lee, H., Lau, S.-L., Kayhanian, M. & Stenstrom, M. K. 2004 Seasonal first flush phenomenon of urban stormwater discharges. *Water Res.* **38**(19), 4153–4163.

McCarthy, D. T. 2008 A traditional first flush assessment of *E. coli* in urban stormwater runoff. *11th International Conference on Urban Drainage*. Edinburgh, Scotland, UK.

Saget, A., Chebbo, G. & Bertrand-Krajewski, J.-L. 1996 The first flush in sewer systems. *Water Sci. Technol.* **33**(9), 101–108.

Sansalone, J. J. & Buchberger, S. G. 1997 Partitioning and first flush of metals in urban roadway storm water. *J. Environ. Eng.* **123**(2), 134–143.

Taylor, G. D. 2006 *Improved Effectiveness of Nitrogen Removal in Constructed Stormwater Wetlands*. Civil Engineering Department. Monash University, Australia.

Wanielista, M. & Yousef, Y. 1993 *Stormwater Management*. John Wiley & Sons, New York, NY, pp. 1–579.