University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2013

Prediction of long-term urban stormwater loads at single sites

Daniel May

University of Wollongong, dbm05@uow.edu.au

Muttucumaru Sivakumar

University of Wollongong, siva@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Prediction of long-term urban stormwater loads at single sites

Abstract

Long-term total phosphorus loads from 17 urban catchments in the USA were predicted using five different measures of central tendency defining site mean concentration (arithmetic mean, geometric mean, median, harmonic mean and flow-weighted mean). Overall, the flow-weighted mean concentration produced the most accurate predictions of long-term loads. The geometric mean produced the second most accurate predictions. Along with the median and harmonic mean, the geometric mean predicted long-term load relatively well at most catchments exhibiting negative correlations between event mean concentration and total event runoff depth. However, they significantly underestimated long-term load at catchments exhibiting a positive correlation between these variables. Better estimates of long-term load at these two catchments were produced using the flow-weighted mean and arithmetic mean. However, the arithmetic mean tended to overestimate long-term load at the remainder of catchments.

Keywords

stormwater, loads, prediction, long, sites, term, single, urban

Disciplines

Engineering | Science and Technology Studies

Publication Details

May, D. & Sivakumar, M. (2013). Prediction of long-term urban stormwater loads at single sites. *Proceedings of the Institution of Civil Engineers: Water Management*, 166 (2), 81-92.

Prediction of long-term urban stormwater loads at single sites

Daniel May BE, PhD

Sustainable Water and Energy Research Group, School of Civil, Mining and Environmental Engineering, University of Wollongong, Wollongong, Australia

Muttucumaru Sivakumar ME, PhD, MIEAust; CPEng

Associate Professor, Sustainable Water and Energy Research Group (SWERG), School of Civil, Mining and Environmental Engineering, University of Wollongong, Wollongong, Australia

Long-term total phosphorus loads from 17 urban catchments in the USA were predicted using five different measures of central tendency defining site mean concentration (arithmetic mean, geometric mean, median, harmonic mean and flow-weighted mean). Overall, the flow-weighted mean concentration produced the most accurate predictions of long-term loads. The geometric mean produced the second most accurate predictions. Along with the median and harmonic mean, the geometric mean predicted long-term load relatively well at most catchments exhibiting negative correlations between event mean concentration and total event runoff depth. However, they significantly underestimated long-term load at catchments exhibiting a positive correlation between these variables. Better estimates of long-term load at these two catchments were produced using the flow-weighted mean and arithmetic mean. However, the arithmetic mean tended to overestimate long-term load at the remainder of catchments.

Notation

C_a	arithmetic mean
C_{fw}	flow-weighted mean
C_g	geometric mean
C_h	harmonic mean
EMC	event mean concentration
L_a	accumulated pollutant load (kg/ha)
L_m	maximum pollutant load accumulating on catchment surface (kg/ha)
L_T	total load at a catchment (mg)
L_t	total pollutant mass washed from catchment surface (kg/ha)
L_0	initial pollutant mass on catchment surface (kg/ha)
N	number of catchments
n	number of sampled storms
TRFD	total runoff depth (mm)
t_s	time since the previous storm event (days)
V_i	total volume of runoff for storm i (mm)
V_T	total runoff volume at a catchment (l)
w	empirical washoff coefficient (per mm)
x	accumulation rate (per day)
y_j	observed long-term load for catchment j (mg)
$\hat{y}_j$	predicted long-term load for catchment j (mg)

1. Introduction

Eutrophication is a major problem in surface waters throughout the world. Urban stormwater runoff is a significant contributor of nutrient loads, which are the major cause of eutrophication in receiving waters (Taylor *et al.*, 2005). The nightly algal respiration associated with algal blooms may deplete dissolved oxygen concentrations to levels that are no longer able to support aquatic life (Paerl, 2006; Thomann and Mueller, 1987). This problem is further exacerbated when the algae die off and are converted back

to their basic constituents. The decreased diversity of aquatic life (Ferreira *et al.*, 2007; Hilton *et al.*, 2006) may then adversely affect the balance of the ecosystem (Burian *et al.*, 2001; Ferreira *et al.*, 2011). As a result, it has become important to quantify nutrient exports into surface waters.

The response time of many surface waters to the impact of nutrient exports may be of the order of several years (Huber, 1992). Therefore, long-term water monitoring programmes are potentially required to adequately quantify nutrient exports. However, attaining water quality data is both time-consuming and expensive (Brezonik and Stadelmann, 2002; Driver and Tasker, 1990; Mourad *et al.*, 2005a, 2006; Sliva and Williams, 2001). This has created the need to forecast long-term nutrient loads.

Process-based models based upon fundamental physical relationships have wider domains of applicability than statistical models (Maier and Dandy, 1996). The most common group of process-based models used in urban stormwater quality applications are buildup–washoff models, which replicate the two main processes influencing the accumulation and transport of pollutants in urban stormwater systems. Exponential buildup is the most common function used to define the accumulation of pollutants on impervious surfaces during dry periods (Wang *et al.*, 2011). Park *et al.* (2008) define the exponential buildup function used in the stormwater management model as

$$1. \quad L_a = L_m [1 - \exp(-xt_s)]$$

where L_a is the accumulated pollutant load (kg/ha), L_m is maximum pollutant load accumulating on the catchment surface

(kg/ha), x is the accumulation rate (per day) and t_s is the time since the previous storm event (days).

Most washoff algorithms assume that pollutant removal from the catchment surface is proportional to the amount of pollutant on the catchment surface (Zoppou, 2001). This gives rise to an exponential washoff function

$$2. \quad L_t = L_0 [1 - \exp(-w\text{TRFD})]$$

where L_t is the total pollutant mass washed from the catchment surface (kg/ha), L_0 is the initial pollutant mass on the catchment surface (kg/ha), TRFD is total runoff depth (mm) and w is an empirical washoff coefficient (per mm).

A variation of this washoff algorithm is used in the stormwater management model (Park *et al.*, 2008). Despite the ability of buildup–washoff models to replicate the dominant processes influencing urban stormwater quality, they are subject to several limitations. Potentially significant processes are ignored, including atmospheric scavenging, pervious area erosion, subsurface contribution (Butcher, 2003) and instream suspension (Corbett *et al.*, 1997). Furthermore, most buildup–washoff models incorrectly assume that the majority of pollutants accumulated on a catchment surface are removed during typical storm events (Vaze and Chiew, 2002). This effectively hinders the calibration of such models.

The difficulties associated with the application of complex process-based models have created the need for simple statistical models. Regression models and rating curves have been applied in the past (Brodie and Dunn, 2010; Davis and Birch, 2010; Jewell and Adrian, 1982; Maniquiz *et al.*, 2010; Mourad *et al.*, 2005b, 2005c; Vaze and Chiew, 2003). However, the accuracy of such models is often restricted by a lack of correlation between stormwater concentrations and commonly used explanatory variables such as runoff volume (Athayde *et al.*, 1983; Huber, 1992). Furthermore, an analysis by Mourad *et al.* (2006) suggested it was no more advantageous using a regression model over a site mean concentration (SMC) model. SMC models are simple, generally requiring less sampling data to calibrate than more complex buildup–washoff or regression models (Mourad *et al.*, 2005a, 2005b, 2005c; Obropta and Kardos, 2007).

A number of different measures of central tendency may be used to define SMC at a given catchment: arithmetic mean (Brodie and Dunn, 2009; Chow *et al.*, 2011; Mourad *et al.*, 2005a); geometric mean (Brodie and Dunn, 2009); lognormal mean (Mourad *et al.*, 2005a); median (Athayde *et al.*, 1983; Leecaster *et al.*, 2002; Mourad *et al.*, 2005a); flow-weighted mean (Francey *et al.*, 2010; Mourad *et al.*, 2005a, 2005b, 2005c). Mourad *et al.* (2005a) compared the ability of four measures of central tendency defining SMC (arithmetic mean, lognormal mean, median and flow-weighted mean) to predict long-term biological

oxygen demand, chemical oxygen demand and suspended solids loads at three catchments. The results varied between catchments according to the data used. Despite this, it was found that the median generally produced long-term estimates of load that did not vary significantly, according to the number of storms used to generate the estimate. Furthermore, it was observed that the long-term load estimates generated using the arithmetic mean had a significant bias compared with the observed load. However, the analysis was limited due to the minimal number of catchments analysed. The main objective of the current study was to identify the optimum measure of central tendency defining SMC to predict long-term total phosphorus loads at 17 urban catchments in the USA.

The 17 catchments analysed in the current work were sampled during the US Nationwide Urban Runoff Program (NURP) and data from the program were first presented by Athayde *et al.* (1983). Subsequently, regression analysis (Driver and Tasker, 1990), additional studies (May and Sivakumar, 2003, 2008, 2009a, 2009b, 2009c, 2009d) and a comparative study (Park *et al.*, 2009) have been carried out using the NURP data.

2. Materials and methods

The objective of this study was to determine the measure of central tendency defining SMC that most accurately predicts long-term total phosphorus loads in urban stormwater runoff. The water quality data used in this study were collected by the United States Geological Survey and the United States Environmental Protection Agency from catchments located throughout the USA as part of NURP. The data were converted into electronic format by the Cahaba/Warrier Student Chapter of the American Water Resources Association (AWRA, 1998) and used in the current study. Sampling was undertaken at the pour point of each catchment, using both discrete and composite samples. Flow records at each catchment were typically collected every 5 min, although they were occasionally collected at other regular intervals.

A total of 17 catchments from the NURP study were analysed, each containing a minimum of 12 sampled storm events. The locations of the catchments and the catchment characteristics are presented in Table 1 and Table 2 respectively. The majority of the analysed catchments were small and highly impervious with large areas of residential and/or commercial land use. Only one catchment (catchment 6) had a large portion of non-urban land use. Despite the low concentrations of total phosphorus associated with storms from this catchment, it was analysed along with the remainder of catchments.

Site mean concentrations were derived from event mean concentration (EMC) data using five measures of central tendency

- (a) arithmetic mean
- (b) geometric mean
- (c) median

Metropolitan area	Mean annual rainfall: mm	Mean annual average temperature:		Catchments
		°F	°C	
Salt Lake City, Utah	432	50	10.0	1
Bellevue, Washington	945	51	10.6	2, 3
Miami, Florida	1500	74	23.3	4–6
St. Paul, Minnesota	737	46	7.8	7–11
Milwaukee, Wisconsin	738	42	5.6	12–17

Table 1. Locations of urban catchments

Catchment	Drainage area: ha	Impervious area: %	Land use: %		
			Commercial	Residential	Non-urban
1	26	52.4	8.0	83.0	9.0
2	39	36.1	6.8	89.8	3.4
3	36	30.9	5.7	90.7	3.6
4	8	97.9	97.9	0.0	2.1
5	16	43.9	0.0	100.0	0.0
6	23	36.0	40.0	0.0	60.0
7	34	11.0	1.0	87.2	11.8
8	85	22.0	0.0	83.0	17.0
9	31	70.0	84.6	9.6	5.8
10	122	35.0	19.5	79.5	1.0
11	57	29.0	3.5	96.5	0.0
12	16	44.0	0.0	100.0	0.0
13	5	99.4	100.0	0.0	0.0
14	5	98.0	100.0	0.0	0.0
15	13	30.0	0.0	100.0	0.0
16	10	76.8	80.0	20.0	0.0
17	18	80.6	70.0	30.0	0.0

Table 2. Characteristics of urban catchments

(d) harmonic mean

(e) flow-weighted mean.

The SMC may be used in the determination of design storms for the construction of best management practices. However, as in this study, predictions of SMC are often used to quantify long-term loads entering receiving waters such as lakes. This may then allow determination of the risk of eutrophication.

The most commonly used measure of central tendency is the arithmetic mean, which may be calculated from

$$3. \quad C_a = \frac{\sum_{i=1}^n EMC_i}{n}$$

where EMC_i is the event mean concentration for storm i (mg/l) and n is the number of sampled storms.

One of the main advantages of the arithmetic mean is that it enables the calculation of associated confidence intervals using an estimate of standard deviation. However, it is commonly known that stormwater constituents are lognormally distributed (Athayde *et al.*, 1983; Duncan, 1999; Van Buren *et al.*, 1997). As a result, the arithmetic mean is likely to overestimate SMC and produce a lower confidence interval bound that is less than zero. This has led to the use of the geometric mean to derive estimates of SMC. This is because the geometric mean is equivalent to the arithmetic mean of logarithmically transformed data raised to the base of the logarithm. Stated otherwise, it provides an accurate measure of central tendency associated with lognormal distributions. The geometric mean C_g may be calculated from

$$4. \quad \log C_g = \frac{\sum_{i=1}^n \log EMC_i}{n}$$

The advantage of the geometric mean is that it is not biased by large, potentially outlying EMCs. Furthermore, an associated standard deviation may be calculated on logarithmically transformed data and used to construct intuitively correct confidence intervals, with lower bounds always greater than zero.

The median is another measure of central tendency that may be used to estimate the centrality of lognormal distributions. Once again, unlike the arithmetic mean, the median is not biased by large, potentially outlying EMCs, making it a representative estimate of a typical value of concentration at a given site.

The harmonic mean C_h can also be used to derive an SMC estimate for a given site; it is given by

$$5. \quad \frac{1}{C_h} = \frac{1}{n} \sum_{i=1}^n \frac{1}{EMC_i}$$

Once again, the harmonic mean is not biased by large, potentially outlying EMCs at a given site. This may enhance its ability to calculate the SMC at catchments with EMCs following lognormal distributions. However, confidence intervals may not be generated when using the harmonic mean.

Potentially the most interesting measure of central tendency defining SMCs used to derive long-term loads is the flow-weighted mean. The flow-weighted mean concentration C_{fw} is generated from

$$6. \quad C_{fw} = \frac{\sum_{i=1}^n EMC_i V_i}{\sum_{i=1}^n V_i}$$

where V_i is the total volume of runoff for storm i (mm).

The flow-weighted mean is a valid approach, particularly in situations where a significant correlation exists between EMC and runoff volume. However, confidence intervals are more difficult to determine for the flow-weighted mean because no universal methodology exists to calculate the associated standard deviation (Mourad *et al.*, 2005a).

At each catchment, two sets of analyses were undertaken. Leave-one-out analyses were used to determine how well the various measures of central tendency estimated long-term load at well-sampled catchments. Single storm events were sequentially isolated to be used for validation, with the remaining storms used to derive an estimate of SMC. This process was repeated until

every storm event from a given catchment had been used for validation. Secondly, a series of cross-validation analyses was used to determine how well the measures of central tendency performed at catchments with limited sampling. Storm events from a given catchment were randomly allocated into two data sets, each containing approximately 50% of the data. One set of data was initially used for calibration (i.e. calculating SMC) and the other used for validation. This process was then repeated, using different sets for calibration and validation.

The ability of the measures of central tendency defining SMC to predict long-term loads was then evaluated. Observed event loads were derived by multiplying EMC with event runoff volume. Observed long-term loads were then calculated at each catchment, by summing all the observed event loads. Similarly, predicted long-term loads were also calculated at each catchment, by summing the predicted event loads. Predicted event loads were derived using both leave-one-out and cross-validation analyses. During the leave-one-out analysis, an SMC estimate was obtained for each validation storm event, using all storms apart from the validation storm. This predicted SMC was then multiplied by the runoff volume associated with the validation storm to derive a predicted event load for the validation storm. This process was repeated until each storm had been used for validation. In this manner, predicted event loads were derived for each monitored storm event at a given catchment. During the cross-validation analysis, an SMC estimate was obtained using storm events in the calibration set. Predicted event loads were then obtained by multiplying the derived SMC by each storm runoff volume associated with storms in the validation set. This process was then repeated using different sets for calibration and validation. All predicted event loads were then summated to obtain an estimate of predicted long-term load at the catchment of interest.

The observed and predicted long-term loads were then entered into Equation 7 to generate an average absolute percentage error (AAPE) associated with long-term load predictions for each measure of central tendency

$$7. \quad AAPE = 100 \times \frac{\sum_{j=1}^N (|y_j - \hat{y}_j|/y_j)}{N}$$

where y_j is the observed long-term load for catchment j (kg), $\hat{y}_j$ is the predicted long-term load for catchment j (kg) and N is the number of catchments. The AAPE was used to assist in the identification of the most appropriate measure of central tendency defining SMC to predict long-term total phosphorus loads at urban catchments.

3. Results

Five measures of central tendency defining SMCs were used to generate estimates of long-term total phosphorus loads from

Figure 1. Long-term load predictions using the arithmetic mean

Figure 2. Long-term load predictions using the flow-weighted mean

stormwater runoff at 17 urban catchments. The long-term load predictions using each measure of central tendency are presented in Figures 1 to 5. The absolute percentage errors associated with the prediction of long-term loads at each catchment were also calculated, and then averaged. The averaged errors are presented in Figure 6. In general, using a flow-weighted mean concentration produced the most accurate predictions of long-term loads. The second most accurate predictions of long-term loads were derived using a geometric mean, while the least accurate predictions were associated with the median, harmonic mean and arithmetic mean.

Theoretically, when all available data from the catchment of interest are used to construct the SMC using the flow-weighted mean, it will produce a perfect prediction of the long-term load. Equation 6 may be rewritten as follows

$$8. \quad C_{fw} = \frac{L_T}{V_T}$$

where L_T is the total load at a catchment (mg) and V_T is the total runoff volume at a catchment (l). Rearranging Equation 8 yields

Figure 3. Long-term load predictions using the geometric mean

Figure 4. Long-term load predictions using the harmonic mean

$$9. \quad L_T = C_{fw} V_T$$

which is identical to the equation used for the calculation of long-term load. However, when all available data are not used to construct the SMC using a flow-weighted mean, perfect prediction of long-term load will generally not eventuate. In particular, when small data sets are used, calibration of the flow-weighted mean may be biased by large, potentially outlying EMCs. This observation was supported by the results of an analysis conducted by Mourad *et al.* (2005a) who observed that the flow-weighted

mean did not produce consistently accurate predictions of long-term load when limited sampling data were used to generate the estimate of SMC. As a result, it was deemed necessary to analyse the effect of data set size upon the calculation of SMC, using all measures of central tendency.

In general, the number of storm events used to derive the SMCs was found to have a minor influence on the prediction of long-term loads. However, it was observed that the flow-weighted mean produced less accurate predictions of long-term load when fewer storms were used to derive this measure. For this measure

Figure 5. Long-term load predictions using the median

Figure 6. Long-term load errors associated with each measure of central tendency, expressed as an average absolute percentage error (AAPE). AM, arithmetic mean; FWM, flow-weighted mean; GM, geometric mean; HM, harmonic mean; Med, median

of central tendency, it was assumed that potentially outlying values may produce non-representative estimates of SMCs at catchments with limited sampling. This phenomenon did not occur for the other measures of central tendency. Surprisingly, the long-term load estimates were generally more accurate when fewer storms were used to derive the measures of central tendency. However, it was assumed that if even fewer storms were used, more inaccurate predictions of long-term load would eventuate.

Pearson correlation coefficients (Owen and Jones, 1990) between untransformed values of total runoff depth and EMC were then calculated at each catchment. The correlations are presented in Table 3, along with their associated *p* values and the number of sampled storms at each catchment (both totals and average numbers per year). The correlation coefficients were assumed to

Catchment	Pearson correlation	p	Number of storms sampled	
			Total	Average per year
1	−0.35	0.196	15	22
2	−0.01	0.977	31	14
3	−0.25	0.232	24	11
4	−0.44	0.018	28	14
5	−0.21	0.272	28	19
6	−0.29	0.063	42	19
7	0.53	0.033	17	39
8	−0.53	0.060	13	32
9	−0.17	0.463	21	48
10	−0.12	0.617	21	48
11	−0.05	0.812	25	68
12	−0.49	0.103	12	10
13	0.23	0.420	14	6
14	−0.42	0.158	13	57
15	−0.11	0.718	14	8
16	−0.31	0.298	13	11
17	−0.33	0.253	14	61

Table 3. Pearson correlation coefficients between total phosphorus EMC and total runoff depth

assist in the determination of the most appropriate measure of central tendency defining SMC for predicting long-term load. Overall, only two of the analysed catchments had correlations between total event runoff depth and EMC that were significant at the 95% level (catchments 4 and 7). Despite this fact, it was found that the correlation coefficients still influenced the selection of the most applicable measure of central tendency, irrespective of their significance.

The signs of the correlation coefficients between total phosphorus EMC and total event runoff depth influenced the accuracy of the estimates of long-term loads derived using the various measures of central tendency. It can be seen from Table 3 that at the majority of catchments there is a negative correlation between EMC and total event runoff depth. This leads to a situation whereby the accurate predictions of relatively low EMCs produce accurate predictions of large loads. These large loads tend to dominate the calculation of long-term loads. Hence, measures of central tendency that accurately predict lower EMCs generally produce accurate predictions of long-term load at these catchments.

In general, the geometric mean, median and harmonic mean systematically place more weight on the low EMCs, whereas the arithmetic mean tends to place more weight on the larger EMCs. The flow-weighted mean concentration, on the other hand, places more weight on the EMCs associated with large total event runoff depths. As a result, the flow-weighted mean concentration tends to accurately predict long-term load, irrespective of the sign of the correlation between EMC and total event runoff depth.

Two of the catchments analysed in the current study (7 and 13) exhibited a positive correlation between EMC and total event runoff depth. The geometric mean, harmonic mean and median significantly underestimated the long-term loads at these catchments. Surprisingly, the arithmetic mean also underestimated the long-term loads at these catchments, though by not as much as the previously mentioned measures of central tendency. In contrast, the flow-weighted mean concentration produced relatively accurate estimates of long-term load at the two catchments.

The remaining 15 catchments analysed in the study exhibited negative correlations between EMC and total event runoff depth. The arithmetic mean systematically overestimated long-term loads at all these catchments. As expected, the degree of overestimation was very low for the two catchments with very low correlations between EMC and total event runoff depth (catchments 2 and 11). In contrast, the degree of overestimation was very large at catchment 8, which had the highest negative correlation between EMC and total event runoff depth.

The harmonic mean underestimated long-term loads at all but one catchment. This catchment (catchment 8) displayed the highest negative correlation between EMC and total event runoff

depth. Surprisingly, the harmonic mean produced a more accurate estimate of long-term load at this catchment than the flow-weighted mean. However, the large systematic underestimations of long-term loads at the majority of catchments were perceived to reduce the practicality of the harmonic mean, particularly in situations where conservative estimates are required.

The median tended to underestimate long-term total phosphorus loads at the majority of catchments. When overestimations were observed, they were at catchments with large negative correlations between EMC and total event runoff depth. The geometric mean also tended to overestimate long-term loads at catchments with large negative correlations between EMC and total event runoff depth. However, the overall number of overestimations of long-term load was almost equivalent to the number of underestimations for this measure of central tendency.

When cross-validation was used to calibrate the flow-weighted mean concentration, it tended to underestimate long-term loads at catchments with large negative correlations between EMC and total event runoff depth. In contrast, when a leave-one-out analysis was used to calibrate the flow-weighted mean concentration, it tended to overestimate long-term loads at these catchments. However, overall, the flow-weighted mean concentration overestimated long-term loads at the majority of catchments. It was therefore concluded that the flow-weighted mean generally produced conservative estimates of long-term load.

Overall, three catchments were assumed to bias the average long-term load errors associated with the various measures of central tendency. Two of these catchments exhibited positive correlations between EMC and total event runoff depth and the other exhibited the largest negative correlation between EMC and total event runoff depth. As a result, the average long-term load errors presented in Figure 6 may not be representative of the accuracy of each measure of central tendency at typical urban catchments. Another method was thus used to compare the accuracy of the various measures of central tendency. This method was based on a comparison between the most accurate measure of central tendency (flow-weighted mean concentration) and the remainder of measures of central tendency. The percentage of catchments where the long-term load error generated using the given measure of central tendency was lower than the long-term load error generated using the flow-weighted mean concentration was calculated. The results are presented in Figure 7.

The flow-weighted mean concentration produced the most accurate predictions of long-term load at the majority of catchments. However, when a cross-validation analysis was used to calibrate the measures of central tendency, the geometric mean estimate was found to produce more accurate estimates of long-term load at more than 40% of the analysed catchments. This suggested that the geometric mean produced relatively accurate estimates of long-term load at typical urban catchments with limited sampling.

Figure 7. Percentage of catchments where the long-term load error associated with the given measure of central tendency is lower than the long-term load error associated with flow-weighted mean concentration. AM, arithmetic mean; GM, geometric mean; HM, harmonic mean; Med, median

Various explanatory variables presented in Tables 1 and 2 were analysed to determine whether or not they could be used to identify the most applicable measure of central tendency for predicting SMC at a given catchment. As mentioned previously, the correlation between EMC and runoff depth at a given catchment was found to influence the selection of the most applicable measure of central tendency. However, it was found that no single explanatory variable could be used to predict the strength or sign of this correlation. Therefore, in order to define SMC at a catchment, stormwater sampling must be undertaken to determine the correlation between EMC and total runoff depth.

This study was based upon the analysis of a single water quality

constituent – total phosphorus. However, it is important to determine whether the conclusions from this study could also be valid for a range of other water quality constituents. In order to do so, Pearson correlations between EMCs and total runoff depths were viewed for a variety of water quality constituents measured during the NURP study, as calculated by Athayde *et al.* (1983). Table 4 presents the number of significant linear correlations (calculated using a two-tailed test at a 90% confidence interval) at catchments with more than five monitored storms. Overall, more significant negative correlations were viewed for the majority of constituents. This suggests that measures of central tendency accurately predicting the lower EMCs (i.e. geometric mean, median, harmonic mean) are generally more applicable than those that systematically place more emphasis on the larger EMCs (i.e. arithmetic mean). However, significant positive correlations between EMCs and total runoff depth were viewed at a number of catchments. In particular, total suspended solids had more positive correlations. For this contaminant, median, harmonic mean or geometric mean estimates would not be applicable. The flow-weighted mean is generally more applicable since it can reasonably handle both positive and negative correlations between EMC and total runoff depth.

Even though more than two-thirds of the analysed correlations presented in Table 4 were not deemed significant, this does not mean that the correlation between EMC and runoff depth should be ignored. It is highly possible that weaker trends still play an important role in determination of the measure of central tendency best suited to forecasting long-term loads at the majority of catchments. This situation was observed in the current study. As a result, it may be concluded that a flow-weighted mean is the most versatile measure of central tendency defining SMCs at urban catchments. However, it is suggested that a more

Water quality constituent	Total number of catchments	Number of significant Pearson correlations ($p < 0.1$)		
		Total	Negative	Positive
Total suspended solids	67	13	4	9
Chemical oxygen demand	64	24	23	1
Total phosphorus	67	20	16	4
Dissolved phosphorus	34	10	6	4
Total Kjeldhal nitrogen	64	19	18	1
Nitrogen oxides	57	17	15	2
Total copper	49	17	15	2
Total lead	59	15	13	2
Total zinc	56	19	18	1
Total	517	154	128	26
%		30	25	5

Table 4. The number of significant Pearson correlations between EMCs and runoff depths at well sampled catchments analysed during the NURP study (adapted from Athayde *et al.* (1983))

comprehensive analysis be undertaken using data from additional catchments and/or water quality constituents in order to prove the universality of the conclusions from this study.

From a practical perspective, the sign of the correlation between EMC and runoff depth is potentially the most important factor in determining which model should be applied in a particular circumstance. In most situations, this correlation is negative. This indicates that, as storms progress, pollutant concentrations become more diluted due to the washoff of supply-limited pollutant built up on the catchment surface. This phenomenon is supported by the underlying theory behind complex buildup–washoff models. However, in some circumstances, the dilution effect does not occur, resulting in higher pollutant concentrations for larger storm events. This may arise from a number of phenomena. Arguably the most likely cause is the contribution of pollutants from pervious surfaces, which only occurs once the catchment has become saturated later on in the storm event. As a result, eroded soil or dissolved constituents leaching from pervious surfaces are more likely to be prevalent in larger storm events. Alternatively, larger storm events may increase the volume of water entering the waterways, increasing the erosive power of the flow, subsequently eroding or resuspending the particulate matter that was deposited in the channels during smaller storms. This is supported by Corbett *et al.* (1997), who showed that the majority of suspended sediment in runoff during large storm events was contributed by stream channels. This result also suggests that the intensity of the rainfall and/or runoff may also influence pollutant concentration, leading to the divergence from the anticipated dilution effect of stormwater runoff assumed by most buildup–washoff models. This limitation may be partially overcome by separating the catchment into distinct land uses and specifying buildup and washoff coefficients according to land use. However, the process is made difficult by the lack of relevant literature available (Butcher, 2003). To complicate matters further, when such models are calibrated using at-site data, a number of non-unique parameter solutions are often found (Obropta and Kardos, 2007), making the calibration process difficult if not impossible. This supports the use of SMC models to predict long-term loads exported from an urban catchment – in particular, the flow-weighted mean.

A major limitation of the current study was associated with the limited number of catchments analysed. This limitation may be overcome by analysing stormwater quality for a range of other catchments and constituents. However, even with additional catchments, simple cross-validation analyses may still struggle to identify the optimum measure of central tendency defining SMC to predict long-term load. This is essentially due to the fact that storm sample size is likely to influence the results of the analyses. This was suggested by the current analysis, when it was found that the results differed depending on the method used for validation (i.e. cross-validation or leave-one-out analyses). To overcome this problem, a reiterative analysis may be undertaken, similar to that conducted by May and Sivakumar (2009a). In such

an analysis, storms may be separated into calibration subsets containing between 1 and N storms (N is the total number of storms measured at a given catchment). The available calibration data may then be used to generate an estimate of SMC, which can be used to generate a prediction of long-term load using data from the validation set. This process is then repeated a large number of times to generate an average long-term load prediction error for each calibration set size. These errors may then be used to identify the optimum measure of central tendency defining SMC used to predict long-term load.

4. Conclusions

Long-term total phosphorus loads from stormwater runoff were derived at 17 urban catchments using five different measures of central tendency defining SMC – arithmetic mean, geometric mean, median, harmonic mean and flow-weighted mean. Overall, the flow-weighted mean produced the most accurate estimates of long-term load. This was due to the fact that the flow-weighted mean tended to place more weight on the values of EMC associated with larger runoff depths. This resulted in the accurate prediction of the large storm loads that dominated the long-term load predictions at a given catchment. The geometric mean produced the second most accurate estimates of long-term load. However, along with the median and harmonic mean, the geometric mean estimate tended to systematically place more weight on the low values of EMC. This resulted in accurate predictions of long-term load at the majority of catchments exhibiting negative correlations between EMC and total runoff depth. In contrast, large underestimations of long-term load were observed at the two catchments exhibiting positive correlations between EMC and total event runoff depth. The arithmetic mean, on the other hand, produced more accurate predictions of long-term loads at these two catchments. However, the arithmetic mean tended to systematically overestimate long-term loads at the remainder of catchments exhibiting a negative correlation between EMC and total event runoff depth. The flow-weighted mean, on the other hand, produced accurate estimates of long-term loads at all catchments, irrespective of the strength or sign of the correlation between EMCs and runoff depth, making it the most versatile measure of central tendency for predicting long-term load from an urban catchment. It is therefore recommended that the flow-weighted mean be used to quantify the export of total phosphorus from urban catchments into receiving waters in order to determine the risk of eutrophication.

Acknowledgements

The authors acknowledge Dr Pam Davy and Associate Professor Ken Russell for assisting with the statistical analyses, and Dr Nancy Driver for providing data for the study.

REFERENCES

- Athayde DN, Shelley PE, Driscoll ED, Gaboury D and Boyd G (1983) *Results of the Nationwide Urban Runoff Program*, Vol. 1. US Environmental Protection Agency, Washington, DC, USA, PB84-185545.

- AWRA (American Water Resources Association) (1998) *NURP Data*. Cahaba/Warrier Student Chapter of the American Water Resources Association, University of Alabama, Tuscaloosa, AL, USA. See <http://www.eng.ua.edu/~awra/download.htm> (accessed 23/11/2001).
- Brezonik PL and Stadelmann TH (2002) Analysis and predictive models of stormwater runoff volumes, loads and pollutant concentrations from watersheds in the Twin Cities metropolitan area, Minnesota, USA. *Water Research* **36**(7): 1743–1757.
- Brodie IM and Dunn PK (2009) Suspended particle characteristics in storm runoff from urban impervious surfaces in Toowoomba, Australia. *Urban Water Journal* **6**(2): 137–146.
- Brodie IM and Dunn PK (2010) Commonality of rainfall variables influencing suspended solids concentrations in storm runoff from three different urban impervious surfaces. *Journal of Hydrology* **387**(3): 202–211.
- Burian SJ, Streit GE, McPherson TN, Brown MJ and Turin HJ (2001) Modeling the atmospheric deposition and stormwater washoff of nitrogen compounds. *Environmental Modelling & Software* **16**(5): 467–479.
- Butcher JB (2003) Buildup, washoff and event mean concentrations. *Journal of the American Water Resources Association* **39**(6): 1521–1528.
- Chow MF, Yusop Z and Mohamed M (2011) Quality and first flush analysis of stormwater runoff from a tropical commercial catchment. *Water Science and Technology* **63**(6): 1211–1216.
- Corbett CW, Wahl M, Porter DE, Edwards D and Moise C (1997) Nonpoint source runoff modeling: a comparison of a forested watershed and an urban watershed on the South Carolina coast. *Journal of Experimental Marine Biology and Ecology* **213**(1): 133–149.
- Davis B and Birch G (2010) Comparison of heavy metal loads in stormwater runoff from major and minor urban roads using pollutant yield rating curves. *Environmental Pollution* **158**(8): 2541–2545.
- Driver N and Tasker G (1990) *Techniques for Estimation of Storm-runoff Loads, Volumes, and Selected Constituent Concentrations in Urban Watersheds in the United States*. United States Geological Survey, Washington, DC, USA, Water supply paper 2363.
- Duncan HP (1999) *Urban Stormwater Quality: A Statistical Overview*. CRC for Catchment Hydrology, Monash University, Victoria, Australia, Report 99/3.
- Ferreira JG, Bricker SB and Simas TC (2007) Application and sensitivity testing of a eutrophication assessment method on coastal systems in the United States and European Union. *Journal of Environmental Management* **82**(4): 433–445.
- Ferreira JG, Anderson JH, Borja A et al. (2011) Overview of eutrophication indicators to assess environmental status within the European Marine Strategy Framework Directive. *Estuarine, Coastal and Shelf Science* **93**(2): 117–131.
- Francey M, Fletcher TD, Deletic A and Duncan H (2010) New insights into the quality of urban stormwater in South Eastern Australia. *Journal of Environmental Engineering* **136**(4): 381–390.
- Hilton J, O'Hare M, Bowes MJ and Jones JI (2006) How green is my river? A new paradigm of eutrophication in rivers. *Science of the Total Environment* **365**(1): 66–83.
- Huber WC (1992) Prediction of urban nonpoint source water quality: methods and models. *Proceedings of an International Symposium on Urban Stormwater Management*, Institution of Engineers, Sydney, pp. 1–16.
- Jewell TK and Adrian DD (1982) Statistical analysis to derive improved stormwater quality models. *Journal of Water Pollution Control Federation* **54**(5): 489–499.
- Leecaster MK, Schiff K and Tiefenthaler LL (2002) Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research* **36**(6): 1556–1564.
- Maier HR and Dandy GC (1996) The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research* **32**(4): 1013–1022.
- Maniquiz MC, Lee S and Kim LH (2010) Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables. *Journal of Environmental Sciences* **22**(6): 946–952.
- May D and Sivakumar M (2003) Techniques for predicting total phosphorus in urban stormwater runoff at unmonitored catchments. *Journal of Australia and New Zealand Industrial and Applied Mathematics (ANZIAM)* **45**(E): C296–C309.
- May D and Sivakumar M (2008) Comparison of artificial neural network and regression models in the prediction of urban stormwater quality. *Water Environment Research* **80**(1): 4–9.
- May D and Sivakumar M (2009a) Optimum number of storms required to derive site mean concentration at urban catchments. *Urban Water Journal* **6**(2): 107–113.
- May D and Sivakumar M (2009b) The prediction of nutrient concentrations in urban stormwater. *Journal of Environmental Engineering* **135**(8): 586–594.
- May D and Sivakumar M (2009c) The prediction of heavy metal concentrations in urban stormwater. *Water and Environment Journal* **23**(4): 247–254.
- May D and Sivakumar M (2009d) Prediction of urban stormwater quality using artificial neural networks. *Environmental Modelling and Software* **24**(2): 296–302.
- Mourad M, Bertrand-Krajewski JL and Chebbo G (2005a) Sensitivity to experimental data of pollutant site mean concentration in stormwater runoff. *Water Science and Technology* **51**(2): 155–162.
- Mourad M, Bertrand-Krajewski JL and Chebbo G (2005b) Stormwater quality models: sensitivity to calibration data. *Water Science and Technology* **52**(5): 61–68.
- Mourad M, Bertrand-Krajewski JL and Chebbo G (2005c) Calibration and validation of multiple regression models for stormwater quality predictions: data partitioning, effect of dataset size and characteristics. *Water Science and Technology* **52**(3): 45–52.
- Mourad M, Bertrand-Krajewski JL and Chebbo G (2006) Design of a retention tank: comparison of stormwater quality models

- with various levels of complexity. *Water Science and Technology* **54**(6–7): 231–238.
- Obropta CC and Kardos JS (2007) Review of urban stormwater quality models: deterministic, stochastic, and hybrid approaches. *Journal of the American Water Resources Association* **43**(6): 1508–1523.
- Owen F and Jones R (1990) *Statistics*, 3rd edn. Pitman, London, UK.
- Paerl HW (2006) Assessing and managing nutrient-enhanced eutrophication in estuarine and coastal waters: interactive effects of human and climatic perturbations. *Ecological Engineering* **26**(1): 40–54.
- Park M, Swamikannu X and Stenstrom MK (2009) Accuracy and precision of the volume-concentration method for urban stormwater modelling. *Water Research* **43**(11): 2773–2786.
- Park SY, Lee KW, Park IH and Ha SR (2008) Effect of the aggregation level of surface runoff fields and sewer network for a SWMM simulation. *Desalination* **226**(1): 328–337.
- Sliva L and Williams DD (2001) Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Research* **35**(14): 3462–3472.
- Taylor GD, Fletcher TD, Wong TH, Breen PF and Duncan HP (2005) Nitrogen composition in urban runoff – implications for stormwater management. *Water Research* **39**(10): 1982–1989.
- Thomann RV and Mueller JA (1987) *Principles of Surface Water Quality Modelling and Control*. Harper Collins, New York, NY, USA.
- Van Buren MA, Watt WE and Marsalek J (1997) Application of the log-normal and normal distributions to stormwater quality parameters. *Water Research* **31**(1): 95–104.
- Vaze J and Chiew F (2002) Experimental study of pollutant accumulation on an urban road surface. *Urban Water* **4**(4): 379–389.
- Vaze J and Chiew F (2003) Comparative evaluation of urban stormwater quality models. *Water Resources Research* **39**(10): 1280–1290.
- Wang L, Wei J, Huang Y, Wang G and Maqsood I (2011) Urban nonpoint source pollution buildup and washoff models for simulating storm runoff quality in the Los Angeles County. *Environmental Pollution* **159**(7): 1932–1940.
- Zoppou C (2001) Review of urban storm water models. *Environmental Modelling and Software* **16**(3): 195–231.

WHAT DO YOU THINK?

To discuss this paper, please email up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial panel, will be published as a discussion in a future issue of the journal.

Proceedings journals rely entirely on contributions sent in by civil engineering professionals, academics and students. Papers should be 2000–5000 words long (briefing papers should be 1000–2000 words long), with adequate illustrations and references. You can submit your paper online via www.icevirtuallibrary.com/content/journals, where you will also find detailed author guidelines.