Journal of Great Lakes Research 41 (2015) 780–793

Contents lists available at ScienceDirect

Journal of Great Lakes Research

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

Estimation of tributary total phosphorus loads to Hamilton Harbour, Ontario, Canada, using a series of regression equations

Tanya Long [a,*], Christopher Wellen [b], George Arhonditsis [c], Duncan Boyd [a], Mohamed Mohamed [a], Kristin O'Connor [d]

[a] Ontario Ministry of the Environment and Climate Change, Environmental Monitoring and Reporting Branch, 125 Resources Road, Toronto, ON M9P 3V6, Canada
[b] Watershed Hydrology Group, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada
[c] Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, 1265 Military Trail, Toronto, ON M1C 1A4, Canada
[d] Hamilton Harbour Remedial Action Plan (HH RAP) Office, Canada Centre for Inland Waters, 867 Lakeshore Road, Burlington, ON L7S 1A1, Canada

ARTICLE INFO

Article history:
Received 30 September 2014
Accepted 16 March 2015
Available online 23 April 2015

Communicated by Joseph Makarewicz

Keywords:
Total phosphorus (TP)
Event mean concentration (EMC)
Hamilton Harbour
Tributary loads
Regression methods

ABSTRACT

Event-based sampling was conducted from July 2010 to May 2012 at four stations in the watersheds of Hamilton Harbour, Ontario, Canada, with the primary objective of estimating total phosphorus (TP) loads. Eighty-seven 24-hour, level-weighted composite samples were collected during a variety of catchment states (rain, snowmelt, baseflow), and TP concentrations were regressed against flow or precipitation in an attempt to mitigate the considerable loading estimation bias arising from event-scale hysteresis. Annual average TP loads were estimated for 2008 to 2012 and were the highest from the Desjardins Canal (17.4 kg/d to 65.6 kg/d), followed by Red Hill Creek (6.4 kg/d to 25.8 kg/d), Grindstone Creek (3.4 kg/d to 33.4 kg/d), and Indian Creek (3.0 kg/d to 7.9 kg/d). Daily TP loads varied by three orders of magnitude between wet and dry conditions, with storm events driving peak daily loads in the urban watersheds, and spring freshet in the agricultural and wetland influenced watersheds. Areal TP loads were higher from the urban relative to the agricultural watersheds. This study demonstrated that the tributaries did not meet the Hamilton Harbour Remedial Action Plan (HH RAP) initial target of 65 kg/d in 2008 to 2011 but did in 2012. Comparison of three loading methods emphasized the vital role of characterizing TP concentrations during high flow events. The higher resolution TP loads generated in this study will assist the HH RAP in forming additional remedial actions in the watersheds for delisting the Hamilton Harbour Area of Concern.

Introduction

Eutrophication of surface waters has long been linked to an overenrichment of phosphorus as it is generally the limiting nutrient for algal growth and biomass in freshwater systems (Schindler, 1977). Such is the case in Hamilton Harbour, a 2150-ha partially enclosed embayment located at the western end of Lake Ontario, Ontario, Canada. Historically, nutrient loads from three wastewater treatment plants (WWTPs), from combined sewer overflows (CSOs), and from industry, urban and agricultural runoff entering Hamilton Harbour via four major tributary inputs resulted in severe eutrophication of the Harbour. In response, Hamilton Harbour was declared an Area of Concern (AOC) under the 1987 Great Lakes Water Quality Agreement.

The Hamilton Harbour Remedial Action Plan (HH RAP) was released in 1992 in part to address nuisance algal growth, reductions in water clarity, and a hypoxic hypolimnion during the summer (HH RAP, 1992). Substantial nutrient loading reductions over the past few decades have led to measurable improvements in the trophic status of the system (Charlton, 1997; Hiriart-Baer et al., 2009); but ambient water quality goals have not yet been achieved (HH RAP, 2012). Recent eutrophication modelling has demonstrated that achievement of the HH RAP TP goal of 20 μg/L is in part contingent on our assumptions of the exogenous loads to the Harbour (Gudimov et al., 2011; Ramin et al., 2012). While it is believed that loads from the point sources have been well characterized, the magnitude of TP loads attributed to the creeks is highly uncertain (HH RAP, 2010).

The TP loads from the creeks were to be revised by the HH RAP utilizing recent monitoring data collected under Ontario's Provincial Water Quality Monitoring Network (PWQMN) (HH RAP, 2010); however, this sampling programme collects monthly samples primarily during baseflow conditions. Accurate characterization of TP dynamics during high flow conditions was deemed critical to increasing the accuracy of TP loading estimates as the majority of annual TP loads occur during brief, high flow events such as storm events and the spring freshet (Booty et al., 2013; Duan et al., 2012; Horowitz, 2013; Macrae et al., 2007; Old et al., 2003; Richards and Holloway, 1987; Sharpley et al., 1993). Thus, not only was an updated event-based monitoring dataset

* Corresponding author. Tel.: +1 416 235 6247: fax: +1 416 235 6235.
E-mail address: tanya.long@ontario.ca (T. Long).

http://dx.doi.org/10.1016/j.jglr.2015.04.001
0380-1330/

needed, but also a simple TP loading estimation method with minimal data requirements will be essential to evaluate compliance with loading objectives and project the future state of the Harbour.

A common technique for estimating tributary loads to the Great Lakes is the stratified Beale Ratio Estimator. When applied to flow stratified data, the basis of this well-described approach (Cochran, 1977; Dolan et al., 1981; Dolan and Chapra, 2012; Dolan and McGunagle, 2005) is the assumption that the ratio of load to flow for all days within a chosen flow stratum will be the same as the ratio of daily load to daily flow on the days when water quality was measured within that flow stratum. Typically, flow strata are selected to cover a range of base flow or low flow conditions, as well as one or more high flow conditions. The Beale Ratio Estimator was originally endorsed by the International Joint Commission (IJC) due to a relatively low bias, high precision, and its robustness (Preston et al., 1989; Rao, 1979; Richards et al., 1996; Richards and Holloway, 1987; Richards, 1998). In reviews of multiple loading methods, however, no group of methods has been found to outperform others for all scenarios examined (Moatar and Meybeck, 2005; Preston et al., 1989; Richards and Holloway, 1987). This suggests that the choice of a loading estimation technique should be driven by the nature of the dataset and end use of the calculated loads.

Due to their relative simplicity and ease of use, regression models that establish the relationship between TP concentration and flow can be a very useful tool. The United States Geological Survey (USGS) has recently established regression models to estimate real time nutrient loads and concentrations of Great Lakes tributaries in an effort to better understand the water quality impact of land management practices and the impact of other restoration activities (Baldwin et al., 2013). By providing daily load estimates, regression methods elucidate variability in the system, thereby offering important information for forming meaningful remedial actions in the watersheds (O'Connor et al., 2011). Such data resolution is not achieved through the current HH RAP method (HH RAP, 2010), an averaging estimator approach, or a ratio approach. Furthermore, acceptable accuracy of estimated loads can be obtained with less resource requirements through regression techniques if concentration and flow are strongly correlated for a wide range of streamflow values (Preston et al., 1989; Quilbé et al., 2006; Richards, 1998).

A major impediment to stronger regression relationships is hysteresis (Williams, 1989) because distinct TP concentration versus flow relationships have been found for the rising and falling limbs of a hydrograph over the course of an event (Aulenbach and Hooper, 2006a,2006b; Hirsch et al., 2010; Macrae et al., 2007; O'Connor et al., 2011). By addressing the issue of event-scale hysteresis, the performance of the regression approach can be further improved. Use of a flow proportional composite sample collected for the full duration of an event mitigates hysteresis in TP concentration versus flow relationships as the samples are integrative of contributors to the variability (e.g., first flush). Very few studies have used a flow-weighted TP concentration dataset in a regression-based approach to estimate tributary TP loads, but results of these studies have suggested that it is a viable modification of the more traditional approach (Booty et al., 2013).

The primary goal of this study was to reduce uncertainty in the tributary TP loading estimates to Hamilton Harbour due to the pivotal role that phosphorus plays in the trophic status and resulting ecology of this system. Accurate loading estimates are needed to ensure that expectations for improvements to the trophic status of Hamilton Harbour are realistic and that the optimal remedial actions in the watersheds will be implemented. The specific objectives of this paper are to:

1) Develop a relatively simple, empirical TP loading estimation method that can be used by the HH RAP to estimate annual TP loads to the Harbour from the four major tributaries;
2) Calculate TP loads that were delivered to Hamilton Harbour from the major tributaries during the July 2010 to May 2012 monitoring period, as well as annual average loads for 2008 to 2012, and evaluate if these sources have met their HH RAP delisting targets; and
3) Compare updated TP loading estimates to those estimated by methods currently endorsed or considered by the HH RAP and recommend methodological changes if warranted.

While this study was primarily conducted to meet the needs of the HH RAP, reducing the uncertainty in Hamilton Harbour's watershed TP loads has wide-ranging benefits to similar systems. For example, few studies in general have assessed the precision and accuracy of TP loads for the large number and diversity of events measured in this study, especially for small, primarily urban watersheds. Further, our empirical approach is user-friendly and adds to the knowledge base on methods to estimate TP loads, an exercise of the utmost importance in the Great Lakes region (2012 Great Lakes Water Quality Agreement, http://www.ec.gc.ca/grandslacs-greatlakes/default.asp?lang=En&n=A1C62826-1, Accessed: May 13, 2014). Our approach also lends insights into some of the processes driving the loading patterns, thus providing suggestions as to how TP loads can potentially be reduced.

Material and methods

Overview of study area

To sample the four main tributary inputs to Hamilton Harbour, four water quality monitoring stations were installed in the summer of 2010 in Burlington and Hamilton, Ontario, Canada, cities with 2011 populations of 175,779 and 519,949, respectively (Statistics Canada, 2011 Census, http://www12.statcan.gc.ca/census-recensement/index-eng.cfm, last accessed April 21, 2014). The monitoring stations were located at the mouths of Red Hill Creek and Indian Creek – two primarily urban watersheds each traversed by three major expressways – and Grindstone Creek and the Desjardins Canal — two primarily agricultural watersheds (Fig. 1; Table 1). Important to note is that the Desjardins Canal station is not technically on a tributary but rather on the canal that hydraulically joins the Cootes Paradise wetland in the west and Hamilton Harbour in the east. The water sampled at the Desjardins Canal reflects what is delivered to the Harbour from a variety of sources as it is integrative of complex wetland processes in Cootes Paradise, numerous tributary inputs, effluent from the Dundas tertiary WWTP located 3.8 km to the west (18 ML/day; HH RAP, 2010), and six CSOs. Samples collected from Red Hill Creek are also intermittently influenced by CSOs; the two CSO points are located 0.8 km and 4 km upstream. Additional details on the City of Hamilton CSOs pertaining to this study are in the Electronic Supplementary Material (ESM Appendix S1).

Event-based water quality data collected July 2010 to May 2012

Between July 5, 2010 and May 8, 2012, 87 24-hour periods during rain events, spring freshet, or baseflow were sampled at the four monitoring stations. The monitoring station setup as well as sample collection, retrieval, and processing are described in detail in Long et al. (2014) but are described briefly below. The core of each monitoring station was a Teledyne ISCO (Model 6712) automatic water sampler equipped with a water level bubbler module (Model 730) as well as power and telephone connections to permit remote programming and data downloads. Water level data were collected in 15-minute intervals and were used for triggering the sampler during an event, as well as for post-event sample processing. Serial correlation is not an issue for the event-based data collected in our study as the average time between the collection of samples was approximately 1 week, a time interval greater than a characteristic correlation time of 1.12 days calculated for Red Hill Creek and 5.37 days calculated for Grindstone Creek (see ESM Appendix S1).

For each station and event, 1-L water samples were collected hourly for 24 h which, during rain events, was generally enough time to capture the rising limbs well as the peak and falling limbs of the hydrograph.

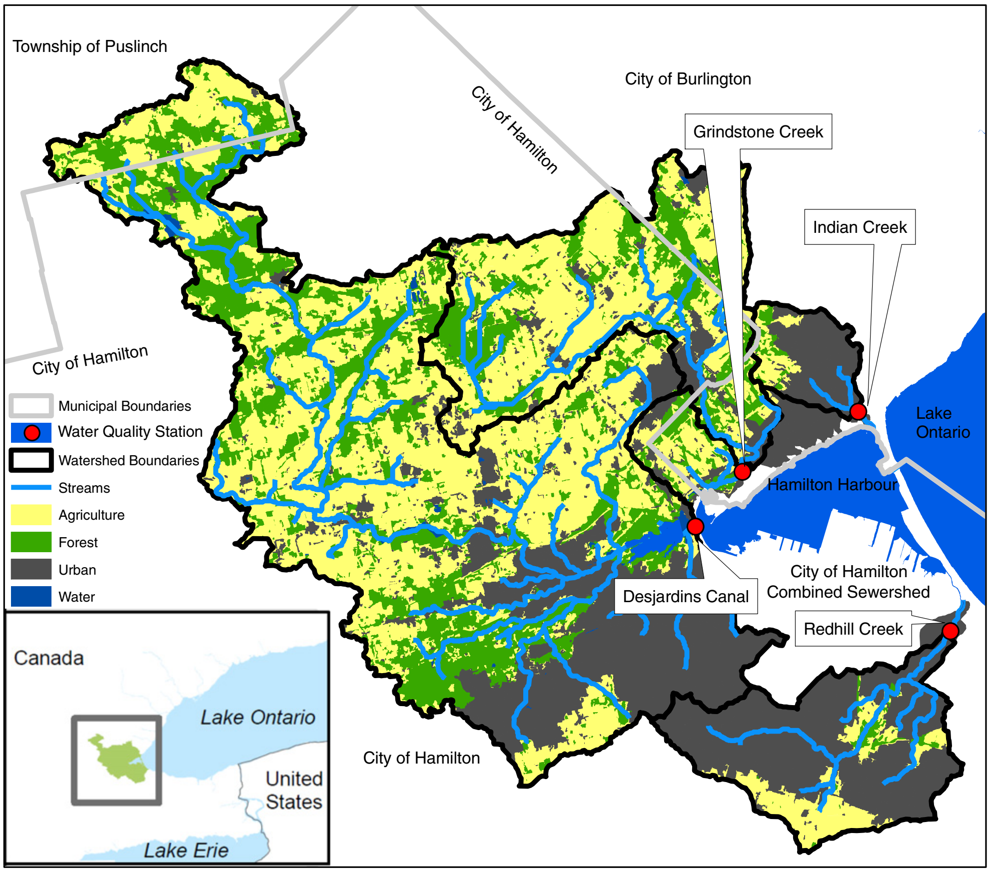


Fig. 1. Location of sampling stations and land use in the watersheds of Hamilton Harbour. Adapted from Long et al. (2014).

Aliquot volumes proportional to water level at the time of sampling were determined for each of the 24 grab samples to form a level-weighted composite sample. We were not able to prepare flow-weighted composite samples as real time flow data and/or rating curves were not available at all our stations. The use of level-weighted composite samples instead of flow-weighted composites slightly underpredicts TP concentrations by an average of approximately 15% based on an analysis of two individual rain events (see ESM Appendix S1); however, this is not expected to have a major impact on the overall interpretation of results. Any bias is expected to be within the range of other sources of variability and error, and any TP loads considered problematic as measured through level-weighted composite samples would also be problematic if measured through flow-weighted composite samples.

Water quality data collected through the Provincial Water Quality Monitoring Network (PWQMN)

Independent to the collection of water quality data in our study but important in our ultimate data analysis, water quality data were also obtained for Red Hill Creek and Grindstone Creek from Ontario's Provincial Water Quality Monitoring Network (PWQMN). The PWQMN is a long-term water quality monitoring programme for Ontario's tributaries managed by the Ontario Ministry of the Environment and Climate Change (OMOECC) and implemented by local conservation authorities who conduct the sampling. Manual grab samples are collected monthly for the ice-free season, a sampling regime that results in approximately six to eight water samples a year reflecting primarily low flow conditions. The PWQMN station for Red Hill Creek (Station ID: 09000100502) is located approximately 2.3 km upstream from our event-based monitoring station, and the PWQMN station for Grindstone Creek (Station ID: 09000902402) is co-located with the station in our study. There are no PWQMN stations on Indian Creek or the Desjardins Canal.

Laboratory analysis

Water samples collected in this study and through the PWQMN monitoring programme were submitted for analysis to the OMOECC Laboratory Services Branch in Toronto, Ontario, accredited by the Canadian Association for Laboratory Accreditation (CALA) and the Standards Council of Canada (SCC). The analysis of TP was made by colourimetry by OMOECC's TOTNUT3367 method (OMOE, 2010, 2012a).

Table 1
Area and land use of the four Hamilton Harbour watersheds (Long et al., 2014).

	Land use (%)[a]				
	Watershed Area (km^2)	Urban	Urban Greenspace	Agricultural (pasture and cropland)	Forest
Red Hill Creek	65	66	15	16	3
Indian Creek	23[b]	54	18	17	10
Grindstone Creek	87	6	3	60	29
Desjardins Canal (Cootes Paradise)[c]	290	17	6	47	28

[a] Land use data from Ontario Ministry of Natural Resources (2008).
[b] 74% Hagar-Rambo watershed and 26% Indian Creek watershed (Conservation Halton, 2006).
[c] Subwatersheds of Cootes Paradise include Spencer Creek (235 km^2), Chedoke Creek (25 km^2), Borers Creek (20 km^2) and other small creeks.

Discharge data

Discharge data for Red Hill Creek and Grindstone Creek were obtained from the Water Survey of Canada (WSC) for Hydat flow stations located in these watersheds. The Red Hill Creek WSC flow station (Station ID: 02HA014) is located approximately 1 km upstream from our event-based monitoring station, and approximately 1.3 km downstream from the Red Hill Creek PWQMN station. Discharge data for Red Hill Creek are not available for 2008 due to the construction of the Red Hill Valley Parkway. The Grindstone Creek WSC flow station (Station ID: 02HB012) is co-located with our event-based monitoring station and the PWQMN station. Discharge data for Grindstone Creek are not available for January 2008 due to vandalism of equipment at the site. Discharge data for Indian Creek during August 2010 to May 2012 were obtained by installation of a Teledyne ISCO 2150 Flow Module at our Indian Creek event-based monitoring station. Outside of the period when the current meter was installed, discharge data were based on an empirical regression with discharge from the Red Hill Creek WSC Hydat flow station ($r^2 = 0.85$; Long et al., 2014, Supplementary Data Appendix A). Estimation of discharge out of the Desjardins Canal was made based on an empirical regression with discharge from the Spencer Creek WSC Hydat flow station (Station ID: 02HB007). This regression was developed from current meter monitoring data collected during the summer of 2009 at the Desjardins Canal ($r^2 = 0.66$; Long et al., 2014, Supplementary Data Appendix A). Important to note is that WSC flow data are collected on an ongoing basis independent from our study and as such, use of these data were considered for forming the core of a TP loading method that could be undertaken in years beyond the cessation of our study.

Generation of TP load estimates and statistical analysis

At each station, TP loads for each of the 87 24-hour sampling periods from July 2010 to May 2012 were calculated as the product of the TP concentration in the level-weighted composite sample and the corresponding 24-hour discharge volume. For sampling periods that were less than the standard 24 h due to equipment failure or other unscheduled changes to the sample programming, the total loads were prorated to 24 h for comparability purposes.

A series of simple linear regression models (Neter et al., 1996) were used to examine the relationships between the TP concentrations and flows using Microsoft Excel (Microsoft Office Professional Plus 2010) with the data Analysis Toolpak add-in. For 2008 to 2012, daily average TP concentrations for Red Hill Creek, Indian Creek, and Grindstone Creek were estimated by regressing log-transformed level-weighted TP concentrations for samples collected July 2010 to May 2012 against log-transformed average event flow (total flow volume of period/duration of that period). Data were log transformed to linearize the relationship and to address the influence of outliers. In addition, log-transformed TP concentrations for grab samples collected 2008 through 2012 from Red Hill Creek and Grindstone Creek through the PWQMN sampling programme were regressed against log-transformed daily average discharge. For the Desjardins Canal, the complexity of the flow and nutrient dynamics at this location required exploration of a variety of approaches to estimate loading. Daily average TP concentrations for 2008 to 2012 were estimated through one of three empirical equations derived from data collected July 2010 to May 2012: (1) a sine wave equation for May to November (for days with precipitation < 15 mm); (2) a log-log TP concentration versus flow regression for December to April (for days with precipitation < 15 mm); and (3) a log-log TP concentration versus precipitation regression for all days with precipitation > 15 mm.

Regression equations in this study were considered strong if the coefficient of determination (r^2) was above 0.5, a threshold that has generally been accepted as representing strong regressions in other tributary loading studies (Booty et al., 2013; Macrae et al., 2007; Moatar and Meybeck, 2005; Quilbé et al., 2006). We used Analysis of Variance (ANOVA) to infer about the significance of the amount of variability explained by our regression models relative to the residual variability. The falsification of the null hypothesis of no relationship was based on the F-test and a 5% level of significance. For the Grindstone Creek station, statistically significant differences in the slopes between the winter/spring and summer/fall regression equations for both the event-based and PWQMN datasets were tested using one-way analysis of covariance (ANCOVA) in the PAST Software (Hammer et al., 2001).

At each station, daily TP loads for 2008 to 2012 were estimated as the product of average discharge over the 24-hour period of interest and the TP concentration as predicted by the applicable equation for each station. Following the estimation of daily TP loads, a correction factor developed by Ferguson (1987) was applied to the loads estimated through log log regressions to increase accuracy when back transforming data:

$$L_{ct} = L_r \exp\left(2.651\ SE^2\right) \quad (1)$$

where L_{ct} is the corrected load, L_r is the load approximated by regression, and SE is the standard error of the estimate of the regression in $\log_{10}$ units (given $\log_{10}$ transformed regressions). This correction factor has been used in similar studies to correct for the retransformation bias that results in an underestimation of loads from use of log-log regressions in loading estimation (Moatar and Meybeck, 2005; O'Connor et al., 2011; Quilbé et al., 2006; Richards, 1998). Annual TP loads for the event-based and PWQMN methods were calculated as the sum of daily loads. Statistically significant differences between annual TP loads estimated from the PWQMN dataset relative to annual TP loads estimated from the event-based dataset were evaluated for both Red Hill Creek and Grindstone Creek by the paired *t*-test in the Excel Data Analysis Toolpak. All statistical tests performed were considered significant at the level of $p \leq 0.05$. Annual average TP loads for the HH RAP method were estimated following methods outlined in the HH RAP Loadings Report (HH RAP, 2010; ESM Appendix S1). Seasonal TP loads for 2008 to 2012 were calculated as the sum of the 2008 to 2012 daily TP loads within each seasonal bin (summer: June 21 to September 20; fall: September 21 to December 20; winter: December 21 to March 20; spring: March 21 to June 20) as estimated from the regression models developed from the event-based data collected July 2010 to May 2012. The percentage of loads delivered from each tributary during the days that comprised the top 10% of events was calculated. Daily TP loads from 2008 to 2012 were sorted in descending order, then the sum of the top 10% of daily loads were calculated over the sum of the daily loads for the entire 5-year period.

We assessed the performance of the empirically derived models using four metrics. The coefficient of determination (r^2) was calculated to evaluate the agreement between the TP loads calculated from measured TP concentration and flow during the 87 sampled events and those predicted from the station-specific regression model(s). In addition, we calculated Nash and Sutcliff's (1970) index of model efficiency (NSE), where a value of one represents a perfect model fit, a value of zero suggests that the predictive capacity of the model is as good as the mean of the measured values, and a negative value means the mean of the measured values is a better predictor than the model itself. We also calculated the relative error (RE) as presented by Arhonditsis and Brett (2004) and the root mean square error (RMSE).

Results

TP load estimates from concentrations and flow measured during 87 sampling periods between July 2010 and May 2012

Water samples in our study were collected over the wide range of flows measured at each of the monitoring stations during the July 2010 to May 2012 study period (Fig. 2a–d). High variability was observed in TP loads among sampling events, with minimum TP loads

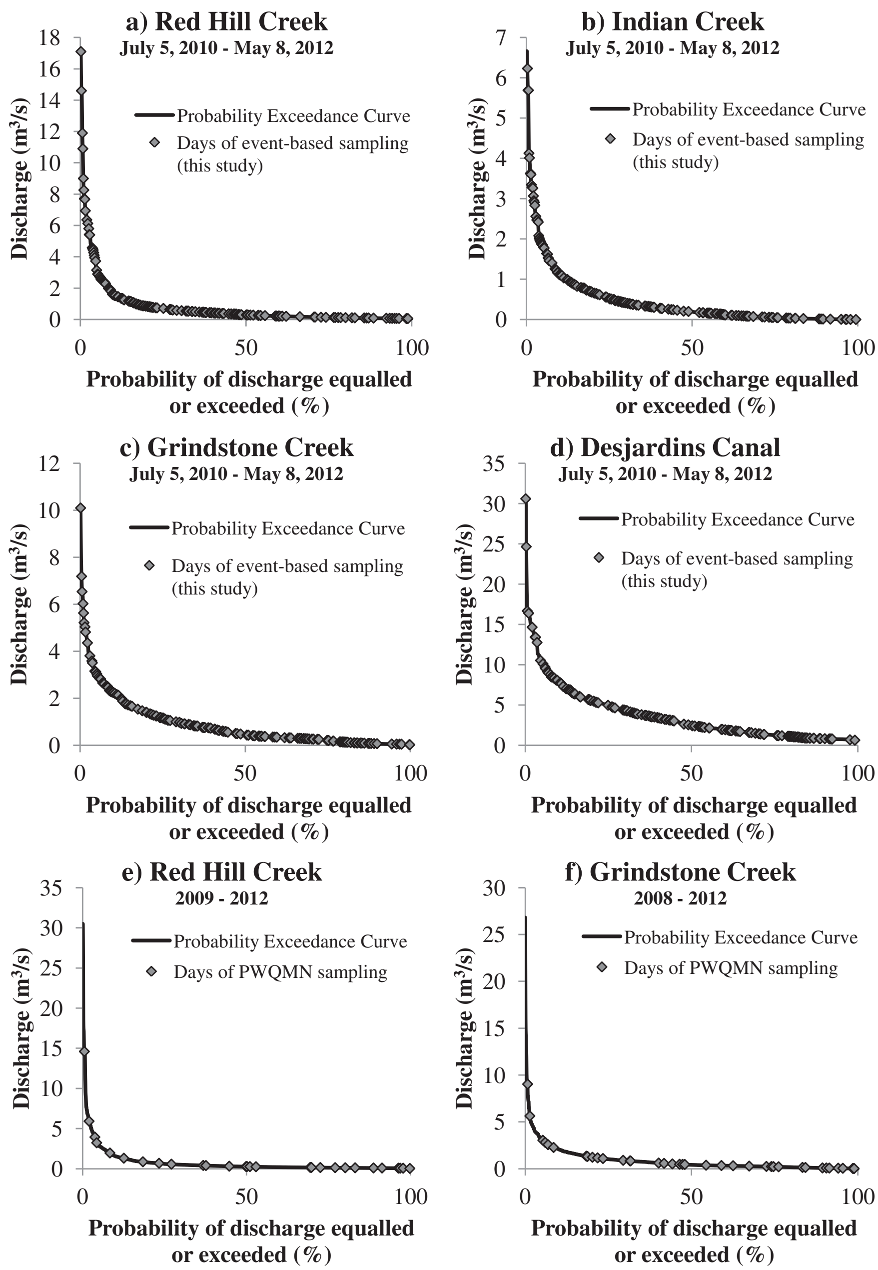


Fig. 2. Discharge probability exceedance curves for July 5, 2010 to May 8, 2012 and dates of event-based sampling (this study) at a) Red Hill Creek, b) Indian Creek, c) Grindstone Creek, and d) Desjardins Canal; 2009 to 2012 and PWQMN sampling at e) Red Hill Creek; 2008 to 2012 and PWQMN sampling at f) Grindstone Creek.

measured during baseflow (23% of sampled events) and maximum TP loads measured during rain/melt events (77% of sampled events). TP loads during rain/melt events were up to three orders of magnitude greater than baseflow loads (Table 2). The event during which the largest 24-hour TP load was measured over the course of the study varied among the four stations. The largest TP load was a fall precipitation event at both Red Hill Creek (841 kg/d) and Indian Creek (152 kg/d), whereas the largest TP load was a combined spring melt/precipitation event at both Grindstone Creek (334 kg/d) and the Desjardins Canal (704 kg/d). As such, the maximum 24-hour TP load of the four stations was measured at Red Hill Creek, although important to note is that median TP loads were the highest from the Desjardins Canal, reflecting the higher variability in TP loads from the former relative to more temporally consistent TP loads from the latter.

2008 to 2012 annual average TP load estimates

Regression models developed from event-based data collected July 2010 to May 2012

During our 22-month period of study, we sampled during the day of peak flow measured at three of the four stations (Fig. 2a–d), an important facet of regression-based load estimates given the challenges of extrapolation in this approach (Quilbé et al., 2006). The log-log TP concentration versus flow regressions using the event-based data collected in this study were strong for Red Hill Creek and Indian Creek (Fig. 3a and b). Implicit in the regression for Red Hill Creek is any TP input from CSOs, as many of the CSO events known to occur during the July 2010 to May 2012 period coincided with sample collections at Red Hill Creek (ESM Appendix S1; ESM Fig. S1). The log-log TP concentration versus flow regression for

Table 2
Summary of TP load estimates based on measured level-weighted TP concentrations and flow for 87 sampling periods between July 5, 2010 and May 8, 2012 at four stations in the Hamilton Harbour watershed.

	Mean (standard deviation) (kg/d)	Median (kg/d)	Minimum (kg/d)	Maximum (kg/d)
Red Hill Creek				
All sampled events (n = 92)	58.3 (126.5)	8.0	0.1	841 (Sep 28–29, 2010; 54.8 mm of rain)
Baseflow (n = 21)	0.5 (0.4)	0.3	0.1	1.5 (Jun 10–11, 2011)
Rain/melt events (n = 71)	75.4 (139.7)	19.5	0.4	841 (Sep 28–29, 2010; 54.8 mm of rain)
Indian Creek				
All sampled events (n = 97)	20.8	6.1	0.1	152 (Nov 29–30, 2011; 40.2 mm of rain)
Baseflow (n = 23)	1.0 (1.2)	0.4	0.1	3.6 (May 24–25, 2011)
Rain/melt events (n = 74)	26.9 (34.8)	12.8	0.2	152 (Nov 29–30, 2011; 40.2 mm of rain)
Grindstone Creek				
All sampled events (n = 89)	31.7 (62.9)	4.8	0.2	334 (Mar 10–11, 2011; melt + 22.1 mm of rain)
Baseflow (n = 19)	2.6 (4.0)	0.9	0.2	14.8 (June 10–11, 2011)
Rain/melt events (n = 70)	39.6 (68.9)	9.8	0.2	334 (Mar 10–11, 2011; melt + 22.1 mm of rain)
Desjardins Canal				
All sampled events (n = 95)	57.1 (108.3)	19.3	1.8	704 (Mar 11–12, 2011; melt + 9.6 mm of rain)
Baseflow (n = 24)	15.7 (16.1)	10.1	1.8	68.4 (Jun 10–11, 2011)
Rain/melt events (n = 71)	71.1 (122)	26.9	4.1	704 (Mar 11–12, 2011; melt + 9.6 mm of rain)
Total sum (all sampled events)	167.9	38.2	2.2	2031

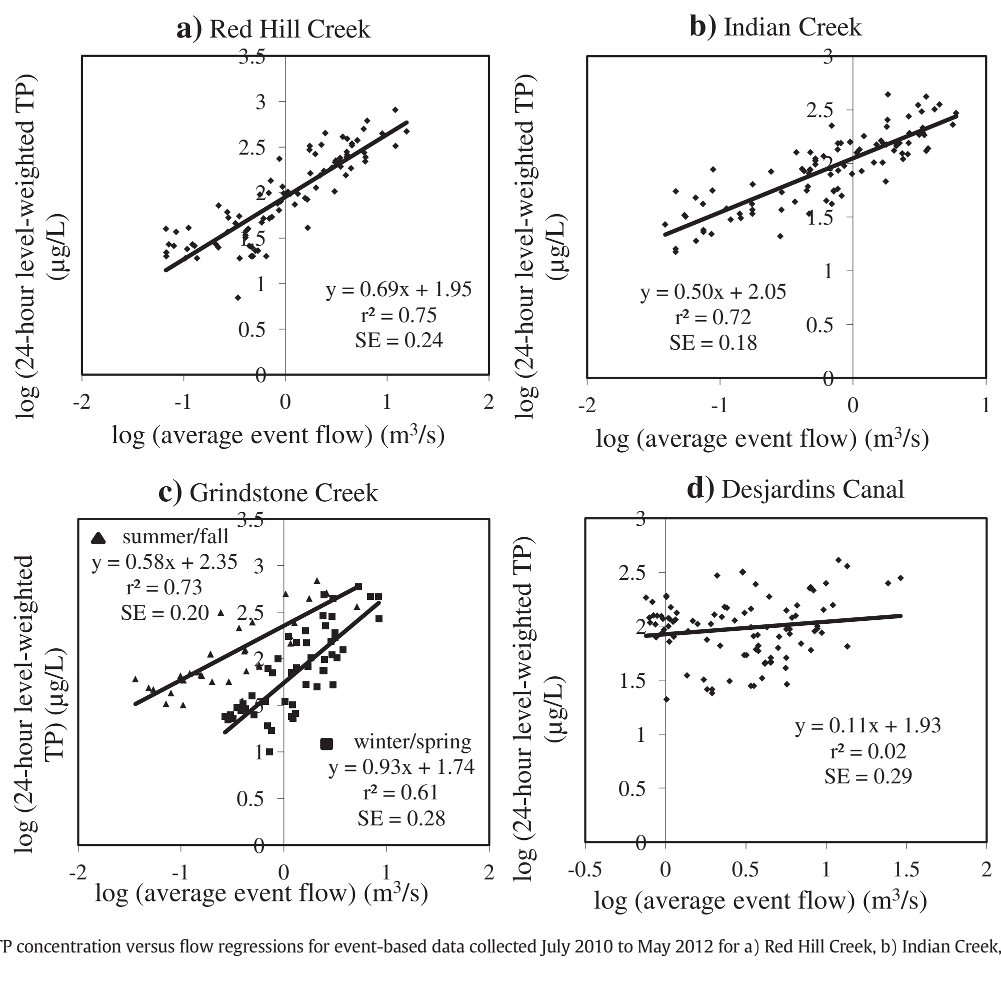


Fig. 3. Log-transformed TP concentration versus flow regressions for event-based data collected July 2010 to May 2012 for a) Red Hill Creek, b) Indian Creek, c) Grindstone Creek, and d) Desjardins Canal.

Grindstone Creek was not strong ($r^2 = 0.26$); however, splitting the data for this station into two seasonal bins (summer/fall: June to October; and winter/spring: November to May) resulted in two strong regressions (Fig. 3c). The summer/fall bin reflected the period where precipitation falls exclusively as rain, and the winter/spring bin always included the spring freshet. Further, the delineation of similar seasonal bins was used previously in a runoff modelling study of Grindstone Creek (Wellen et al., 2014b) and was also found to describe TP concentration versus flow relationships in a similar study on an agricultural watershed located approximately 100 km to the north (O'Connor et al., 2011). All log log TP concentration versus flow regression equations shown in Fig. 3a–c for Red Hill Creek, Indian Creek, and Grindstone Creek are statistically significant ($p < 0.05$). For Grindstone Creek, the slopes of the summer/fall and winter/spring regression equations were significantly different (ANCOVA, $F = 8.4$, $p < 0.05$).

The log-transformed TP concentration versus flow relationship for all event-based data collected at the Desjardins Canal station during July 2010 to May 2012 did not yield a strong regression and was not statistically significant ($r^2 = 0.02$; $p = 0.14$; Fig. 3d). This finding is consistent with the location of this station at the outlet of the Cootes Paradise wetland, where flow dynamics are not primarily controlled by storm events as they would in a true tributary. A sine wave equation developed by visual inspection describes the seasonal TP concentration trend of the highest TP concentration in summer and the lowest in winter at the Desjardins Canal station:

$$[TP] = 50 \times \sin\left(\frac{2\pi \times \mathrm{DOY}}{365} + 136\right) + 88 \tag{2}$$

where DOY is day of year. Factors hypothesized to be contributing towards the sinusodial seasonal relationship include local ecological and/or hydrological processes as described in Long et al. (2014). The temporal TP concentration trends observed in this study are believed to be representative of long-term seasonal TP concentration trends at the Desjardins Canal. May to November water quality monitoring data collected by the Royal Botanical Gardens (RBG, Hamilton ON) from 2004 to 2013 generally demonstrate the lowest TP concentrations in May, an increase in summer, followed by a decline in September back to spring levels (RBG, unpublished data).

For many events, TP concentrations at the Desjardins Canal were highly elevated relative to those expected according to the sine wave model. Data obtained from the City of Hamilton on the dates of known CSO events to Cootes Paradise during the July 2010 to May 2012 period suggested that many of these unusually high TP concentrations were driven by CSO overflow events (see ESM Appendix S1). A relatively strong ($r^2 = 0.44$) and statistically significant ($p < 0.05$) linear regression was formed between log-transformed TP concentrations on days of known CSO events and log-transformed daily precipitation totals measured at the Government of Canada's RBG meteorological station (National Climate Data and Information Archive, http://climate.weatheroffice.gc.ca/Welcome_e.html, last Accessed April 30, 2014) (Fig. 4a). This regression equation was used to reproduce TP concentration pulses at the Desjardins Canal from CSO events on days that exceeded a daily precipitation threshold of 15 mm, the approximate minimum rainfall amount observed to have occurred during known CSO events outside of the spring melt period.

Many elevated TP concentrations during the winter and spring melt period were not described well either by the sine wave model [Eq. (2)] or the TP concentration versus precipitation regression (Fig. 4a). During the December to April period when melting of the snowpack occurs, a relatively strong ($r^2 = 0.47$) and statistically significant ($p < 0.05$) linear regression was formed between log-transformed TP concentrations versus flow at the Desjardins Canal (Fig. 4b). Important to note is that although the r^2 values for both the Desjardins Canal linear regression models were less than the specified threshold of 0.5 (Fig. 4), predictability of TP concentrations at the same location was improved relative to scenarios where these additional equations were not used to improve the predictive capacity of the sine wave model alone [Eq. (2)].

The performance assessment of the empirical models derived from data collected July 2010 to May 2012 from all four stations demonstrated that the models performed very well for describing the TP loads estimated through measured concentrations and flows (NSE ≥ 0.76; RE ≤ 0.42; $r^2 \geq 0.8$; Table 3; Fig. 5). This is especially true if we consider that estimates of particulate flux tend to be less precise for small-sized watersheds due to higher variability (Horowitz, 2013; Moatar et al., 2006). Validation of the models with an independent dataset was not possible as there are no other year-round, event-based datasets for these watersheds.

The annual average TP loads for 2008 to 2012 as estimated by the regression models developed through the July 2010 to May 2012 event-based data were the highest for the Desjardins Canal (17.4 kg/d to 65.6 kg/d), followed by Red Hill Creek (6.4 kg/d to 25.8 kg/d) and/or Grindstone Creek (3.4 kg/d to 33.4 kg/d), and finally Indian Creek (3.0 kg/d to 7.9 kg/d) (Table 4). Generally, the 2008 loads were the highest and the 2012 loads the lowest, reflecting total annual precipitation in 2008 that was 114% relative to the 1981 to 2010 normal at the Government of Canada's RBG meteorological station and precipitation in 2012 that was only 66% of the normal (National Climate Data and Information Archive, http://climate.weatheroffice.gc.ca/Welcome_e.html, last accessed April 30, 2014). In contrast, the 2009, 2010, and 2011 annual precipitation totals were 97% to 101% of the 1981 to 2010 climate normal. For 2008 to 2012, 89%, 73%, and 78% of the annual loads from Red Hill Creek, Indian Creek, and Grindstone Creek, respectively, were delivered to Hamilton Harbour in 10% of the time. For the Desjardins Canal, only 52% of the annual load was delivered in 10% of the time due to the less variable nature of TP loads from the Cootes Paradise wetland relative to the creeks.

When the annual average TP load estimates were normalized by watershed area, the annual average areal TP loads were

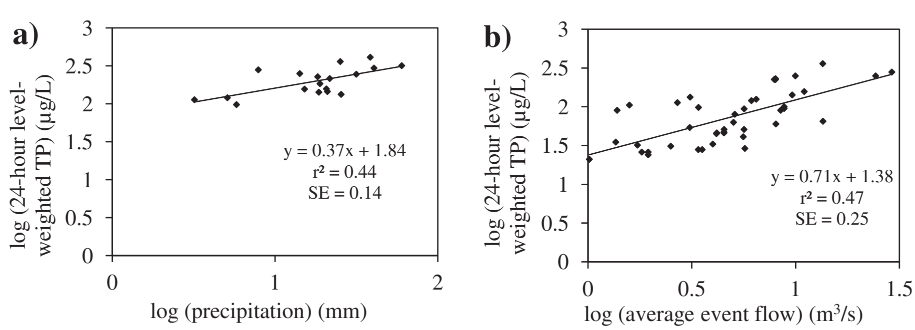


Fig. 4. Regressions for log-transformed 24-hour level-weighted TP concentrations collected July 2010 to May 2012 at the Desjardins Canal with a) log-transformed daily total precipitation at the Government of Canada's Royal Botanical Gardens (RBG) Meteorological Station on days of known CSO events, and b) log-transformed average event flow at the Desjardins Canal station during the December to April period.

Table 3
Assessment of model performance for estimating TP loads (kg/d) for 87 periods sampled during the July 5, 2010 to May 8, 2012 study period. Headings for columns are as follows: Meas is Mean daily TP load based on measured level weighted TP concentrations and flow; Mod is mean daily TP load based on empirical regression models developed from event-based data collected July 2010 to May 2012; NSE is index of model efficiency; RE is relative error; RMSE is root mean square error and Meas:Mod regress is linear regression between Meas and Mod TP loads.

	Meas	Mod	NSE	RE	RMSE	Meas:Mod regress
Red Hill Creek	58.3	60.4	0.82	0.34	53.4	TP load$_{(mod)}$ = 0.99 (TP load$_{(meas)}$) + 2.5 $r^2 = 0.85$
Indian Creek	20.9	20.7	0.86	0.29	12.2	TP load$_{(mod)}$ = 0.85 (TP load$_{(meas)}$) + 2.9 $r^2 = 0.86$
Grindstone Creek	31.7	32.3	0.76	0.42	30.4	TP load$_{(mod)}$ = 0.96 (TP load$_{(meas)}$) + 2.0 $r^2 = 0.80$
Desjardins Canal	57.1	61.1	0.91	0.29	32.1	TP load$_{(mod)}$ = 0.95 (TP load$_{(meas)}$) + 7.0 $r^2 = 0.91$

the highest at the two most urban watersheds of Red Hill Creek (0.36 kg/ha/year to 1.4 kg/ha/year; mean of 1.1 kg/ha/year) and Indian Creek (0.48 kg/ha/year to 1.3 kg/ha/year; mean of 1.0 kg/ha/year) (ESM Table S1). Relatively lower annual average areal TP loads were estimated for the more agricultural and rural watersheds of Grindstone Creek (0.14 kg/ha/year to 1.4 kg/ha/year; mean of 0.80 kg/ha/year) and the Desjardins Canal (0.22 kg/ha/year to 0.83 kg/ha/year; mean of 0.56 kg/ha/year).

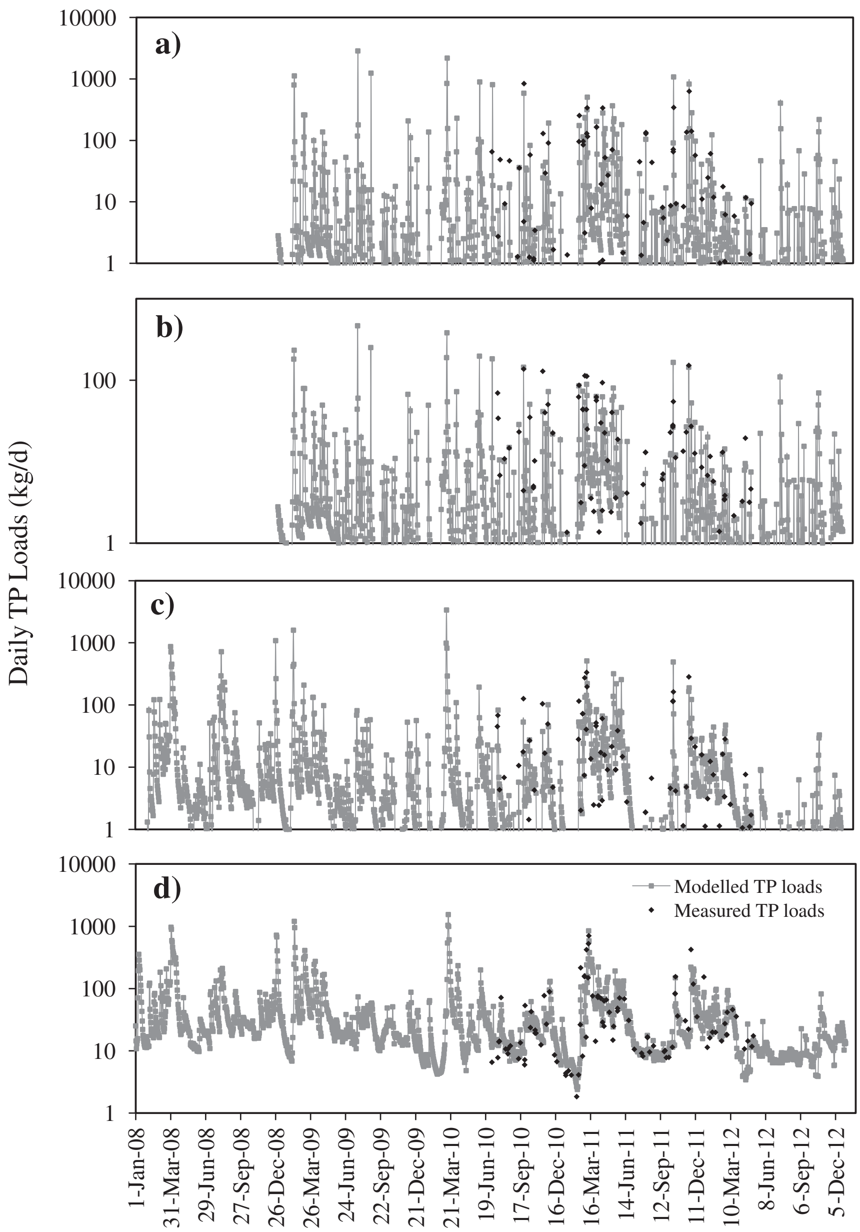


Fig. 5. TP loads estimated through measured TP concentrations and flow during 87 sampling periods between July 2010 and May 2012 (measured TP loads) and through empirically derived equations developed in this study (modelled TP loads) for a) Red Hill Creek, b) Indian Creek, c) Grindstone Creek, and d) Desjardins Canal. There is some degree of error between measured and modelled TP loads as measured daily TP loads are based on 24-hour periods corresponding to event sampling, whereas modelled daily TP loads are based on consistent midnight to midnight time intervals.

Table 4
2008 to 2012 annual average TP loads (kg/d) to Hamilton Harbour (and 95% confidence intervals) as estimated based on (A) regression models developed from event-based data collected July 2010 to May 2012; (B) HH RAP loading estimation methods (HH RAP, 2010; ESM Appendix S1); and (C) regression models developed from 2008 to 2012 data collected through the PWQMN programme.

	Red Hill Creek			Indian Creek		Grindstone Creek			Desjardins Canal	
	A	B	C	A	B[a]	A	B	C	A	B[b]
2008	n/a	n/a	n/a	n/a	n/a	33.4 (11.4)	44.3 (18.5)	18.8 (4.9)	65.6 (11.5)	49.9
2009	25.8 (18.4)	24.2 (10.4)	7.5 (3.9)	7.9 (3.4)	8.1 (3.5)	17.2 (9.4)	23.5 (12.9)	10.6 (3.7)	50.4 (9.9)	44.1
2010	21.7 (14.6)	21.0 (9.3)	6.5 (3.3)	7.1 (2.9)	7.0 (3.1)	22.2 (19.4)	24.7 (18.6)	10.7 (6.4)	39.6 (12.4)	36.4
2011	23.0 (9.1)	28.5 (8.3)	7.8 (2.4)	7.1 (1.8)	9.5 (2.8)	19.5 (5.3)	28.3 (9.9)	12.7 (2.7)	48.9 (8.0)	41.8
2012	6.4 (2.9)	10.4 (3.4)	2.8 (0.89)	3.0 (0.88)	3.5 (1.1)	3.4 (0.7)	5.2 (1.1)	3.2 (0.5)	17.4 (1.6)	25.8

[a] The HH RAP method does not explicitly include load estimates for Indian Creek (HH RAP, 2010; ESM Appendix S1); however, the HH RAP applies a 4/3 area ratio to Red Hill Creek TP loads to account for additional TP loads that enter the Harbour from other smaller creeks, such as Indian Creek, as shown here for illustrative purposes.
[b] 95% confidence intervals cannot be calculated for annual average DC loads following the HH RAP method (see ESM Appendix S1).

HH RAP method

The 2009 to 2012 total Harbour annual average TP loads estimated through the HH RAP method were characterized relatively well as were 99%, 98%, 110%, and 149%, respectively, of the annual average TP loads estimated through the regression methods developed from the event-based monitoring data (Table 4). Similar to the spatial trends described for the event-based regression methods, the annual average TP loads estimated through the HH RAP method were also the highest for the Desjardins Canal (25.8 kg/d to 49.9 kg/d), followed by Red Hill Creek (10.4 kg/d to 28.5 kg/d), Grindstone Creek (5.2 kg/d to 44.3 kg/d), and other small creeks to the Harbour, represented here by Indian Creek (3.5 kg/d to 9.5 kg/d) (Table 4). The largest discrepancy between the two methods was for the dry year of 2012, when the TP loads estimated through the HH RAP method were overestimated by close to 50%.

Regression models derived from PWQMN data

The water quality samples collected through the PWQMN sampling programme during 2008 to 2012 tended to be biased towards low flow sampling with no extreme flow events captured (Fig. 2e–f). The log log TP concentration versus flow regressions for all 2008 to 2012 PWQMN data were statistically significant ($p < 0.05$) but weak for Red Hill Creek ($r^2 = 0.22$; Fig. 6a) and Grindstone Creek ($r^2 = 0.27$). However, regressions were both strong and statistically significant ($p < 0.05$) for Grindstone Creek when data were split into summer/fall (June to October; $r^2 = 0.55$) and winter/spring (November to May; $r^2 = 0.48$) bins (Fig. 6b). Unlike the event-based dataset, equality of the summer/fall and winter/spring regression slopes for the PWQMN dataset cannot be rejected (ANCOVA, $F = 0.9$, $p = 0.35$). A single regression equation for Red Hill Creek and two seasonal regression equations for Grindstone Creek using the PWQMN dataset were carried forward to remain consistent with the approach taken with the event-based dataset.

The annual average TP loads based on regressions developed from the PWQMN dataset were consistently higher for Grindstone Creek (3.2 kg/d to 18.8 kg/d) relative to Red Hill Creek (2.8 kg/d to 7.8 kg/d) (Table 4), in contrast to the event-based results (except for 2010). The 2008 to 2012 annual average TP loads derived from the PWQMN regressions were significantly lower relative to TP loads derived from the event-based regressions for both Red Hill Creek ($t = 4.0$, $p < 0.05$) and Grindstone Creek ($t = 3.2$, $p < 0.05$). The annual average PWQMN-based TP loads were only 29% to 44% and 48% to 65% of event-based TP loads for Red Hill Creek and Grindstone Creek, respectively, except for the 2012 Grindstone Creek TP loads which were 94% of paired event-based TP loads. Thus, TP loads in both creeks were consistently underestimated when based upon the PWQMN datasets relative to loads estimated through the event-based dataset, although this bias was reduced during the dry year of 2012.

2008 to 2012 seasonal TP loads

The seasonal distribution of TP loads delivered to Hamilton Harbour from the four watersheds demonstrated both intra- and inter-annual variability, as well as many differences in seasonal loading patterns among the watersheds (Fig. 7). Trends at Red Hill Creek and Indian Creek were similar and demonstrated substantial inter-annual variability in the seasonal distribution of annual loads, especially for summer. The summer 2009 TP loads at these two stations were 3 to 21 times higher than summer loads estimated for 2010 to 2012 and contributed to 52% and 40% of the 2009 total annual TP load at Red Hill Creek and Indian Creek, respectively. In contrast, summer TP loads at these two stations for 2010 to 2012 only contributed 2% to 33% of total annual TP loads. The drivers of higher TP loads in summer 2009 were two large precipitation events of 60.6 mm on July 25–26, 2009 and 31.4 mm on August 28–29, 2009 (National Climate Data and Information Archive, http://climate.

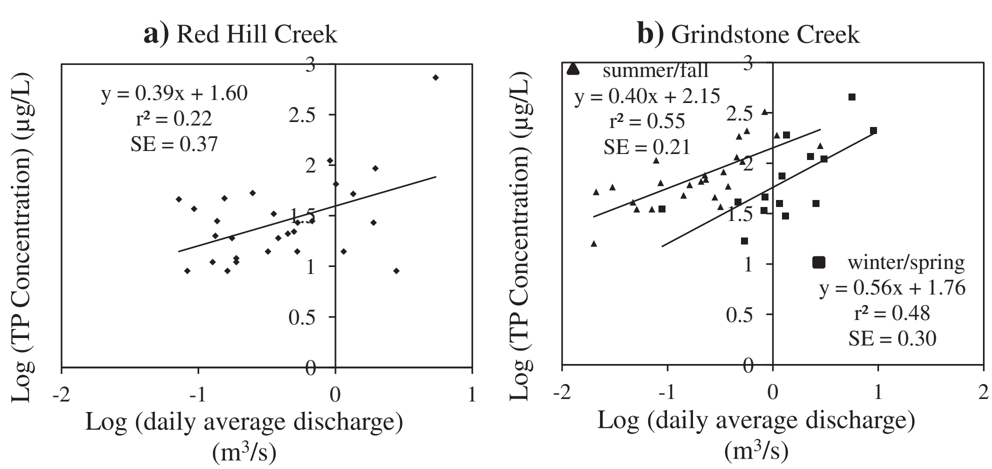


Fig. 6. Log-transformed TP concentration versus flow regressions for PWQMN data collected from a) Red Hill Creek (2009 to 2012), and b) Grindstone Creek (2008 to 2012).

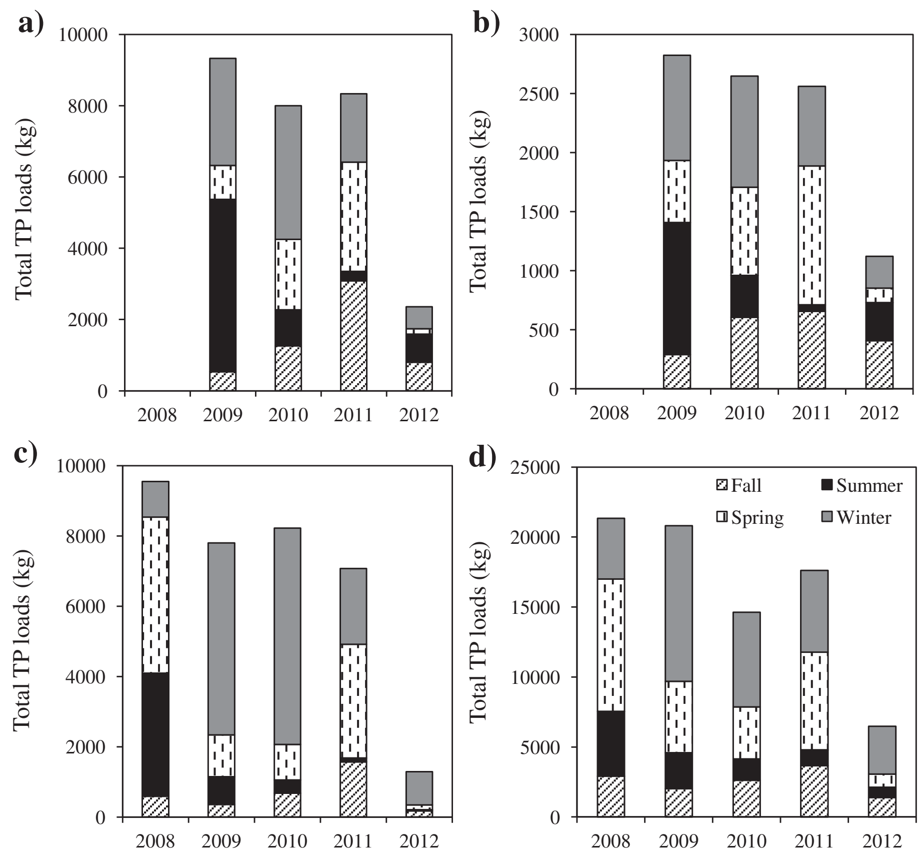


Fig. 7. 2008 to 2012 seasonal distribution of total TP loads to Hamilton Harbour for a) Red Hill Creek, b) Indian Creek, c) Grindstone Creek, and d) Desjardins Canal. Note differences in range of y-axis among panels.

weatheroffice.gc.ca/Welcome_e.html, last accessed April 30, 2014). The two storm events were responsible for 88% and 65% of the total summer TP loads in Red Hill Creek and Indian Creek, respectively, with the July 25–26, 2009 storm event contributing 62% and 42% of summer 2009 TP loads alone.

The seasonal distribution of TP loads at Grindstone Creek was different from that observed at the other stations, but akin to that previously noted for Red Hill Creek and Indian Creek, there was large inter-annual variability. The relatively high TP loads observed during the summer of 2009 in Red Hill Creek and Indian Creek were not observed in Grindstone Creek as the two large storm events did not cause a response in flow in the Grindstone Creek watershed, likely because the storms were localized in nature. On the other hand, the summer 2008 loadings in Grindstone Creek were up to two orders of magnitude larger than TP loads estimated for the summers of 2009 to 2012, reflecting precipitation in 2008 that was 114% the 1981 to 2010 climate normal (National Climate Data and Information Archive, http://climate.weatheroffice.gc.ca/Welcome_e.html, last accessed April 30, 2014) and high summer flows (up to 10 m^3/s) not seen in the summers of 2009 to 2012 (ESM Fig. S2). Nonetheless, in all 5 years examined, the annual TP load was dominated by that contributed through the spring freshet, falling into either the winter season due to an early melt of the snowpack (2009, 2010, 2012) or the spring season due to a delayed melt (2008, 2011).

At the Desjardins Canal, the seasonal distribution of TP loads was similar to that observed at Grindstone Creek as the winter and/or spring TP load dominated the annual total. In contrast to the other three stations, however, inter-and intra-annual variability in seasonal TP loads was comparatively diminished. For example, summer TP loads at the Desjardins Canal in 2008 to 2012 comprised 6 to 22% of annual totals; however, the range was relatively wider at Red Hill Creek (3% to 52%), Indian Creek (2% to 40%), and Grindstone Creek (2% to 37%). Such results are likely due to consistently high background TP concentrations from the wetland during the late spring, summer, and early fall.

Discussion

Hamilton Harbour RAP delisting target for creeks

Using the series of relatively simple, empirical TP loading estimation methods developed from event-based monitoring data, we estimated the 2008 to 2012 annual average TP loads to Hamilton Harbour from the four major tributaries. In 2009 to 2011, the total estimated TP loads from all four watersheds were well above the HH RAP TP loading target for the creeks of 65 kg/d (HH RAP, 1992; ESM Appendix S1) but were below the target in 2012. The total Harbour TP load for 2008 was likely even greater than that for 2009 given the relatively high 2008 estimates for Grindstone Creek and the Desjardins Canal. As precipitation totals in 2009 to 2011 were consistent with the long-term precipitation average in the Hamilton area, the TP loads from the creeks delivered during these 3 years may reflect more typical annual loads from the watersheds, suggesting that the HH RAP target is not consistently being met. While the Hamilton Harbour watershed TP load estimate is of direct relevance to the status of the Hamilton Harbour AOC due to its recognized contribution to eutrophication of the Harbour, TP loading intensity is of interest to the HH RAP and also relevant to other remedial action plans in the Great Lakes basin. Areal TP loads estimated for each watershed can help determine if TP loads are regionally variable given the predominant land use, or alternatively, if TP loads are consistent with those expected in well-managed watersheds. Comparisons of this nature will help remedial action plans determine the nature of mitigation efforts needed in each watershed to reach ambient TP goals in the AOCs.

Regional context of Hamilton Harbour watershed TP loads and implications to mitigation

Areal TP loads for Hamilton Harbour watersheds were consistent with those recently estimated in both impaired AOC watersheds and in non-AOC watersheds of similar land use (ESM Table S1). The 2009 to 2012 areal TP loads in Red Hill Creek and Indian Creek were similar to those estimated in other urban watersheds which ranged from 0.15 kg/ha/year (Duan et al., 2012) to 1.6 kg/ha/year (Boyd, 1999). Similarly, 2008 to 2012 areal TP loads in Grindstone Creek were within the range for other agricultural watersheds of 0.01 kg/ha/year (Diamond, 2011) to 1.89 kg/ha/year (OMOE, 2012b). The 2008 to 2012 areal TP loads for the Desjardins Canal were also within this range, although areal TP loading estimates at this station are not directly comparable to any land-use category due to the strong influence of wetland processes, input of a WWTP, and CSOs upstream from this station.

Although these results illustrate that TP loads in the local watersheds are not anomalously high in a regional context, mitigation is still needed to reduce TP loads to Hamilton Harbour. Annual precipitation is a large determinant in the total annual TP load, a factor beyond the control of local management actions, yet TP loading intensity is still amenable to reduction as less nutrient mass on the landscape equates to less that can be transported by a given amount of precipitation, and a relatively lower areal TP load. Ultimately, it needs to be demonstrated that all reasonable efforts have been made to reduce TP loads to the Hamilton Harbour AOC prior to delisting, especially if algal blooms in the Harbour continue to remain problematic. How to reduce TP loads is an ongoing challenge in the Great Lakes basin given these same issues, however, lessons learned in this study may have broad scale applications especially in areas with similar land use to the mixed but predominantly urban Hamilton Harbour AOC.

Due to the traditional viewpoint that agricultural export rates are higher than those from urban areas (Moore et al., 2004; Soldat and Petrovic, 2008; Soldat et al., 2009), many watersheds in the Great Lakes basin including those in this study have focused on reducing agricultural sources of TP though implementation of best or beneficial management practices (BMPs). In contrast, higher urban TP loads relative to agricultural TP loads found in this study are consistent with results of recent modelling studies on Hamilton Harbour (Wellen et al., 2014a) as well as the Bay of Quinte Area of Concern (Kim et al., Submitted for publication a; Kim et al., Submitted for publication b) and other empirical studies (Beaulac and Reckhow, 1982; Duan et al., 2012; Rast and Lee, 1983; Winter and Duthie, 2000). Such data suggest that a strong focus should be placed on actively mitigating urban TP sources in the Hamilton Harbour AOC, and potentially other AOCs.

High urban TP loading has been attributed to storm water-induced bank erosion as well as a lack of phosphorus retention on impervious surfaces (Withers and Jarvie, 2008) and potentially high fertilizer use on residential lawns (Pfeifer and Bennett, 2011). In the Red Hill Creek watershed, a substantial portion of the TP load can likely be attributed to CSO inputs, as all large loading events (>100 kg/d) in Red Hill Creek coincided with CSO events, seemingly independent of precipitation amount (ESM Fig. S1). The contribution of TP from CSOs, however; is expected to decline given the ongoing upgrades being made to the City of Hamilton combined sewer system. Other potential mitigation measures in the urban systems could include low impact development (LID) practices such as bioretention, permeable pavement, and swales which have all been demonstrated to reduce offsite TP transport (Ahiablame et al., 2012).

TP mitigation in urban areas is particularly important given that these areas tend to respond strongly to large storm events. Catchment storage is bypassed in urban areas such as Red Hill Creek, resulting in higher surface runoff generation relative to more agricultural watersheds like Grindstone Creek (Wellen et al., 2014b). In the July 2010 to May 2012 monitoring period of our study, four of the five largest daily precipitation totals occurred during fall. Although it is not known if this observation is representative of projected seasonal precipitation trends with climate change, it is of relevance if we consider that the lack of leaves on trees results in a greater proportion of rainfall becoming runoff relative to a scenario under full leaf canopy and associated biological uptake of water (Kim et al., 2014). Thus, the predicted increase in storm event intensity in the region (Kunkel et al., 2013), especially if such storms occur during the fall as observed during our study, suggests that the TP loads in urban watersheds may be particularly prone to increases with climate change.

An increase in extreme events with climate change may further exacerbate the issue of urban TP loading, as hydrological behaviour, and hence TP loads, may change above certain discharge or precipitation thresholds. In a hydrological modelling study conducted on Red Hill Creek and Grindstone Creek, Wellen et al. (2014b) hypothesized that the watershed responses to precipitation differed above a threshold corresponding to an extreme state. A greater amount of rainfall was converted to surface runoff relative to the proportion of runoff generated below the critical threshold. The vulnerability of urban areas to TP load increases with changing meteorological conditions is particularly problematic considering the trend of increasing urbanization in areas around the Great Lakes and globally.

Differences in TP loads estimated through three methods examined in this study

A large part of determining mitigation measures that will result in meaningful change in watersheds is the use of accurate TP loads in the assessments. Although TP loads would ideally be estimated using continuously measured TP concentrations and flow, this is not economically feasible. Therefore a loading estimation method must be chosen, and in this study, three loading estimation methods were compared. Annual average TP loads for Red Hill Creek and Grindstone Creek estimated from regressions derived from data collected through the PWQMN programme were generally two to three fold lower relative to TP loads estimated from regressions derived from event-based data. The underestimation bias from use of PWQMN data for estimating TP loads was also found in a similar comparison made on Duffins Creek, a tributary to the north shore of Lake Ontario (Booty et al., 2013). Thus, care must be taken not only with the loading method selection but also with the nature of the dataset used, given that the same regression methods were applied to both the event-based and grab sample data in this study. This is important considering that in many areas where event-based data are not available, tributary TP loads are updated through monthly grab sample data which could be resulting in a systematic and pervasive underestimation of TP loads.

The difference between TP loads estimated in this study relative to those estimated through use of PWQMN data emphasizes the importance of characterizing TP concentrations above baseflow. TP concentrations during high flow are inadequately represented in the PWQMN database (Fig. 2e–f). The bias in estimating TP loads for Grindstone Creek from the PWQMN dataset was less for 2012, likely because it was a dry year and hence, characterization of TP concentrations or loads during high flow conditions would have been comparatively less important relative to other, more wet years. Characterizing TP concentrations during high flow conditions is essential in the establishment of accurate concentration versus flow relationships and subsequently, TP load estimates. It is the brief but intense events which occurred less than 10% of the time when 52% to 89% of TP loads were delivered to Hamilton Harbour from its tributaries.

A surprising outcome of this study was that total Harbour annual average TP loads were more accurate when estimated through the outdated and uncertain HH RAP method relative to annual average TP loads estimated through recently collected PWQMN data. This comparison is somewhat deceiving, however, as the source apportionment breakdown reveals that a series of over-and underestimations of

TP loads from the four watersheds as estimated through the HH RAP method leads to the appearance of accuracy at the scale of total Harbour TP load (see following discussion). Also, the explicit exclusion of Indian Creek from the HH RAP method is a major data gap considering the high urban land use of this watershed and the resulting turbidity plumes in the northeast corner of the Harbour following major storm events.

For Red Hill Creek, the annual average TP loads estimated through the HH RAP method were within an acceptable margin of error (Table 4), reflecting relatively well-characterized TP concentrations during high flow conditions (>4 m^3/s; HH RAP TP = 500 μg/L; event-based TP = 372 μg/L (n = 18)). Caution would still need to be exercised for use of the HH RAP method at Red Hill Creek, however, as it overestimates TP concentrations by a factor of two during low flow conditions (<4 m^3/s; HH RAP TP = 190 μg/L; event-based TP = 97 μg/L (n = 74)).

At Grindstone Creek, annual average TP loads were consistently overestimated by the HH RAP method by approximately 50%. Like Red Hill Creek, TP concentrations during all flow conditions were overestimated by the HH RAP method in Grindstone Creek, but were particularly overestimated during peak flow conditions. The peak flow TP concentration of 1190 μg/L used in the HH RAP method is over three times the event-based mean TP concentration of 401.7 μg/L (n = 3) measured in this study. This makes use of the HH RAP method at Grindstone Creek particularly problematic since it is during peak flow when the majority of the annual TP loading occurs, thus explaining the overestimated annual average TP loads in this watershed.

In contrast to Red Hill Creek and Grindstone Creek, annual average TP loads estimated at the Desjardins Canal through the HH RAP method were only 76% to 92% of TP loads estimated through the event-based dataset, except during the dry year of 2012 when HH RAP TP loads were 148% of those estimated though this study. The relative differences in TP loads between methods are likely on account of inadequate representation of TP concentration dynamics in the HH RAP method. The July 2010 to May 2012 event-based monitoring demonstrated that high TP concentrations are measured at the Desjardins Canal following large precipitation events and spring freshet, resulting in high TP loads intermittently delivered to the Harbour. Intermittent pulses in TP loads are not accounted for in summer ambient monitoring in the centre of Cootes Paradise, the dataset on which the HH RAP loading method is based. Thus, the overestimation bias in annual average TP loads at Grindstone Creek is counteracted by a general underestimation bias at the Desjardins Canal to produce little apparent overall bias in the total Harbour annual average TP loads through use of the HH RAP method. Although the erroneous HH RAP TP loads may be negligible at the scale of a total Harbour loading, use of these loads has implications for our ability to understand how local areas of Hamilton Harbour may respond to inputs from each of the watersheds, or how the Harbour may be expected to respond to watershed loads delivered either during dry or wet weather.

In addition to the site specific concerns stemming from use of the HH RAP method to determine TP loads for the watersheds of Hamilton Harbour, lessons learned from evaluating this method are applicable beyond Hamilton Harbour. The apparent accuracy of a method can be misleading if based upon a problematic metric such as an annual average, as it can lead to an inaccurate understanding of the sources and impacts of tributary TP loads which are important in establishing potential mitigation measures. For example, use of an annual average does not reflect the pulse nature of tributary inputs. Also, a few large events a year can substantially increase an annual average, effectively eradicating any observation of progress which would have otherwise been made through the implementation of management actions. Additionally, an annual average also does not indicate what events or times of the year are most problematic in each watershed, important in focusing mitigation efforts.

Our study demonstrated that the most problematic high flow TP loading events in agricultural Grindstone Creek and the Desjardins Canal were due to the spring freshet, whereas rain storm events had the highest event TP loads in urban Red Hill Creek and Indian Creek. This distinction, made possible through daily rather than annual average TP load estimates points to the need for more stormwater-mitigation strategies in the urban watersheds with flashy hydrology and for further study into ways that early spring TP loading can be reduced in agricultural watersheds. Use of a higher resolution TP loading estimation method will also be able to better identify problematic seasonal TP loads. While there was only one TP concentration versus flow regression each for Red Hill Creek and Indian Creek, there was a distinct difference in TP mobilization behaviour between summer/fall and winter/spring in Grindstone Creek. Reasons for this could be process-based, source-based, or even hydrology-based, because modelling work by Wellen et al. (2014b) hypothesized that the dominant source of water to Grindstone Creek shifted during the growing season to almost exclusively urban runoff, even given the relatively small proportion of this land use in this watershed.

Use of a higher resolution TP loading estimation method will also be able to better identify the role of climate change on any changes to seasonal loading patterns in watersheds across the Great Lakes basin. This is especially true for winter given the distinctly different precipitation and temperature patterns observed between winter 2010 and 2011 in this study. Most of the winter TP loads in 2011 were comprised of a spring freshet loading, whereas sporadic winter runoff events comprised much of the winter TP loads in 2012. Use of a daily or seasonal metric in a watershed such as Hamilton Harbour will make it possible to draw potential causal connections between changes in winter TP loads from the watersheds, the magnitude of the spring algal bloom, and the subsequent impact of the autochthonous material on the summer hypoxia patterns in the Harbour (Gudimov et al., 2010).

Recommendations

The empirical TP loading estimation methods based on the July 2010 to May 2012 event-based monitoring programme are recommended for use by the HH RAP for estimating TP loads from the watersheds of Hamilton Harbour. This series of revised methods has increased accuracy over the previously endorsed method and produces TP loading data at a meaningful temporal resolution, beneficial in determining daily or seasonal TP loads.

In addition to revising the HH RAP TP loading estimation method, event-based water quality monitoring similar to methods undertaken in 2010 to 2012 is also recommended to be repeated in future years. The TP loading estimation method in this study is based on a series of static relationships, as such, the regressions established for the years 2010 to 2012 will need to be reexamined, especially as remedial measures in the watersheds are implemented. In particular, the frequency and volume of CSO events are expected to decline as several additional CSO holding tanks have been brought online by the City of Hamilton since the beginning of the study. Recalibration of relationships is particularly important for the Desjardins Canal station given that CSO events were explicitly accounted for through the TP concentration versus precipitation regression, and this station is estimated to have the largest of the tributary TP loads to Hamilton Harbour. Future event-based monitoring will not only help to ensure ongoing accuracy in TP load estimation but also to ensure that any nutrient management actions implemented are having the desired effect. Some natural, inherent inter-annual variability in the relationships is expected (Horowitz, 2013); however, this should be distinguished from loading increases or decreases which are a result of anthropogenic factors.

Finally, the revised TP load estimates should be examined for their implications on the predicted achievement of the HH RAP ambient TP goal for Hamilton Harbour, as well as the predicted biological response. A Bayesian set of eutrophication models demonstrated that Hamilton Harbour should meet the ambient TP goal of 20 μg/L but with the caveat that the exogenous TP loads used in the simulations have been well characterized (Ramin et al., 2012; Gudimov et al., 2011). TP loads input into these models were based on the HH RAP method (HH RAP, 2004), and follow-up modelling work on the Harbour by Kim et al. (2014) used TP loads for Red Hill Creek (2.7 kg/d to 9.4 kg/d) and Grindstone Creek (3.4 kg/d to 11.5 kg/d) which were approximately two to three fold lower than estimates made in this study. The change to both annual loading estimates and the resolution of the TP loads to daily estimates has the potential to change the outlook for the Harbour, and problematic large-loading events previously not explicitly identified can continue to be addressed through watershed management programmes.

Conclusions

Following this study, the HH RAP has an improved understanding of the TP loads entering Hamilton Harbour from the four watersheds due to the development of a simple, empirically-based TP loading estimation method for each of the watersheds. By increasing the resolution of TP loads from the tributaries, water managers are able to more fully understand the seasonal and event-based loading patterns in each watershed. This information may be critical in assisting remedial action plans across the Great Lakes basin to determine potential management actions in the watersheds necessary to achieve AOC delisting goals. Likewise, lessons learned in this study may prompt remedial action plans to reevaluate and potentially revise any current TP loading targets for their watersheds, such as the HH RAP 65 kg/d target. An important consideration in an assessment of this nature is the ongoing reduction of TP loads from the wastewater treatment plants; as the major point source contributions to Hamilton Harbour and other areas of the Great Lakes become proportionally less with technological improvements, there is an increasing likelihood that inputs from the watersheds may modulate the nutrient dynamics of the nearshore in the Great Lakes. Adding to the challenge in an area such as Hamilton Harbour – the recipient of TP inputs from a mixed land use 465-km^2 area – remedial action plans will need to consider what loading reductions are achievable in the watersheds given the current land use and the regional nature of nutrient issues in urban and agricultural tributaries. The strong role of climatic conditions on tributary TP loads and what remedial measures are achievable through local management actions will also need to be considered so that efforts are made where they are expected to make the most positive change in both the watersheds and the Great Lakes.

Acknowledgments

The authors thank Dave Supper at the OMOECC for monitoring station installation and sample collection, and the numerous staff and summer students who assisted with sample collection, preparation, and analysis. We also wish to thank Mushtaq Hussain who provided the Ontario PWQMN data, as well as Derek Smith, Veronique Hiriart-Baer, Joe Makarewicz and two anonymous reviewers for helpful comments on an earlier draft of the manuscript. Logistical and other data support from Mark Bainbridge (City of Hamilton), Tys Theysmeyer (Royal Botanical Gardens), Tom Arsenault (Water Survey of Canada branch of Environment Canada), Conservation Halton, and Hamilton Conservation Authority is also much appreciated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jglr.2015.04.001.

References

Ahiablame, L.M., Engel, B.A., Chaubey, I., 2012. Effectiveness of low impact development practices: literature review and suggestions for future research. Water Air Soil Pollut. 223, 4253–4273.

Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Mar. Ecol. Prog. Ser. 271, 13–26.

Aulenbach, B.T., Hooper, R.P., 2006a. The composite method: an improved method for stream-water solute load estimation. Hydrol. Process. 20, 3029–3047.

Aulenbach, B.T., Hooper, R.P., 2006b. Improving stream solute load estimation by the composite method: a comparative analysis using data from the Panola Mountain Research Watershed. Proceedings of the 2005 George Water Resources Conference.

Baldwin, A.K., Robertson, D.M., Saad, D.A., Magruder, C., 2013. Refinement of regression models to estimate real-time concentrations of contaminants in the Menomonee River drainage basin, southeast Wisconsin, 2008–11. U.S. Geological Survey Scientific Investigations Report 2013-5174. U.S. Geological Survey, Reston, Virginia.

Beaulac, M.N., Reckhow, K.H., 1982. An examination of land use-nutrient export relationships. Water Resour. Bull. 18, 1013–1024.

Booty, W.G., Wong, I., Bowen, G.S., Fong, P., McCrimmon, C., Leon, L., 2013. Loading estimate methods to support integrated watershed-lake modelling: Duffins Creek, Lake Ontario. Water Qual. Res. J. Can. 49, 179–191.

Boyd, D., 1999. Assessment of six tributary discharges to the Toronto Area Waterfront, Volume 1: project synopsis and selected results. Report Prepared for The Toronto and Region Remedial Action Plan.

Charlton, M.N., 1997. The sewage issue in Hamilton Harbour: implications of population growth for the Remedial Action Plan. Water Qual. Res. J. Can. 32, 407–420.

Cochran, W.G., 1977. Sampling Techniques. third ed. John Wiley and Sons, New York.

Conservation Halton, 2006. North Shore Watershed Study. Conservation Halton, Milton, Ontario.

Diamond, J., 2011. The Welland River Eutrophication Study in the Niagara River Area of Concern in Support of the Beneficial Use Impairment: Eutrophication and Undesirable Algae. Niagara River RAP, Welland River Eutrophication Study Technical Working Group.

Dolan, D.M., Chapra, S.C., 2012. Great Lakes total phosphorus revisited: 1. Loading analysis and update (1994–2008). J. Great Lakes Res. 38, 730–740.

Dolan, D.M., McGunagle, K.P., 2005. Lake Erie total phosphorus loading analysis and update: 1996–2002. J. Great Lakes Res. 31, 11–22.

Dolan, D.M., Yui, A.K., Geist, R.G., 1981. Evaluation of river load estimation methods for total phosphorus. J. Great Lakes Res. 7, 207–214.

Duan, S., Kaushal, S.S., Groffman, P.M., Band, L.E., Belt, K.T., 2012. Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed. J. Geophys. Res. 117, G01025.

Ferguson, R.I., 1987. Accuracy and precision of methods for estimating river loads. Earth Surf. Process. Landf. 12, 95–104.

Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour: system analysis and evaluation of nutrient loading scenarios. J. Great Lakes Res. 36, 520–539.

Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions: a modeling analysis of the "ecological unknowns". J. Great Lakes Res. 37, 494–506.

Hamilton Harbour Remedial Action Plan (HH RAP), 1992. Goals, options and recommendations. Volume 2 — Main Report. RAP Stage 2. Burlington, Ontario.

Hamilton Harbour Remedial Action Plan (HH RAP), 2004. Contaminant Loadings and Concentrations to Hamilton Harbour: 1996–2002. Burlington, Ontario.

Hamilton Harbour Remedial Action Plan (HH RAP), 2010. Contaminant Loadings and Concentrations to Hamilton Harbour: 2003–2007 Update. Burlington, Ontario.

Hamilton Harbour Remedial Action Plan (HH RAP), 2012. Hamilton Harbour Remedial Action Plan Beneficial Uses 2012 Fact Sheets. Burlington, Ontario.

Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4 (9 pp.).

Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. J. Great Lakes Res. 35, 293–301.

Hirsch, R.M., Moyer, D.L., Archfield, S.A., 2010. Weighted regression on time, discharge and season (WRTDS), with an application to Chesapeake Bay River inputs. J. Am. Water Resour. Assoc. 1–24.

Horowitz, A.J., 2013. A review of selected inorganic surface water quality-monitoring practices: are we really measuring what we think, and if so, are we doing it right? Environ. Sci. Technol. 47, 2471–2486.

Kim, D.K., Zhang, W., Hiriart-Baer, V., Wellen, C., Long, T., Boyd, D., Arhonditsis, G.B., 2014. Towards the development of integrated modelling systems in aquatic biogeochemistry: a Bayesian approach. J. Great Lakes Res. 40, 73–87.

Kim, D.K., Zhang, W., Mugalingam, S., Kaluskar, S., Watson, S., McClure, C., Keene, B., Long, T., Geater, K., Morley, A., Arhonditsis, G.B., 2015a. A Bayesian SPARROW Application for Estimating Phosphorus Export From the Bay of Quinte Watershed (Submitted for publication).

Kim, D.K., Zhang, W., Mugalingam, S., Kaluskar, S., Watson, S., McClure, C., Keene, B., Long, T., Geater, K., Morley, A., Arhonditsis, G.B., 2015b. Towards the Development of an Integrated Modelling Framework for Evaluating the Water Quality in the Bay of Quinte: Watershed Physiography, Land-use Patterns, and Phosphorus Loading Dynamics (Submitted for publication).

Kunkel, K.E., Karl, T.R., Easterling, D.R., Redmond, K., Young, J., Yin, X., Hennon, P., 2013. Probable maximum precipitation and climate change. Geophys. Res. Lett. 40, 1402–1408.

Long, T., Boyd, D., Wellen, C., Arhonditsis, G., 2014. Evaluation of stormwater and snowmelt inputs, land use and seasonality on nutrient dynamics in the watersheds of Hamilton Harbour, Ontario, Canada. J. Great Lakes Res. 40, 964–979.

Macrae, M.L., English, M.C., Schiff, S.L., Stone, M., 2007. Capturing temporal variability for estimates of annual hydrochemical export from a first-order agricultural catchment in southern Ontario, Canada. Hydrol. Process. 21, 1651–1663.

Moatar, F., Meybeck, M., 2005. Compared performances of different algorithms for estimating annual nutrient loads discharged by the eutrophic River Loire. Hydrol. Process. 19, 429–444.

Moatar, F., Person, G., Meybeck, M., Coynel, A., Etcheber, H., Crouzet, P., 2006. The influence of contrasting suspended particulate matter transport regimes on the bias and precision of flux estimates. Sci. Total Environ. 370, 515–531.
Moore, R.B., Johnston, C.M., Robinson, K.W., Deacon, J.R., 2004. Estimation of Total Nitrogen and Total Phosphorus in New England Streams Using Spatially Referenced Regression Models. United States Geological Survey, Pembroke, New Hampshire.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 10, 282–290.
Neter, J., Kutner, M., Wasserman, W., Nachtsheim, C., 1996. Applied Linear Statistical Models. fourth ed. McGraw-Hill College.
O'Connor, E.M., McConnell, C., Lembcke, D., Winter, J.G., 2011. Estimation of total phosphorus loads for a large, flashy river of a highly developed watershed — seasonal and hysteresis effects. J. Great Lakes Res. 37, 26–35.
Old, G.H., Leeks, G.J., Packman, J.C., Smith, B.P., Lewis, S., Hewitt, E.J., Holmes, M., Young, A., 2003. The impact of a convectional summer rainfall event on river flow and fine sediment transport in a highly urbanised catchment: Bradford, West Yorkshire. Sci. Total Environ. 314–316, 495–512.
Ontario Ministry of Natural Resources, 2008. Landcover Greater Toronto Area (GTA) (2002).
Ontario Ministry of the Environment (OMOE), 2010. Method E3367: The Determination of Total Kjeldahl Nitrogen and Total Phosphorus in Water and Precipitation by Colourimetry, Version 2.0. Laboratory Services Branch. Ontario Ministry of the Environment, Toronto, Ontario.
Ontario Ministry of the Environment (OMOE), 2012a. Method E3367: The Determination of Total Kjeldahl Nitrogen and Total Phosphorus in Water and Precipitation by Colourimetry, Version 3.0. Laboratory Services Branch. Ontario Ministry of the Environment, Toronto, Ontario.
Ontario Ministry of the Environment (OMOE), 2012b. Water Quality of 15 Streams in Agricultural Watersheds of Southwestern Ontario 2004–2009. Seasonal Patterns, Regional Comparisons, and the Influence of Land Use. Queen's Printer for Ontario, Toronto, Ontario.
Pfeifer, L.R., Bennett, E.M., 2011. Environmental and social predictors of phosphorus in urban streams on the Island of Montréal, Québec. Urban Ecosyst. 14, 485–499.
Preston, S.D., Bierman, V.J., Silliman, S.E., 1989. Evaluation of methods for the estimation of tributary mass loads. Water Resour. Res. 25, 1379–1389.
Quilbé, R., Rousseau, A.N., Duchemin, M., Poulin, A., Gangbazo, G., Villeneuve, J.P., 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: application to the Beaurivage River (Québec, Canada). J. Hydrol. 326, 209–310.
Ramin, M., Labencki, T., Boyd, D., Trolle, D., Arhonditsis, G.B., 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecol. Model. 242, 127–145.
Rao, A.M., 1979. Pollution from rural, transportation, extractive, and undisturbed land uses in the Grand and Saugeen Watersheds. Task Group C (Canadian Section) Activities 3 and 4 International Reference Group on Great Lakes Pollution from Land Use Activities. Ontario Ministry of the Environment.
Rast, W., Lee, G.F., 1983. Nutrient loading estimates for lakes. J. Environ. Eng. 109, 502–517.
Richards, R.P., 1998. Estimation of pollutant loads in rivers and streams: a guidance document for NPS programs. Prepared for U.S. Environmental Protection Agency, Region VIII.
Richards, R.P., Holloway, J., 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. Water Resour. Res. 23, 1939–1948.
Richards, R.P., Baker, D.B., Kramer, J.W., Ewing, D.E., 1996. Annual loads of herbicides in Lake Erie tributaries of Michigan and Ohio. J. Great Lakes Res. 22, 414–428.
Schindler, D.W., 1977. Evolution of phosphorus limitation in lakes. Science 21, 260–262.
Sharpley, A.N., Daniel, T.C., Edwards, D.R., 1993. Phosphorus movement in the landscape. J. Prod. Agric. 6, 492–500.
Soldat, D.J., Petrovic, A.M., 2008. The fate and transport of total phosphorus in turfgrass ecosystems. Crop Sci. 48, 2051–2065.
Soldat, D.J., Petrovic, A.M., Ketterings, Q.M., 2009. Effect of soil total phosphorus levels on total phosphorus runoff concentrations from turfgrass. Water Air Soil Pollut. 199, 33–44.
Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2014a. Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring. Hydrol. Process. 28, 1260–1283.
Wellen, C., Arhonditsis, G.B., Long, T., Boyd, D., 2014b. Accommodating environmental thresholds and extreme events in hydrological models: a Bayesian approach. J. Great Lakes Res. 40 (S3), 102–116.
Williams, G.P., 1989. Sediment concentration versus water discharge during single hydrologic events in rivers. J. Hydrol. 111, 89–106.
Winter, J.G., Duthie, H.C., 2000. Export coefficient modeling to assess phosphorus loading in an urban watershed. J. Am. Water Resour. Assoc. 36, 1053–1061.
Withers, P.J.A., Jarvie, H.P., 2008. Delivery and cycling of phosphorus in rivers: a review. Sci. Total Environ. 400, 379–395.

Estimation of tributary total phosphorus loads to Hamilton Harbour, Ontario, Canada, using a series of regression equations

Tanya Long[a*], Chris Wellen[b], George Arhonditsis[c], Duncan Boyd[a], Mohamed Mohamed[a], Kristin O'Connor[d]

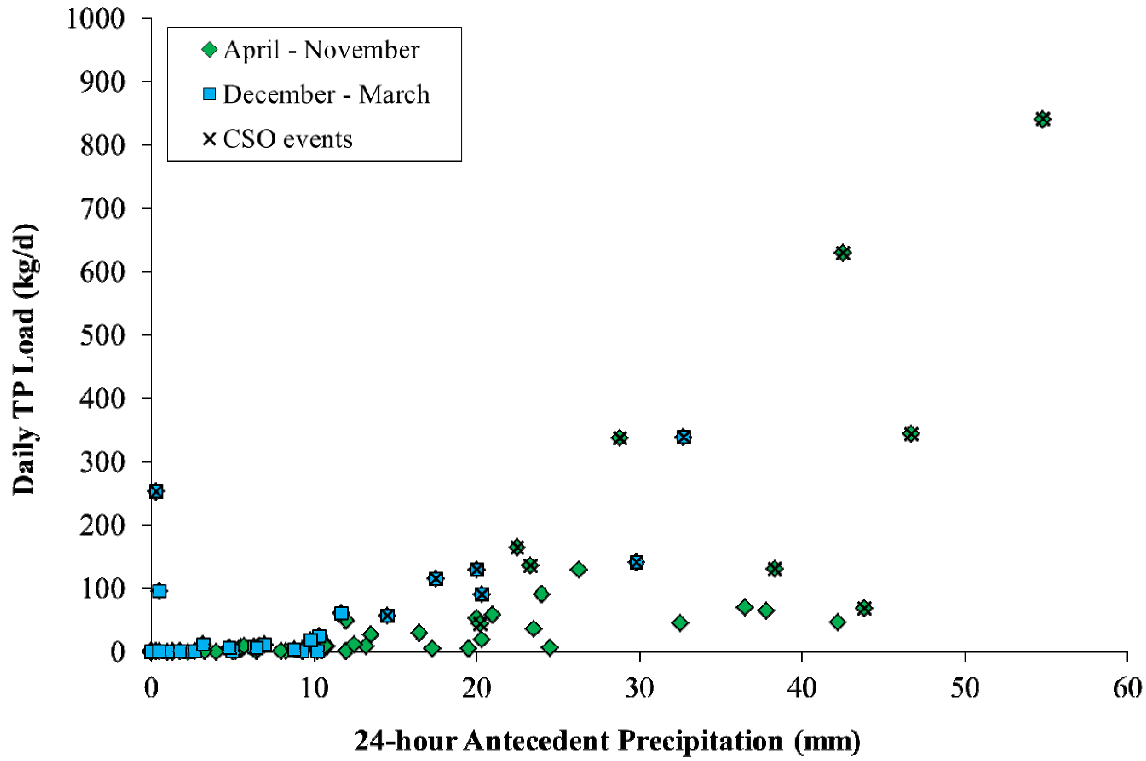
*corresponding author, tanya.long@ontario.ca, 416-235-6247 (telephone), 416-235-6235 (fax)

a – Ontario Ministry of the Environment and Climate Change, Environmental Monitoring and Reporting Branch, 125 Resources Road, Toronto, ON, M9P 3V6, Canada
b – Watershed Hydrology Group, McMaster University, 1280 Main St W, Hamilton, ON, L8S 4L8, Canada
c – Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, ON, M1C 1A4, Canada
d – Hamilton Harbour Remedial Action Plan (HH RAP) Office, Canada Centre for Inland Waters, 867 Lakeshore Road, Burlington, ON, L7S 1A1, Canada

Electronic Supplementary Material (ESM) Appendix S1

Details on the City of Hamilton's Combined Sewer Overflows (CSOs) During the Course of this Study

During the course of this study, the City of Hamilton undertook upgrades to the combined sewer system which improved the number and volume of overflows both to Red Hill Creek and Cootes Paradise. At the end of the study period (May 2012), Red Hill Creek had two CSOs points, located 0.8 km and 4 km upstream from the event-based monitoring station (both now controlled CSO points), and Cootes Paradise had a total of six CSO points (three of which were controlled CSO points). Controlled CSO points are CSOs at which the City of Hamilton has installed infrastructure to mitigate overflow events, such as holding tanks. Of the three controlled CSO points to Cootes Paradise, two discharge to Cootes Paradise via Chedoke Creek (Royal tank, Main/King tank) and one directly to the Cootes Paradise wetland (McMaster tank). Between the July 5, 2010 and May 8, 2012 time period of our study, there were 28 overflow events from the Greenhill tank on Red Hill Creek, 30 overflow events from the Main/King tank, and 21 from the Royal tank; the McMaster tank was not brought online until April 2012 so little overflow data are available on this CSO location for the majority of this study (M. Bainbridge, City of Hamilton, 2014, pers. comm.). Many of these overflow events coincide with sample collection times at the Red Hill Creek (ESM Fig. S1) and the Desjardins Canal stations.



ESM Figure S1: 24-hour antecedent precipitation amount relative to daily total phosphorus (TP) loads estimated through measured 24-hour level-weighted TP concentrations and discharge for 87 sampling periods collected from Red Hill Creek during the July 2010 to May 2012 study period. Winter (December – March) data points are indicated by square symbology, and events which were sampled during known CSO events are indicated by "x" symbology.

Notes: Precipitation data were collected in 15 minute intervals by Hamilton Conservation Authority at their Stoney Creek monitoring station, located approximately 2.6 km from the Red Hill Creek monitoring station. Due to the availability of high resolution precipitation data, precipitation totals were calculated to correspond to precipitation that fell 24 hours prior to sampling in addition to precipitation that fell during the time windows of each sampling events. CSO event data for the Greenhill tanks on Red Hill Creek provided by the City of Hamilton (M. Bainbridge, 2013, pers. comm.).

Serial Correlation of Water Quality at Red Hill Creek and Grindstone Creek

We quantified the hydrologic response time using measured streamflows. Using the daily flows measured at Red Hill and Grindstone Creeks between the years 1988 and 2009, we compute a 1-day correlation coefficient of $\rho = 0.43$ for Redhill Creek and $\rho = 0.83$ for Grindstone Creek. Following Yang et al. (2007a,b), we may transform these estimates of daily correlation to a characteristic correlation time using the equation $\rho = \exp\left(-\frac{\Delta t}{\tau}\right)$, where Δt is the time step (1 day) and τ is the characteristic correlation time in days. This yields $\tau = 1.15$ days for Red Hill Creek, which we rounded up to two days to incorporate some degree of memory in the system. For Grindstone Creek, this method gives us $\tau = 5.37$ days.

One may think of the equation $\rho = \exp\left(-\frac{\Delta t}{\tau}\right)$ as a continuous exponential decay of information in the time series, where the mean lifetime of the information is equal to τ. After the passage of time τ days, the original information content is reduced to 1/e ~ 0.37 of what it was at day 1. Thus after one week (the average gap between collection of samples) there is minimal correlation left in the data.

Time- Versus Level- versus Flow-Weighted Composite Samples

In our study, hourly grab samples corresponding to key points on the hydrograph were submitted for analysis for select rain events in addition to the submission of 24-hour level-weighted composite samples which were analyzed for all 87 events sampled. Two such select events were September 28-29, 2010 (54.8 mm of rain; Hamilton Conservation Authority (HCA) unpublished data), and November 22-23, 2010 (16.5 mm of rain; HCA unpublished data), the former event being the largest event sampled during the July 2010 – May 2012 study. For these two events, we compared the measured level-weighted composite sample TP concentrations to the theoretical time-weighted and flow-weighted composite TP concentrations calculated for each station based on the TP concentrations measured in the grab samples.

Results of this comparison demonstrated that level-weighted composite samples averaged 85% and ranged from 57% to 105% of paired flow-weighted composite samples; the bias was larger for the larger sized event (T. Long, unpublished data). Although the level-weighted composite samples tended to underestimate flow-weighted TP concentrations, this bias is however markedly improved relative to use of time-weighted composite samples which averaged 71% and ranged from 38% to 102% of paired flow weighted composite samples (T. Long, unpublished data).

The use of level-weighted composite samples underestimates TP loads for the same reason that use of time-weighted composites underestimates TP loads. Save for the Desjardins Canal station, there is a strong positive correlation between flow and TP concentration. An event-mean concentration will be more influenced by high flows than low flows, and if the high flows also have a high concentration, then an averaging procedure which does not take this correlation into account will underestimate the event-mean concentration. Level-weighting addresses this problem somewhat by giving more weight to the high flow/high concentration periods. However, most stage-discharge relationships are non-linear concave, so a small increase in water level can mean a large increase in discharge. This non-linearity is not addressed by the level-weighting. Level weighting composite samples is a preferred approach relative to time-weighting composite samples, as time-weighting does not give a higher weight to the high flow periods and results in an even larger underestimation bias than level-weighting.

Hamilton Harbour Remedial Action Plan (HH RAP) TP Load Estimation Methods

Annual average TP loads to the Harbour from the tributaries have been estimated by the HH RAP through use of the “Draper Method” (D. W. Draper & Associates Ltd., 1993), an averaging method that applies average “wet” and “dry” TP concentrations to daily average tributary discharge values binned into two or three flow strata (HH RAP, 2004; HH RAP, 2010; Vogt, 1998). The “Draper Method” is based on limited monitoring data collected in the late 1980s/early 1990s. Daily average flows from the Water Survey of Canada (WSC) Red Hill Creek (Station ID 02HA014) and Grindstone Creek (Station ID 02HB012) gauging stations are obtained for each year of interest, and annual loads estimated for the HH RAP “Loading Report” (HH RAP, 2010; HH RAP, 2004) are as follows:

Red Hill Creek:

$$Load_{y,d} = k \times Flow_{y,d} \times 190 \mu g / l \mid Flow_{y,d} < 4.0 m^3 / s$$

$$Load_{y,d} = k \times Flow_{y,d} \times 500 \mu g / l \mid Flow_{y,d} \geq 4.0 m^3 / s$$

$$Load_{y} = \sum_{d \in y} Load_{y,d}$$

Grindstone Creek:

$$Load_{y,d} = k \times Flow_{y,d} \times 100 \mu g / l \mid Flow_{y,d} < 2.1 m^3 / s$$

$$Load_{y,d} = k \times Flow_{y,d} \times 300 \mu g / l \mid Flow_{y,d} 2.1 \geq 6.1 m^3 / s$$

$$Load_{y,d} = k \times Flow_{y,d} \times 1190 \mu g / l \mid Flow_{y,d} > 6.1 m^3 / s$$

$$Load_{y} = \sum_{d \in y} Load_{y,d}$$

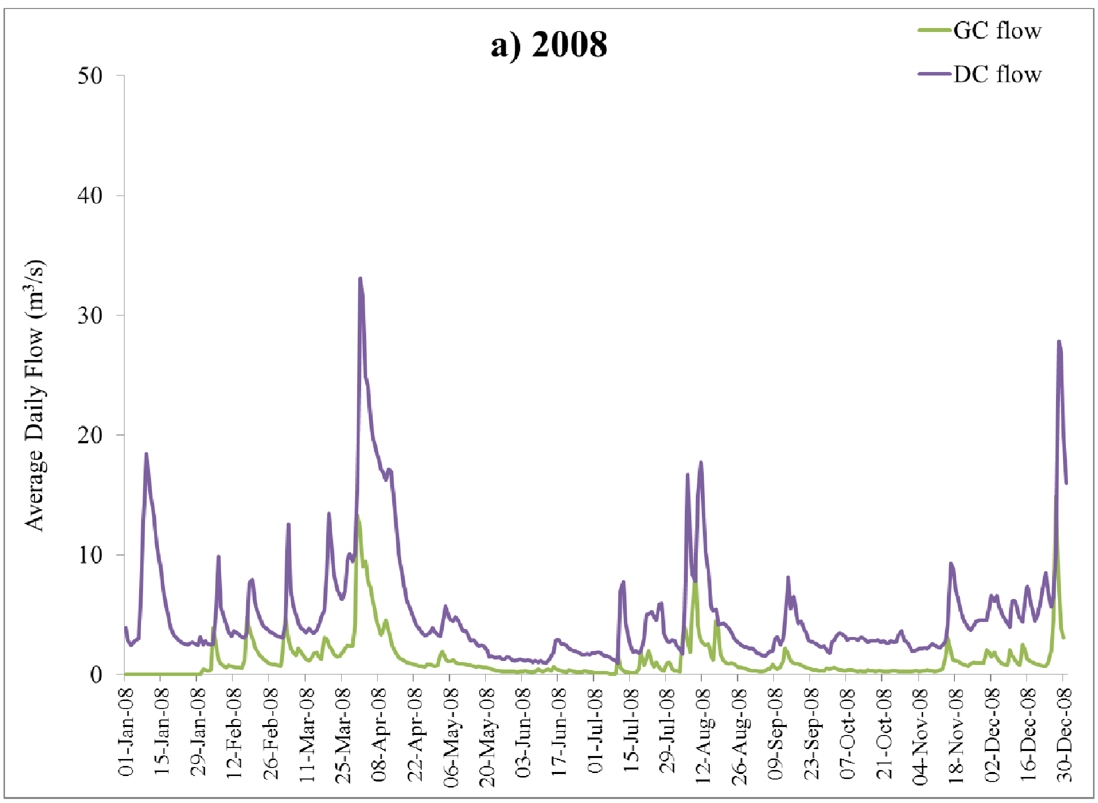
where y refers to a year, d to a day in a year, $Load_{d,y}$ to a daily load , $Load_y$ to an annual load, $Flow_{y,d}$ to the daily flow (m^3/s), and k is a unit conversion factor. No loadings were determined for Indian Creek as it is not included in the HH RAP Loadings Reports (HH RAP, 2010).

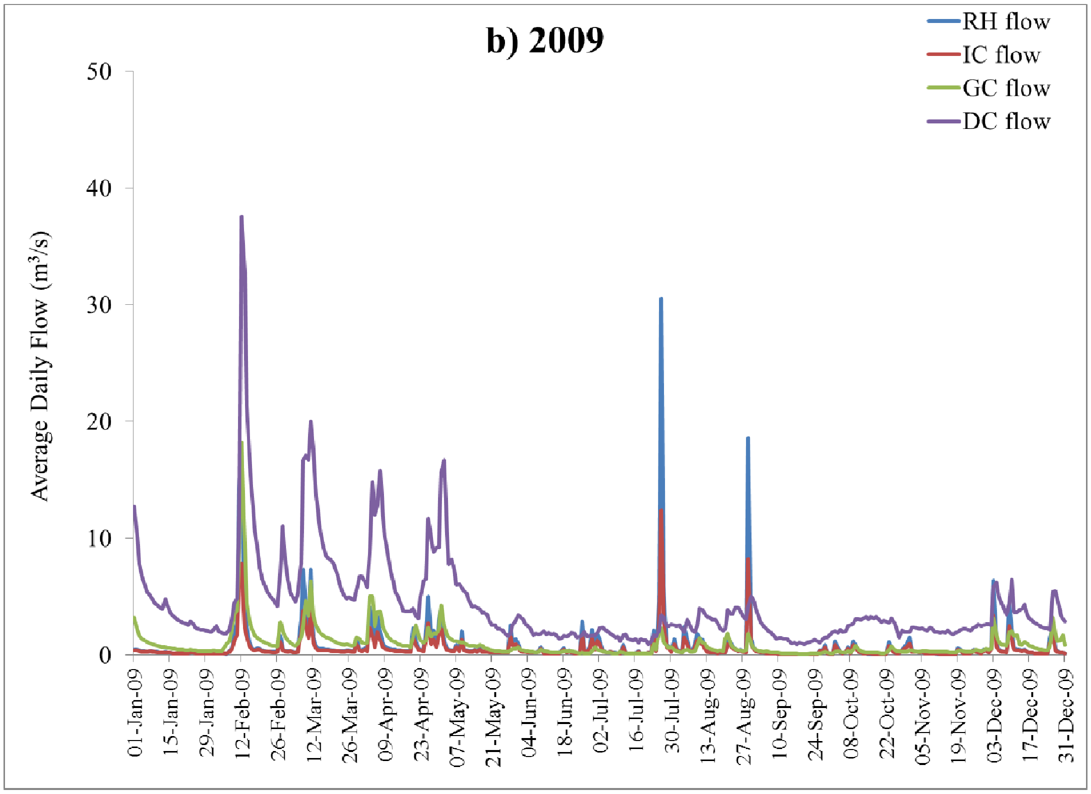
The HH RAP method for estimating the annual TP load from Cootes Paradise to the Harbour is based on a water balance whereby the total flow into Cootes Paradise is assumed to equal the flow out of Cootes Paradise to the Harbour via the Desjardins Canal (HH RAP, 2010). First, the average flow into Cootes Paradise was calculated as the sum of the average daily flow from the Dundas waste water treatment plant (WWTP) (data from City of Hamilton) and the average annual daily flow from the Spencer Creek WSC gauging station (Station ID 02HB007). The flow from Spencer Creek was multiplied by an area ratio of 1.44 to account for other smaller, ungauged creeks that also flow into Cootes Paradise (HH RAP, 2010). The annual average TP load to the Harbour was estimated as the average daily flow into Cootes Paradise multiplied by the annual average TP concentration of Royal Botanical Garden’s CP1 and CP2 Cootes Paradise monitoring stations. These samples were collected biweekly approximately 11 to 12 times a year between May and September.

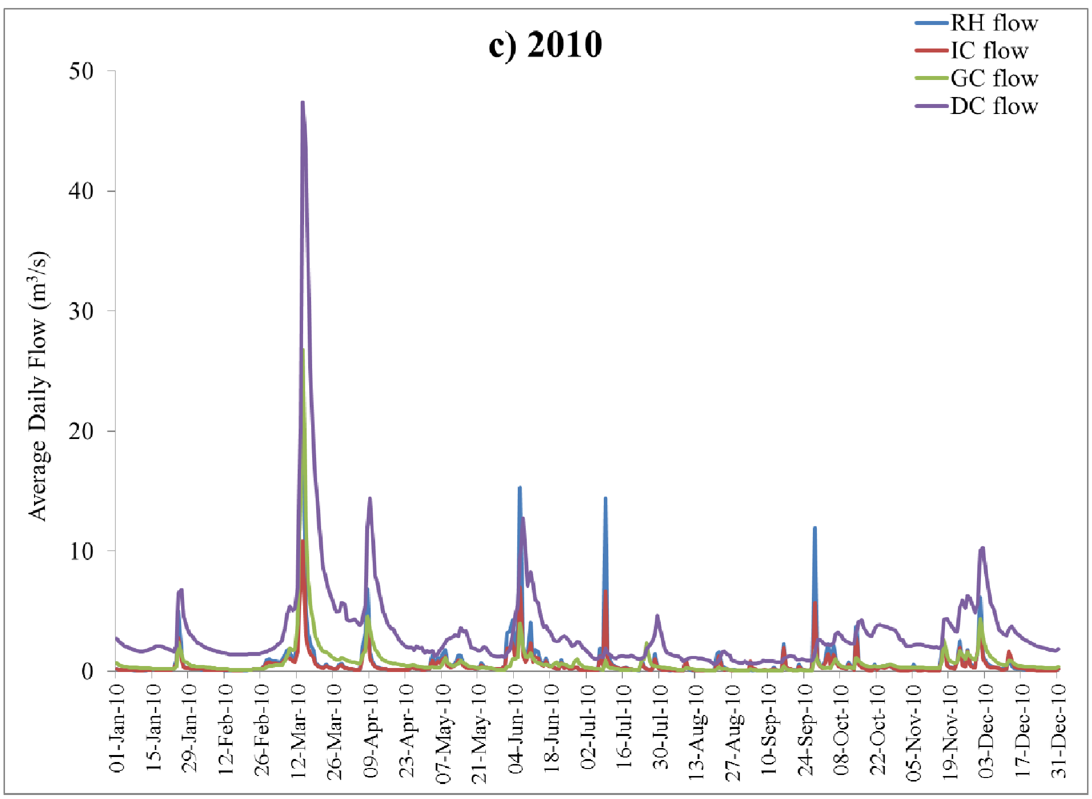
Hamilton Harbour Remedial Action Plan (HH RAP) Delisting Target for Creeks

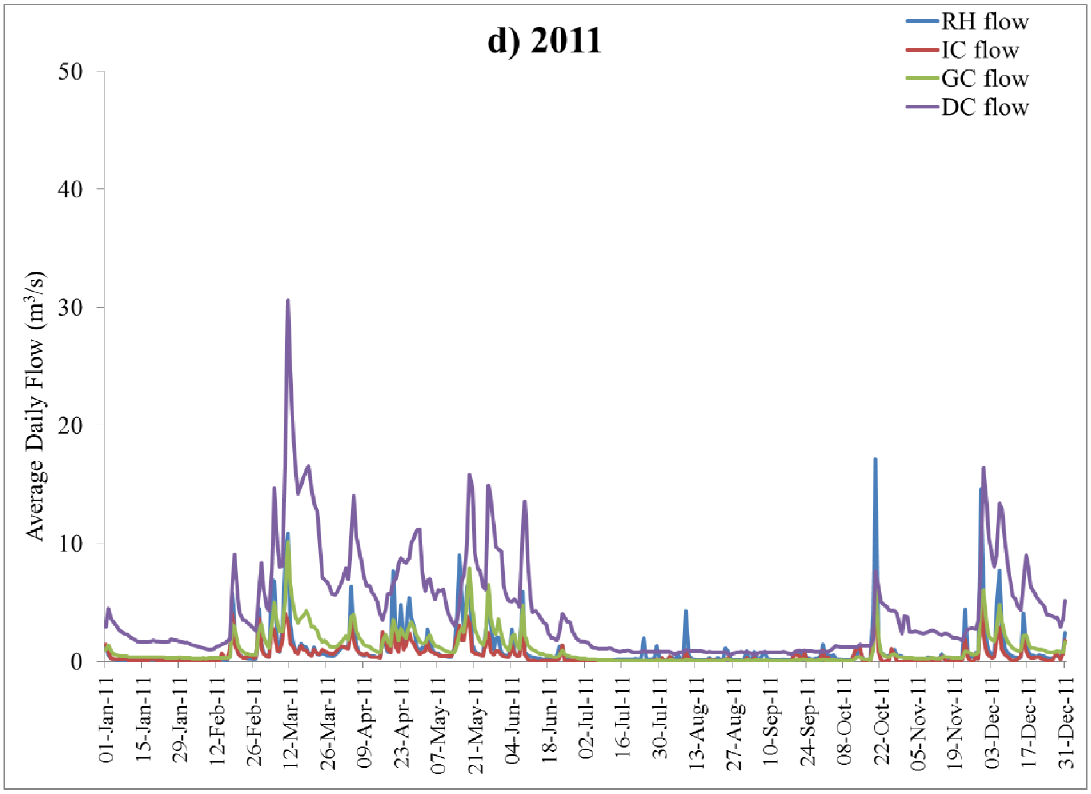
Although the assessment of TP loads from the creeks against the pertinent HH RAP target of 65 kg/d appears relatively straightforward, there are a few additional factors to consider which may impact whether the creeks can be considered in compliance with the HH RAP TP loading target. Cootes Paradise is not technically a stream, but does not have a separate Harbour TP loading target so loadings from this source are assumed to be included in the 65 kg/d TP loading target. This inclusion seems reasonable considering that Spencer Creek is included in the creeks target, and Spencer Creek is the largest hydraulic input to Cootes Paradise. In addition, the HH RAP has a delisting target for all Hamilton CSOs of 5 kg/d, and while some CSOs discharge directly to Hamilton Harbour, some discharge to Red Hill Creek and Cootes Paradise. As such, a fraction of the 5 kg/d target is applicable to the TP loads estimated from Red Hill Creek and Cootes Paradise, although the four creeks still would not have met a TP loading target of 70 kg/d in 2009 – 2011.

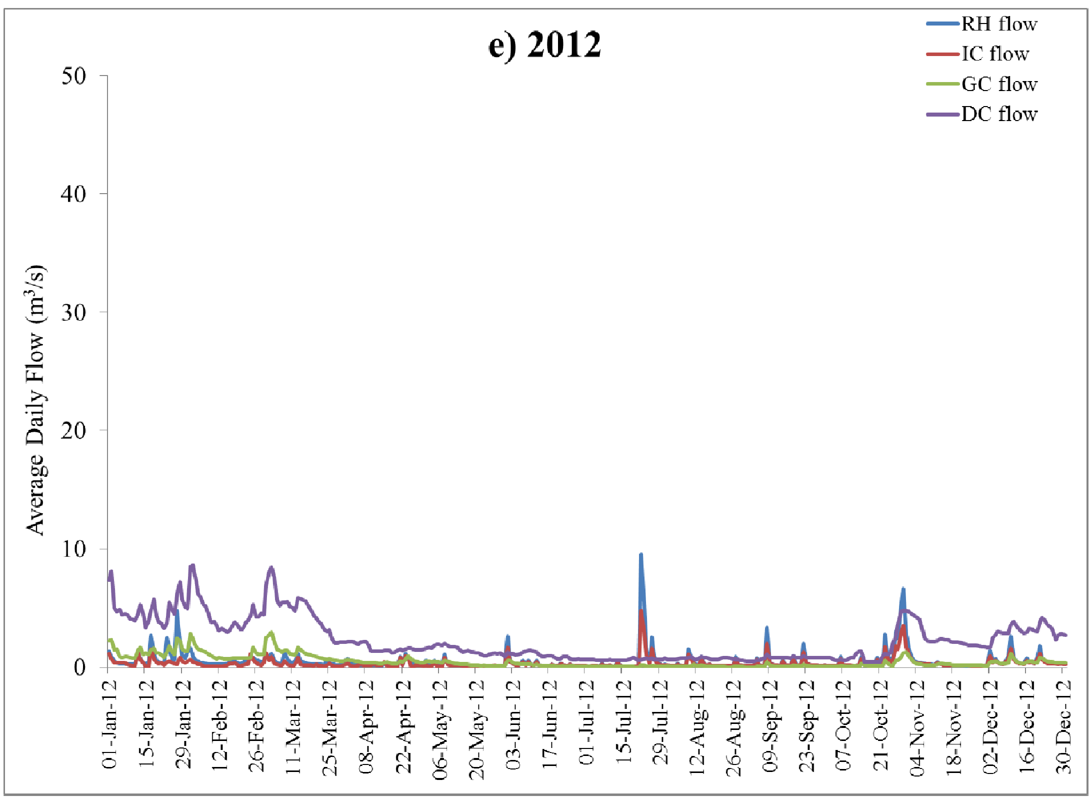
2008 – 2012 Average Daily Flows at the Four Monitoring Stations




b) 2009
RH flow
IC flow
GC flow
DC flow
Average Daily Flow (m3/s)
0
10
20
30
40
50
01-Jan-09
15-Jan-09
29-Jan-09
12-Feb-09
26-Feb-09
12-Mar-09
26-Mar-09
09-Apr-09
23-Apr-09
07-May-09
21-May-09
04-Jun-09
18-Jun-09
02-Jul-09
16-Jul-09
30-Jul-09
13-Aug-09
27-Aug-09
10-Sep-09
24-Sep-09
08-Oct-09
22-Oct-09
05-Nov-09
19-Nov-09
03-Dec-09
17-Dec-09
31-Dec-09


c) 2010
RH flow
IC flow
GC flow
DC flow
Average Daily Flow (m3/s)
0
10
20
30
40
50
01-Jan-10
15-Jan-10
29-Jan-10
12-Feb-10
26-Feb-10
12-Mar-10
26-Mar-10
09-Apr-10
23-Apr-10
07-May-10
21-May-10
04-Jun-10
18-Jun-10
02-Jul-10
16-Jul-10
30-Jul-10
13-Aug-10
27-Aug-10
10-Sep-10
24-Sep-10
08-Oct-10
22-Oct-10
05-Nov-10
19-Nov-10
03-Dec-10
17-Dec-10
31-Dec-10





ESM Figure S2. Average daily flows at Red Hill Creek (RH), Indian Creek (IC), Grindstone Creek (GC) and Desjardins Canal (DC) for a) 2008; b) 2009; c) 2010; d) 2011 and e) 2012.

ESM Table S1: Areal TP loads in Hamilton Harbour watersheds relative to recent values measured in watersheds of similar land use

Range (kg/ha/year)	Mean (kg/ha/year)	Reference	Notes
Urban land use			
0.36 (2012) – 1.4 (2009)	1.1	Red Hill Creek (event-based sampling, this study)	2009 to 2012
0.48 (2012) – 1.3 (2009)	1.0	Indian Creek (event-based sampling, this study)	2009 to 2012
0.145 – 0.837	0.41	Duan et al. (2012)	Urban & suburban land use, Chesapeake Bay area, USA
0.5 – 1.6	1.1	Boyd (1999)	6 Toronto tributaries, Ontario, Canada
	1.26	Arhonditsis et al. (*In press*)	Modelled urban watersheds in Bay of Quinte AOC, Ontario, Canada
Agricultural land use			
0.14 (2012) – 1.4 (2008)	0.80	Grindstone Creek (this study)	2008 to 2012
0.22 (2012) – 0.83 (2008)	0.56	Desjardins Canal (this study)[a]	2008 to 2012
	0.246	Duan et al. (2012)	Agricultural land use, Chesapeake Bay area, USA
	0.35	Macrae et al. (2007)	Cultivated land; Grand River basin, Ontario, Canada
0.06 – 0.45	0.16	Winter et al. (2007)	Forest/agricultural/urban; Lake Simcoe Basin, Ontario, Canada
0.22 – 0.57	0.33	AECOM (2009)	Forest/agricultural; St. Lawrence (Cornwall) AOC Watersheds, Ontario, Canada
0.01 – 0.13 (empirical) 0.35 – 1.2 (modelled)	0.073 0.69	Diamond (2011)	Welland River agricultural subwatersheds; Niagara AOC, Ontario, Canada
0.2 – 0.3		Chambers and Dale (1997)	Crops, pasture, best attainable; 198 North

			American watersheds
0.2 – 1.89	0.92	OMOE (2012)	15 agricultural watersheds in southwestern Ontario, Canada
0.3 – 1.27	0.70	Arhonditsis et al. (*In press*)	Modelled agricultural watersheds in Bay of Quinte AOC, Ontario, Canada

[a] note that the Desjardins Canal station is not directly comparable to other agricultural watersheds due to the influence of wetland processes, input of a WWTP, and CSOs to this station

References

AECOM, 2009. Evaluation of Remedial Action Plan Tributary Nutrient Delisting Criteria for the St. Lawrence River, Cornwall, Area of Concern. Prepared for the Ontario Ministry of the Environment.

Arhonditsis, G.B., Kim, D.K., Shimoda, Y., Zhang, W., Watson, S., Mugalingam, S., Dittrich, M., Geater, K., McClure, C., Keene, B., Morley, A., Richards, A., Long, T., Rao, Y.R., Kalinauskas, R., In press. Integration of best management practices in the Bay of Quinte watershed with the phosphorus dynamics in the receiving water body: What do the models predict? In Munawar, M. K., Minns, and M. Koops (Eds.): Ecology of the Bay of Quinte: Health, Management and Global implications.

Boyd, D., 1999. Assessment of Six Tributary Discharges to the Toronto Area Waterfront, Volume 1: Project Synopsis and Selected Results. Report prepared for The Toronto and Region Remedial Action Plan.

Chambers, P. A., Dale, A.R., 1997. Contribution of Industrial, Municipal, Agricultural and Groundwater Sources to Nutrient Export, Athabasca, Wapiti and Smoky Rivers, 1980 to 1993. Northern River Basins Study Report No. 110. Northern River Basins Study, Edmonton, Alberta, Canada.

Diamond, J., 2011. The Welland River Eutrophication Study in the Niagara River Area of Concern in Support of the Beneficial Use Impairment: Eutrophication and Undesirable Algae. Niagara River RAP, Welland River Eutrophication Study, Technical Working Group.

Duan, S., S. S. Kaushal, P. M. Groffman, L. E. Band, Belt, K.T., 2012. Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed, J. Geophys. Res., 117, G01025, doi:10.1029/2011JG001782.

D.W. Draper & Associates Ltd., 1993. Hamilton Harbour Tributaries Storm Event Monitoring Study. Final Report to: Hamilton Region Conservation Authority. Ancaster, Ontario. June 9, 1993.

Hamilton Harbour Remedial Action Plan (HH RAP), 2004. Contaminant Loadings and Concentrations to Hamilton Harbour: 1996 – 2002. ISBN 0 - 9733779 - 3 - 3. March 2004.

Hamilton Harbour Remedial Action Plan (HH RAP), 2010. Contaminant Loadings and Concentrations to Hamilton Harbour: 2003-2007 Update. ISBN: 978-0-9733779-8-9. January 2010.

Macrae, M.L., English, M.C., Schiff, S.L. Stone, M., 2007. Capturing temporal variability for estimates of annual hydrochemical export from a first-order agricultural catchment in southern Ontario, Canada, Hydrological Processes, 21, 1651-1663.

Ontario Ministry of the Environment (OMOE), 2012. Water quality of 15 streams in agricultural watersheds of Southwestern Ontario 2004-2009. Seasonal patterns, regional comparisons, and the influence of land use. Queen's Printer for Ontario. August 2012.

Vogt, J., 1998. Summary Report. 1990 – 1996 Contaminant Loadings and Concentrations to Hamilton harbour. Prepared for: Hamilton harbour Remedial Action Plan Office & Environment Canada. Prepared by: Water Technology International Corporation a division of Conor Pacific Environmental Technologies Inc., John Vogt & Hamilton Harbour RAP Office. June 1998.

Winter, J.G., Eimers, M.C., Dillon, P.J., Scott, L.D., Scheider, W.A., Willox, C.D., 2007. Phosphorus inputs to Lake Simcoe from 1990 to 2003: Declines in Tributary Loads and Observation on Lake Water Quality, J. Great Lakes Research, 33, 381-396.

Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007a. Hydrological modelling of the Chaohe basin in China: statistical model formulation and Bayesian inference. J. Hydrol. 340, 167–182, http://dx.doi.org/10.1016/j.jhydrol.2007.03.006.

Yang, J., Reichert, P., Abbaspour, K.C., 2007b. Bayesian uncertainty analysis in distributed hydrologic modeling: a case study in the Thur River basin (Switzerland). Water Resour. Res. 43, W10401, http://dx.doi.org/10.1029/2006WR005497