Inventorying Urban Phosphorus Loads

A report to the Bi-National Annex 4 Municipal and Rural Non-Farm Task Group

Shawn P. McElmurry, PhD, PE

August 1, 2014

Executive Summary

Based on increasing concern over eutrophication of surface waters within the Great Lakes basin, better estimates of phosphorous (P) loads from non-point sources are needed. Accordingly, through contract from Environment Canada (#KW405-13-1446 0) a series of workshops were conducted to develop an approach for inventorying urban and non-farm rural non-point source P loads to Lake Erie. This report describes major findings from these workshops and makes recommendations on how to develop reliable P estimates.

Critical knowledge gaps and problems associated with estimating P loads from non-point sources include:

1. The use of disparate land cover/use classifications schemes that inhibit model comparisons
2. A lack of spatial and temporal estimates of atmospheric P
3. Large errors in estimating sediment transport
4. Difficulty in accurately modelling urban hydrology
5. The inability to integrate datasets
6. The uncertain impact of structural beneficial/best management practices (BMPs)
7. Limited understanding of P biogeochemistry in urban environments

To develop robust estimates of P loads from urban and non-farm rural non-point sources it is recommended that the Ministry of Environment coordinate the long-term exchange of nutrient data and conduct an immediate 3-5 year study to develop accurate P load estimates. The first phase of this effort should provide funding for 3-4 projects to refine scientific understanding of fundamental processes that are necessary to accurately estimate P loads from non-point sources. Phase two of this study should focus on providing robust estimates (via modelling) of urban and non-farm rural non-point source loads of P to Lake Erie and connecting waters.

Developing a complete picture of P loads produced from urban and non-farm rural non-point sources requires a combination of research, modelling and monitoring that will take time. With leadership from the Ministry of Environment, many of the existing knowledge gapes can be filled and robust estimates of P loads can be obtained.

Table of Contents

1. Introduction

Elevated levels of phosphorus (P) entering the Great Lakes is recognized as a critical threat to the health of this unique freshwater ecosystem. Much research and monitoring has focused on the amount of P entering Lake Erie, one of the Great Lakes' most productive fisheries, as it is a critical limiting nutrient and excessive loads to the lake have been linked to anoxia and toxic algae blooms [1, 2]. While agricultural practices have long been identified as primary source of P, urban areas are recognized to be dominant sources in some watersheds [3, 4]. For example, the total P load entering the western end of Lake Ontario from the mixed agricultural and urban watershed feeding Hamilton Harbor was estimated to be 346±45 kg d^{-1}, approximately 126 metric tons per year, from 1996 to 2007 [5, 6]. Another study in Madison, Wisconsin USA has estimated urban land cover alone produces 0.64 kg P ha^{-1} yr^{-1} of dissolved reactive P and 1.1 kg P ha^{-1} yr^{-1} of total P [7].

To gain a better understanding of urban P loads and develop an approach for inventorying urban and non-farm rural non-point source P loads to Lake Erie three workshops were conducted (Table 1). During these workshops many competing interests/objectives and problems related to estimating P loads were discussed. Invited guests represented an array of disciplines including academics, public utility directors and operators, conservation officers, civil and environmental engineers, environmental scientists, economists and others. A complete list of attendees is provided in the appendix for reference.

Table1. Technical meetings focused on inventorying urban phosphorus loads

February 5, 2014	University of Windsor	Windsor, ON
March 13, 2014	Ministry of the Environment	Toronto, ON
March 14, 2014	University of Waterloo	Waterloo, ON

This report attempts to document the lessons learned from these workshops and facilitate the development of monitoring and modelling strategies to derive accurate estimates of P loads from urban non-point sources.

2. Sources of phosphorus in urban areas

Primary sources of urban P include discharges from municipal sewage treatment plants, failing septic systems, industrial discharges and the application of fertilizer [8]. Treated sewage effluent and industrial sources tend to provide more static loads and are considered point sources. Stricter effluent standards for sewage treatment plants and bans on phosphate detergents have resulted in significant reductions in P loads from point sources [9].

Major variable sources of P loads in non-agricultural environments identified by workshop participants included the application of lawn fertilizers, atmospheric deposition, enhanced release of soil and sediment, and runoff of animal/pet waste. Similar to agricultural land, the use of fertilizers in urban constitutes a significant source of P. The impact of P fertilizer can be gauged by reductions observed in areas that have eliminated P in their fertilizer. For example, the implementation of a municipal ordinance limiting the application of lawn fertilizers containing P in Ann Arbor, Michigan USA has resulted in immediate and long-term (3 years) reductions in soluble reactive P (24-5%) and total P (11-23%) in urban surface waters [10, 11]. Atmospheric loadings of P may also be the source of important and underappreciated loads. Estimates of atmospheric P loadings to Lake Erie are generally in the range of 10-18% of the total P load and less than 15% of the total load for most of the Great Lakes [12]. Importantly, atmospheric P loads are found to be more variable than point sources but less variable than loadings from surface water tributaries. Estimates of atmospheric P loads are limited by the lack of monitoring. The amount of atmospheric dry deposition is a major focus of current research.

Enhanced transport of soil or sediment bound P is likely to be one of the greatest episodic sources of P in urban settings. Construction activities are found to temporarily generate larger P loads (per area) than agricultural row crops [13]. During construction the amount of soil bound P entering surface waters is on the order of 50,000 metric tons of P per square kilometer per year. This loading is approximately 10-50 times more than the annual yields observed in agricultural areas and more than 500 times greater than from areas with undisturbed vegetation [14]. Discharges from large reservoirs intended to reduce sediment accumulation are another potential variable source of P loads to rivers. These seasonal events may serve as large pulse loadings to rivers and streams. It is important to point out that it remains unclear how internal loading (i.e. P re-suspension) within waterways and surface waters may impact water quality farther downstream and in receiving waters. Sediment dynamics are a critical source of uncertainty in trying to estimate the P load discharging tributaries [15].

Other significant sources of P in temperate regions include seasonal inputs of detritus and regular inputs of animal/pet waste. Leaves from deciduous trees (e.g. oak, poplar) are reported to leach 54-230 µg P g^{-1}, approximately 85% of which is reactive P [16]. Additionally, nearly 3 times as much P is released when leaves are cut, such as during mulching. Pet waste on the other hand has been found to responsible for 84% of P inputs to surface waters in the Minneapolis-Saint Paul, Minnesota metropolitan area [17].

Many workshop participants questioned if discharges of P from groundwater sources may be also be significant; they certainly are not included in most accounting. This is likely because the amount of P that is able to percolate through soils is generally low due to sorption. Exceptions may occur in organic or peat soils, where complexation and cation exchange from organic acids, iron (Fe) and aluminum may facilitate enhanced transport [18-20]. Similarly, P is more susceptible to movement through sandy soils which have natural low binding capacities and in soils which have become waterlogged, where conversion of Fe (III) to Fe (II) and organic P mineralization occurs [21, 22]. However, excessive discharge of nutrients and organic matter may essentially overwhelm the binding capacity of soils (e.g. septic fields) and result large fluxes of dissolved P. Therefore, it is plausible that leakage from old sewer lines (100+ years in for some cities) in densely populated areas may provide a scenario where soils no longer have the capacity to bind P. Along these lines, workshop participants repeatedly lamented that very little is known regarding urban groundwater systems. One of the few studies examining urban groundwater as a potential source of P was conducted by a U. Waterloo graduate student in the late 1990s under the supervision of Prof. Mike Stone. This study provided at least circumstantial support for this hypothesis when groundwater (1.5-3m depth) discharges to streams near residential area were found to contain 150-375 μgP L^{-1}. These surprising measurements were repeatedly observed and were thought to be due to leakage from septic tanks or sewer lines from nearby residential homes. Overall, a better understanding of how non-point sources impact P dynamics in surface waters, particularly during base flow conditions [23], appears to be critically important.

Urban drain tiles may represent an underappreciated pathway for enhanced P transport. Many urban areas have extensive tile drainage structures and, as indicated by a participant at the University of Waterloo meeting, “almost every pond has a direct tile drain from residential lots...they aren’t supposed to be there”. Studies in agricultural landscapes confirm these are conduits of P transport [24, 25], but the U. Waterloo group felt that loads from these sources would be minor compared to surface runoff. This assumption may not be valid considering significant differences between agricultural land and urban land. In agricultural landscapes, the amount of soil exported, even under no-till and other conservation practices, is considerably more than most urban settings – with the exception of construction sites [13]. Urban landscapes typically have extensive ground cover (e.g., grass) and impervious surfaces that are less likely to generate soil particulate matter. Under these conditions, shallow drain tiles intended to keep grassy surfaces dry and useable may have a greater impact on P transport than in agricultural areas where soil erosion is more significant. Regardless, deposition of P fertilizers onto urban surfaces, whether impervious (e.g. sidewalks, roadways) or pervious, is likely to result in P loads during stormwater runoff events.

3. Existing data characterizing urban phosphorus loads

When considering existing data sources, the first entity typically identified by workshop participants was Conservation Ontario. Conservation authorities have extensive knowledge regarding water quality. Other valuable resources identified by participants were rain and stream gauge networks present throughout Canada. These systems offer a valuable data for evaluating P dynamics since P loads are precipitation event driven. Additionally, workshop participants indicated that these systems have proven useful for detecting changes in hydrologic conditions (i.e. changes in stream flow) during the late 1980's that are believed to result from the implementation of beneficial/best management practices (BMPs), such as stormwater detention basins.

In the US, a similar set of data is available through The USGS National Streamflow Information Program (NSIP) (http://water.usgs.gov/nsip/). One of the more comprehensive sources of P data for US surface waters is the National Water Quality Assessment (NAWQA) Program Data Warehouse [26]. However, searching for all P records for three major Great Lake states – Michigan, New York, and Ohio – produced only 130 records, and this includes tributaries not connected to lakes that border Canada. Obviously, this simple exercise highlights the lack of coordinated data collection and historical records that can be used for evaluating long-term trends in P dynamics on the American side of the lakes.

One of the primary sources of P data (usually total P) is provided by wastewater treatment plants (WWTPs). This is true on both sides of the border. However, this data describing receiving water quality is typically obtained by WWTP operator that collect a small number (one to two dozen per year) of grab samples above and below wastewater treatment plant discharges. In contrast, WWTP operators typically monitor total P in the wastewater influent and effluent very closely, providing accurate loadings from these point sources. Point sources of P has historically been considered to be representative of the urban load to Lake Erie (e.g. Dolan and McGunagle [27]). Data from other monitoring studies do exist but they are usually for brief periods of time (1 to 2 years), predominately conducted during the "sampling season" (i.e. during warm weather) under dry conditions (e.g. Ontario Ministry of the Environment's Provincial Water Quality Monitoring Network or PWQMN), and focus on specific locations and objectives that may not be useful for other studies. Most important, many of these monitoring studies utilize grab sampling rather than flow- or time-weighted sampling schemes. These biased sampling schemes often miss episodic P loads. As a result, significant differences in total P load estimates are observed when measures are obtained via grab sampling versus event-based sampling [28].

Other potential sources of data describing P loads in Canada include the Ministry of Natural Resources (MNR), the Department of Fisheries and Oceans (DFO), Ministry of the Environment (MOE), Essex Region Conservation Authority (ERCA), Lake Simcoe Region Conservation Authority (LSRCA) and other conservation authorities. In the US, the National Oceanic and Atmospheric Administration (NOAA), the US Environmental Protection Agency (EPA) and other state environmental agencies, such as the Ohio EPA, the Michigan Department of Environmental Quality (MDEQ), and the New York State Department of Environmental Conservation (NYSDEC). Extensive monitoring operations have historically been conducted in the Lake Simcoe tributaries, the London River, and the north and south branches of the Thames River and these areas therefore sources of extensive data describing P dynamics. For example, continuous flow measurements are taken by Environment Canada for the HYDAT network or the Lake Simcoe Region Conservation Authority's flow monitoring network. The monitoring and modelling activity in the Lake Simcoe region are discussed in detail later (Section 6.2).

Finally, data describing urban BMP performance is likely to be useful for developing a better understanding of urban P loads. The instillations of BMPs are designed to alter the hydrologic behavior of tributaries and they are increasingly being implemented to specifically modify stormwater chemistry. Not all BMPs are structural. For example, street sweeping is a common management practice that if done routinely in urban areas and is reported to reduce total P loads 40-70% [29]. Understanding how BMPs modify typical loads is critical to developing a robust P load estimate. Two sources of urban BMP performance include the *National Pollutant Removal Performance (NPRP) Database* [30] and the *International Stormwater BMP Database* (www.bmpdatabase.org). Winer [30] has summarized the results of more than 150 pollutant removal studies using data from the NPRP. The *International Stormwater BMP Database* is an active database that currently contains data on more than 30 types of BMPs from 530 sites for more than 8,500 events. Unfortunately, less than 5% of these sites are within states or providences (Michigan, Ohio, Pennsylvania, New York, Ontario) adjacent to Lake Erie and most of these sites lie outside the Lake Erie watershed.

Overall, a general consensus emerged across all meetings that there are not a lot of data on the concentration of P in raw stormwater. Based on previous experiences with data gather and sharing efforts, two critical points were made: 1) it is very unlikely an entity conducting monitoring will alter their existing data collection in order to accommodate integration with outside data sources without significant incentives or regulatory requirements, and 2) a framework is required to translate data currently being collected to a common format so that long-term evaluation and intra-study comparisons can be performed. Participants at the U. Windsor workshop, in particular, felt that increased monitoring of raw stormwater from different landscapes would be beneficial to developing a better understanding of P dynamics and allow for more sound monitoring schemes.

4. Monitoring

When considering methods for monitoring P in urban surface waters and from urban non-point sources, three key factors must be considered: (1) P biogeochemistry, (2) the spatial and temporal resolution required, and (3) logistic, financial and other constraints. It is important to keep in mind that the information obtained through monitoring programs (or modelling) are limited and are likely not useful for answering questions outside the original project objectives. For example, a monitoring program designed to verify P loads from multiple types of watersheds would look very different than a monitoring program designed to verify P load reductions due to BMPs. Regardless of the sampling scheme, proper quality assurance/quality control (QA/QC) measures must be implemented.

4.1. Phosphorus biogeochemistry

Traditionally, total P measurements are the predominant measure of P in Lake Erie and its tributaries. Generally, a large fraction of total P is composed of particulate bound P [24]. Based on this phenomena sedimentation basins are commonly implemented to reduce P loads. While this is a reasonable approach, removal of P via sedimentation in structural stormwater BMPs is limited. Due to the larger surface area to mass ratios for small particles, smaller fractions of sediments preferentially adsorb P [31]. Based on stokes law, the length of time required to achieve settling is inversely proportional to the particle size (i.e. as the particle size decreases, more time is required to achieve removal via settling). The vast majority (65-85%) of total suspended solids (TSS) discharging from stormwater ponds and wetlands are composed of particles less than 4 μm in diameter [32]. The preferential sorption of smaller particles that are resistant to settling decreases the removal efficiency of P via sedimentation basins. Therefore, the removal of P via structural BMPs requires extended detention times and low flow velocities, something not always found in stormwater ponds during runoff events. As a result, detention basins are commonly found to exhibit a lower removal efficiency for P than for total suspended solids [32, 33].

In most urban systems, the number of stormwater detention basins and other forms of structural BMPs currently installed (i.e. the extent of BMP saturation)– estimated to be 5-10% of the total load [34] – is not likely to have a significant impact on P loads at the watershed scale. While stormwater detention basins and other structural BMPs may not be necessary to estimating P loads at the watershed level, how they function is critically important to understand P transport, particularly during episodic events. For example, approximately 80% of the sediment load entering Lake Erie is less than 63 µm in diameter [31]. If P is available from groundwater or tile drains, it is likely to be bound to suspended particulate matter that will not settle out and will easily migrate downstream to lakes and reservoirs. Due in part to the desire to better understand P dynamics within these complex systems, the number of studies that measure dissolved reactive phosphorus (DRP) have increased. Studies evaluating DRP, also called soluble reactive phosphorus (SRP), are responsible for identifying an apparent increases in the amount of DRP feeding algal blooms in western Lake Erie [35].

Based on current studies evaluating urban P dynamics and impacts to Lake Erie, fundamental questions concerning the distribution of dissolved and particulate bound P include:

1. do these fractions vary seasonally?
2. are they source dependent (i.e. do they vary based on land cover/use)?
3. do these fractions change within structural BMPs?

If the focus of monitoring or modelling was to evaluate how the implementation of BMPs or how climate change may impact P loads in the future, then a more mechanistic approach must be considered and understanding speciation becomes more critical.

4.2. Spatial and temporal resolution

The most common assessments of P loads to Lake Erie provide annual estimates with varying levels of spatial resolution. These estimates are not ideal. Assessments of P loads need to be conducted at times scales that are more in line with ecosystem dynamics. Stamm, et al. [23] highlighted how important it is to consider the receiving water when determining the impact of P loads. For instance, large stagnant water bodies, such as lakes or bays, are likely to be susceptible to algal growth resulting from large episodic P inputs as well as long-term, low inputs of P. Whereas elevated concentrations of P under baseflow conditions during the growing season may be more critical for streams that may continuously receive pulse P loads. Therefore, mitigating eutrophication of in each of these surface water systems may require different approaches.

The challenges of monitoring water quality and P in urban systems cannot be ignored. In urban areas, dynamic stormwater events require rapid sampling, commonly performed using auto-samplers. Participants at the U. Waterloo workshop felt that even if auto-samplers were collecting samples throughout a rain event, this type of sampling requires a change in flow or another trigger to begin sample collection which can be problematic. One common approach to properly setup sample collection under these conditions is to monitor the system hydrology *prior to* sampling and set the automated sampler to collect samples when changes in flow exceed typical fluctuations in baseflow.

Even in a well characterized system, this approach requires multiple measures of flow to be obtained prior to P sampling. This delay in triggering sample collection inevitably results in a lag that can lead to insufficient characterization of stormwater early in the hydrograph (i.e., missed sampling early in the rising limb). Not accounting for the rising limb of the hydrograph can significantly impact P load estimates as this is the portion of the event which is likely to contain the “first-flush” of sediments and nutrients. While likely not critical to assessing P loads, defining the end or tail of a runoff event is also challenging due to logistic constraints of collecting discrete samples. As a result, participants at the U. Waterloo workshop felt estimates of urban P loads are often underestimated. More advanced methods, such as triggering sample collection based on rainfall, can be employed to overcome this limitation but these approaches require additional instrumentation, modelling and effort.

One area where nearly all workshop participants agreed was on the need to understand seasonal effects on P loads. Seasonally averaged total P concentrations are generally not useful and certainly do not capture important temporal variability [15]. Participants of the meeting at the U. Waterloo believed a significant amount of non-point source P loading occurs during the winter but goes unmeasured. This is supported by the Stormwater Assessment Monitoring and Performance (SWAMP) program [32], which found structural BMP performance to be lower during the cold season relative to the warm season. This was through to be due to changes in hydrology. In early spring, when the ground and detention basins are still frozen but surface temperatures are warm enough to produce surface runoff, it is likely that P exports are higher than would typically be observed. In addition to hydrologic changes at these sites, P uptake by plants does not occur during the cold season when plant growth is negligible and may actually be negative due to decay. Clearly understanding seasonal effects on different land covers/uses and structural BMPs is critical to developing robust estimates of P loads from urban areas.

4.3. Logistic and other constraints

It is important to have a realistic understanding of the costs associated with the level of sampling that will be required to fully achieve monitoring objectives. The SWAMP project [32] provides a useful description of costs that can be expected for monitoring a single BMP performance: “The estimated cost for monitoring and data analysis at a site with 1 rain gauge, 1 inlet and 1 outlet, with continuous flow monitoring and automated sampling, allowing for 30-40 flow proportioned composite samples is approximately $100,000 per year.” In addition to the simple costs of monitoring, it is important to remember that monitoring in an urban environment is extremely challenging due to vandalism. Often auto-samplers are damaged, disconnected, or stolen, creating a challenge for any monitoring effort.

Assessing P loads in urban systems might benefit from the model utilized to set total maximum daily loads (TMDLs) where targeted, intensive studies are conducted in order to fill critical data gaps and this information is assumed to be constant across similar (unmonitored) sites. However, uncertainty concerning how well land cover/use classifications are able to characterize P loads remains. Therefore, this approach needs to be verified for non-point source pollutants from urban areas with different geology and consistent, or at lease comparable, land cover/use classifications. Ultimately, collecting detailed data in a few select systems, performing a sensitivity analysis on model parameters, and verifying the validity of critical assumptions is required to accurately characterize P load estimates and more efficiently collect data.

All data collected should undergo strict quality assurance/quality control (QA/QC) to ensure confidence in the results. The QA/QC methods must be defined in the context of project requirements and objectives. Appropriate QA/QC methods should describe:

1. Precision (a measure of random error)
2. Bias (a measure of systematic error)
3. Accuracy
4. Representativeness
5. Completeness,
6. Comparability, and
7. Sensitivity

Based on standard practice [36], QA/QC procedures should include the regular analysis of blanks (method and field), duplicates (replicated laboratory analysis of split samples and replicated field samples), spiked samples (matrix matched and standard addition, if possible) and standard reference materials. Field instrumentation, such as flow meters and real-time water quality monitoring devices require regular maintenance and calibration that is often overlooked. While a real-time P meter is not currently available, other sensors, such dissolved oxygen probes and particle counters, can provide valuable information in assessing the mobility of P. A minimum of bi-weekly maintenance and cleaning of equipment deployed in the field is recommended [32, 36, 37].

Regarding the representativeness and bias of sampling, the inconvenience and problems associated with collecting samples during cold weather is a significant problem that should be addressed. Because tubing on sampling devices can freeze and hands get cold, many entities view the warm season as the “sampling” season. As discussed previously, there is preliminary evidence to suggest that P loads and BMP performance may vary seasonally. For this reason, historical data providing annual and seasonal estimates based only on sampling during warm periods may be invalid [28]. This point was raised repeatedly by workshop participants.

5. Modeling

Thus far, no one has successfully developed a full closed-loop mass balance model of urban P cycling that can characterize temporal variability or provide reliable predictions (e.g. Zhang, et al. [15]), highlighting the gaps in our understanding of urban P behavior. Despite this shortcoming, there are many models that have been applied to quantify P loadings from non-agricultural sources. Specific models are not the focus of this report and therefore a general discussion of modelling approaches is presented. At the workshops, there was a tendency to classify models as one of three broad types: (1) simple/generalized, (2) moderately complex, or (3) complex.

5.1. Model Complexity

Much the workshop discussions were devoted to gauging participants' opinions on what level of model complexity was required to accurately estimate P loads from urban systems. The tradeoff between simple models that are generally easier to understand but provide limited information versus more comprehensive models that are more labor and data intensive but offer more extensive analysis is common [38]. One comment that was made by participants at the U. Windsor meeting in support of utilizing a simple modelling approach worth noting was that it promotes intra-agency collaboration. Participants assumed that simple models would utilize data that are easier to obtain. As data became more difficult to obtain, participants believed agencies would be reluctant to share this more "valuable" information. This benefit of simplistic modelling should not be underestimated as collaboration is likely to ensure the acceptance and long-term sustainability of developing robust P load estimates.

Generally, participants seemed comfortable with a relatively simple modelling approach that utilizes a modest amount of field observations to empirically estimate unknown values. One example of this approach is the use of export coefficients. Winter and Duthie [3] provide an example of an export coefficient model developed for an urban watershed in Southern Ontario. This modelling approach was used to forecast the impact of P loads based on urban development. The model was developed based on pre-1978 data when urban and agriculture (combination of crops and pasture) land cover/use constituted approximately 30% and 51% of all land cover/use, respectively. By 1995-1996, the model prediction year investigated, the composition of land cover/use had shifted to 41.3% urban and 45.5% agriculture. Under this scenario, runoff from the urban area was found to contribute most of P load (1,537 kg yr^{-1}), followed by commercial crop production (816 kg yr^{-1}). A potential limitation of this approach is that P loading from urban areas originates from a variety of sources, including runoff from roads, parking lots, roofs, lawns, leaf fall, and driveways. Therefore, if a higher degree of spatial and temporal resolution is required, a more refined land cover/use classifications that accurately capture these variations is required.

It should be noted that land cover/use classifications may also not capture major differences in infrastructure. For instance, old urban areas are commonly served by combined sewers, whereas newer urban areas are served by separate storm and sanitary sewers. Some workshop participants felt that a greater refinement of land cover/use classifications (e.g., rather than just urban, classifying urban areas based on density and type of development) is useful for characterizing P loads. This call for more refined land use classifications is something that not all participants believed was necessary. Currently, a variety of different land cover/use classification schemes are used, most without extensive validation or intra-comparison. Ultimately, the level of refinement on land cover/use classification and spatial resolution are a function of the system and questions being evaluated. Because there are a variety classification schemes used, comparing export coefficients between studies and sites is difficult, if not impossible, in most cases. Developing a standard land cover/use classification scheme would be helpful to evaluating modelling approaches and P load estimates.

An approach to developing an appropriate model that was offered by participants at the U. Windsor was to begin characterizing urban loads to Lake Erie by extensively reviewing the literature to identify P sources and the relative magnitude of loads. Even rough estimates would be initially useful. Existing data could then be used to develop a simple model that describes observations. While participants may have not be aware of it, extensive work has already been performed within the Lake Erie, Lake Simcoe and Lake Ontario basins to characterize P loads, making this initial step unnecessary (e.g. Dolan and Chapra [12], Zhang, et al. [15], SWAMP [32], Eimers [39], Han, et al. [40], Ramin, et al. [41], Wellen, et al. [42], Wellen, et al. [43], Whitehead, et al. [44], Winter [45]). Regardless, the next step in analysis would be to evaluate different cities or municipalities so that a diverse array of non-point sources is represented. Multiple linear regression analysis could then be used to identify primary variables responsible for observed changes in P.

An example of this type of mid-level model is the SPAtially Referenced Regression On Watershed attributes (SPARROW) model developed by the U.S. Geological Survey [46, 47]. Regional SPARROW models incorporate geospatial data on geology, soils, land use, fertilizer, wastewater treatment facilities, temperature, precipitation and other watershed characteristics. While the SPARROW model provides an estimate of average annual P loads at landscape and regional scales, others have extended this analysis to account for variations in annual loads [42]. Results identify sources and sinks of nutrients within tributaries proximate to Hamilton Harbour (discussed later). For example, results suggest that wetlands likely serve to attenuate total P transport and P loads vary significantly depending on flow conditions [43]. High flow years tend to generate higher loads, and more of the load comes from further upstream which is believed to be due to the lower in-stream attenuation rates during high-flow events [43]. Recently, further model refinement has facilitated the estimation of daily P inputs that feed directly into a coupled eutrophication model [48].

If the model is required to be predictive for urban environments, it must have a fully coupled hydrologic and biogeochemical model (complex model). With urban systems being primarily driven by episodic events, most urban models have focused on hydraulics. For example, Ontario Conservation authorities already have existing hydrologic model for most watersheds. However, the US EPA's SWMM model is currently the only modelling environment that is capable of accurately describing combined sewer overflow events (CSOs) which will likely impact episodic P loads for many older urban areas.

Multiple models have been developed to accurately describe P loads at different spatial scales within the Great Lakes basin (e.g. SPARROW [48], SWAT [49, 50], INCA [44], HSPF [51]). Most of these models are deterministic. Deterministic models are most useful for comparing how, based on current understanding of P dynamics, changes in regulation or infrastructure would alter existing (baseline) P loads. However, if the purpose of the model is to evaluate compliance, then they will likely have to account for the variability of these complex systems and utilize a more probabilistic approach (i.e. stochastic models). Stochastic models may also be useful for forecasting the likelihood of future scenarios. For example, if it is determined there is a threshold P load that will cause unacceptable harm to receiving waters, a stochastic modelling approach is likely warranted. In this scenario, we may accept that there is a 95% probability that a given exceedance criterion will not be achieved. Therefore, one of the key items that must be determined before additional monitoring or modelling is conducted is to determine the desired outcome of the activity.

It is worth noting that participants in these workshops indicated a closed mass balance model for P cycling in urban systems has not been demonstrated. The inability to formulate completely closed mass balance for urban systems is not unique as "*no fully closed annual mass balance for the major biologically important elements has yet been achieved for a marine system of any size and complexity*" *[52]* either. Ideally, all models would be based on completely closed mass balances in which we have direct measures of inputs and outputs. However, this is unrealistic and unnecessary. For example, if the primary outcome of the model is to provide an annual P load to Lake Erie, then defining processes that account for less than 1% of the total load estimate is not warranted. In this case, a closed-loop mass balance provides no greater insight than a simple model which omits unimportant parameters. The objectives of the modelling effort need to clearly define what level of resolution and accuracy is needed prior to model development.

The idea of creating a large-scale complex models (e.g., completely coupled biogeochemical model) was not popular with most workshop participants. Workshop participants generally felt that as spatial size increases, you have to sacrifice some level of resolution. Very few water quality models are available to predict P concentrations with an adequate level of spatial and temporal resolution. The level of temporal resolution necessary to accurately define P loads is unclear.

Generally, larger spatial scales can be characterized with coarser time resolution without losing accuracy. However, as the spatial resolution increases, differences between types of land cover/use and hydrology become more important. In urban watersheds, loadings are likely to be more episodic and defining flow paths is more complex. Defining the appropriate sampling scales for smaller systems becomes increasingly dependent on the system (how does the stormwater infrastructure route flow, etc.). Accurately characterizing flow or sediment loads across different storm events is difficult, if not spurious, without some amount of initial monitoring. Further complicating monitoring efforts is that all P fractions (dissolved, particulate bound, etc.) are not transported equally and are variable based on hydrologic events [53]. As a result, determining the appropriate sampling interval *a priori* for urban watersheds based on land cover/use and general hydrology characteristics is likely to yield inaccurate estimates. Within this context, a comment that was provided at the U. Windsor meeting and echoed at the meeting in Toronto was that, when evaluating P loads, if the level of time resolution required is undefined from previous study, then the initial sampling should be conducted with the maximum resolution that is logistically feasible. After evaluating initial data, an optimal time resolution should be identified. This approach seems logical for developing a robust analytical strategy for monitoring but is needed prior to modelling.

It is worth explicitly pointing out that more complex models do not necessarily ensure more accurate results [54]. One problem with a complex, mechanistic model is that not all mechanistic processes are well quantified. For example, sediment re-suspension and transport are critical mechanisms of P transport. However, current models are only able to predict sediment loads within *1 or 2 orders of magnitude*. While suspended particulates are not sources of P, rather vectors which transport P, this error will dominate P load estimate error. Additionally, there is an implicit assumption in most models that when you remove one fraction of P, dissolved or particulate bound, that the remaining fraction is fixed. Like all chemical processes, P dynamics are based on equilibrium and kinetics. Therefore, it is unclear if this assumption is valid.

When considering the structural complexity of modelling, we ultimately reach a point when we must ask if it is necessary to enhance the mathematical description or obtain additional information. Until it is necessary, "the gradual incorporation of complexity, where possible and relevant, is the most prudent strategy" [48]. Constructing the least complex model possible that can be calibrated and verified by monitoring data was generally considered to be ideal. One of the comments regarding the modelling approach that came out of the U. Waterloo meeting was fairly consistent with the approach recommended at other workshops, to paraphrase based on notes:

> *Should we go from basic to intermediate?*
> *We need to identify the different sources in an urban setting.*
> *What are the sources of P?*
> *They are likely industrial, residential, and proposed developments.*
> *Then we need to go from there.*
> *Not on a daily basis, but maybe on a monthly scale.*

5.2. Data requirements

Land cover/use and geology are major factors impacting the amount and type of suspended sediments and associated P loads. For example, despite only constituting only 3% of the land cover within the Lake Champlain watershed (62 % forest and 28 % agricultural land), urban sources were estimated to contribute a disproportionate 18% of the P load [4]. Workshop participants believed the primary pathway responsible for generating urban P loads involved the transport of particulate bound P over impervious surfaces (parking lots, roadways, etc.) to receiving waters. Particulate bound P generated during construction activities or through the misapplication of fertilizers (e.g. applying fertilizer to impervious sidewalks, driveways or roadways) are likely two of the primary sources of P. Quantifying particle size distributions, or some fraction of suspended particles, may also be required to develop better estimates of P dynamics in urban systems. Estimating P loads from these sources may be possible using supplementary data sources. For instance, a linear regression analysis may be able to link building permits to P loads associated with construction activity. Fertilizer sales records are likely to be a good indicator of P loads from suburban lawns. As suggested in one of the workshops, it may even be possible to indirectly estimating the portion of the urban non-point P resulting from pet waste based on dog registrations/licenses. Developing an accurate prediction of P loads is most likely to be based on land cover/use and geology but will also include ancillary data characterizing other variables. Ultimately, it is critical that a reliable land cover/use classification scheme which is transferable across studies be available and utilized by researchers to effectively compare model performances.

To provide an illustrative example of challenges in utilizing multiple sources of data, Prof. Darko Joksimovic from Ryerson University presented research utilizing USEPA SWMM models to identify the optimal placement of lot-specific structural BMPs. The modelling utilized urban hydrologic response units based on multiple assumptions derived from values reported in the literature. This approach is one of the most rigorous efforts conducted to date, but highlighted data constraints. These constraints included: limited access to datasets, disparate sources of data, lack of coordinated data standards, incomplete spatial coverage, and a general lack of measurements for parameters required to accurately model urban systems. While Dr. Joksimovic's presentation highlighted challenges using existing data, it also sheds light on what is needed for future modelling efforts.

If additional data is to be collected there needs to be proper structure in place to ensure it long-term viability. Two general statements can be made regarding data: (1) municipalities are generally hesitant to commit to monitoring for financial as well as liability reasons, and (2) municipal drains, in general, are poorly monitored. Unless regulations require monitoring or a strong incentive exists for users to contribute data and there is sustainable long-term funding, history shows that parameters will not be measured in a manner that will enable sound, science based decisions.

6. Monitoring Programs

It is worth pointing out a few examples of communities that have implemented long-term P monitoring and modelling.

6.1. Hamilton Harbour

The Ontario MOE collected event-based water quality samples from the watersheds of Hamilton Harbour between July 2010 and May 2012. This project was undertaken to more accurately estimate non-point source total P loads to Hamilton Harbour, an identified need of the Hamilton Harbour Remedial Action Plan. Automated water samplers were installed at four stations located on downstream sections of Red Hill Creek (primarily urban), Indian Creek (primarily urban), Grindstone Creek (primarily agricultural, rural), and on the Desjardins Canal (primarily agricultural, rural), the hydraulic connection between the Cootes Paradise wetland and Hamilton Harbour. The automated samplers were set to trigger at the onset of a high flow event, although some baseflow sampling was also conducted, particularly during the data-poor winter period. For each station and sampling event, samples were collected once an hour for 24 hours, and then a level-weighted composite sample was prepared from the 24 grab samples, and subsequently submitted for analysis of nutrients (total P, orthosphosphate, total kjeldahl nitrogen, nitrate and nitrite, ammonia), TSS, metals and other general chemistry parameters. For select events, grab samples collected on the rising limb, at the peak, and during the falling limb of the hydrograph were also submitted for analysis in anticipation that observation of nutrient dynamics at a high temporal resolution will assist with identification of event loading processes.

In total, 87 events were sampled over the 22-month period, representing water quality from urban and agricultural land uses during a range of flow conditions (storm events, spring freshet, baseflow) and seasons. Preliminary total P loads for each of the creeks have been estimated through the development of a series of empirically-based regression equations. This work has identified many important aspects of total P dynamics in urban creeks, including the importance of year-round sampling. For example, total P concentrations during rain/melt events demonstrated no significant difference among seasons for the two urban creeks studied [55]. This is contrary to differences in hydrology, especially the amount of surface runoff generated, observed in watersheds with contrasting land cover/use [50]. Further, urban non-point source total P loads in winter can be as high as or even higher than summer loads (Long et al., in prep.). This work also identified the importance of accurately characterizing total P loads during high flow events, especially large storm events in urban creeks which may be a relatively greater contributor to total loads relative to spring runoff in more agricultural and rural areas. Although the sampling program was designed to estimate total P loads entering Hamilton Harbour, and was not directed towards the identification of specific urban total P sources, ongoing analysis of trends in this large dataset should assist with elucidation of major loading contributors, such as CSO events or direct runoff/overland flow. This dataset is available by contacting Tanya Long at the Ontario MOE.

6.2. Lake Simcoe

P loads, including loads from major tributaries, to Lake Simcoe are monitored through a partnership between the Lake Simcoe Region Conservation Authority (LSRCA) and the Ontario MOE. P loads to Lake Simcoe from major watershed tributaries continue to be evaluated by monitoring the flow and P concentrations, which are then used to calculate load [57]. For each river, the concentration is monitored as close to the mouth of the river as possible. Continuous flow measurements are taken by Environment Canada for the HYDAT network or the Lake Simcoe Region Conservation Authority's flow monitoring network. Nutrient sampling is bi-weekly during the ice-free season and every three weeks during the winter with additional event based sampling and supplemented with some continuous sampling at select sites. The flow and concentration measurements are used to determine the load from the tributary including any urban area captured upstream.

The LSRCA has also been conducting studies on stormwater treatment facilities. In a recent study of 135 stormwater detention ponds they noted that sedimentation of stormwater detention basins and ponds is occurring at a much faster rate than commonly assumed. Typically, these detention ponds were designed for 15 years of sediment accumulation. However, the median life span before maintenance was needed was found to be 9.5 years based on the loss of hydraulic retention due to sedimentation. Overall, less than 1% of the ponds studied were found to be performing as assumed (i.e., a Level 1 classification). As a result, dissolved oxygen levels in the ponds were often found to drop to below 2 mgO_2 L^{-1}, a level that induces P release from sediments. Ponds were also found to be stratified with a large fraction (2-44.6%) of loosely bound P. Overall, results of the Lake Simcoe study indicated BMP performance is less than currently assumed and suggests the need for treatment train approaches.

In another example, Baulch, et al. [56] describe results of intensive monitoring in the Beaver River, a subcatchment of Lake Simcoe. While multiple models were employed, depending on the season, landscape and climate, the Branched-INCA-P model was used most often and predicted the discharge of P on a daily time step. Maximum total P values were observed during the summer, with an average observed total P of 0.023 mg L^{-1}. Model simulations represented water quality at some sites well, although on the whole, agreement between model results and observed data was relatively weak. For this system, modelling suggested that high concentrations of P were not driven by hydrologic events and were likely due to internal processing (e.g. P release from sediments). Additionally, average P reductions were relatively small (3 to 12.7%) with relatively low concentrations (0.020 mg total P L^{-1} and 0.019 mg dissolved P L^{-1}) being reported. These small reductions and low concentrations make monitoring difficult but quantifying P reductions appears to be feasible. Finally, the study by Baulch, et al. [56] points out the need for additional research to better understand internal P dynamics and quantify the effectiveness of agricultural BMPs in reducing nutrient loads as they are large sources of uncertainty.

6.3. Milton

The city of Milton, Ontario has focused monitoring and planning on the Sixteen Mile Creek sub-watershed [58]. For this monitoring they are incorporating a "holistic EMP (Environmental Monitoring Program)" that was completed over a 10 year period to evaluate the impact of implementing the "Secondary Plan" has had on water quality and if they are achieving their expected outcomes. The city recently "*completed two secondary plans (Bristol Survey, which added 30,000 residents and Sherwood Survey, added 40,000 residents)*" and "*have just completed a third residential secondary plan (Boyne) and expect an increase in population of 50,000 residents.*" All of these monitoring programs are long-term, requiring 10 years of monitoring.

6.4. Toronto

The Toronto and Region Conservation Authority (TRCA) has collected water samples in the nearshore environment of Lake Ontario in western Durham (Pickering to Ajax) from 2007 to the present. Gary Bowen, TRCA Watershed Specialist, indicated the authorities' interest in working with researchers to address water quality issues that plague the near shore environment of Lake Ontario. Mr. Bowen characterized the following data collected by the TRCA (excerpts quoted directly from written communication):

> *From 2007-2009 (and in spring 2010), 8 transects were sampled taking mostly surface samples, although there were some samples with depth. The coverage with these 8 transects was spatially extensive, reaching from the Rouge River in the west to Carruthers Creek in the east. Samples were taken at 6 locations along each transect from the shoreline extending out to 3 km (0 m, 100 m, 400 m, 1000 m, 1500 m, 2000 m, and 3000 m). Water samples were analyzed for: E. coli, alkalinity, conductivity, ammonia, nitrate+nitrite, nitrite, total phosphorus, pH, soluble reactive phosphorus, TKN, and suspended solids. Water samples were also taken at the 4 marshes located by the shoreline of the lake (Rouge Marsh, Frenchman's Bay Marsh, Duffins Marsh and Carruthers Marsh) from 2007-2009.*
>
> *From 2011-2013, the number of transects was reduced to 4, but samples were taken at the surface, middle and bottom of the water column, and extended to 5 km. With the added sampling, the number of stations was reduced to 100 m, 1000 m, 3000 m, and 5000 m, and the 4 transects extended from the Pickering Nuclear Generating Station (to the east of the Rouge River) to Carruthers Creek. Additional samples were taken at the Drinking Water Intake and at an instrument which the MOE deploys for the ice off season. Water samples were analyzed for: alkalinity, conductivity, ammonia, nitrate+nitrite, nitrite, total phosphorus, pH, soluble reactive phosphorus, TKN, suspended solids, calculated nitrate, and turbidity.*
>
> *TRCA has committed to sampling the western Durham nearshore approximately 5-8 times per year in the ice off season, for the next 5 years. Changes to our sampling program will include reinstating a transect by the Rouge River, adding sampling in the Frenchman's Bay area, and adding shoreside sampling locations which were omitted from 2011 to 2013. We will also include dissolved phosphorus in our water quality measurements, and plan to profile the temperature, conductivity, turbidity, and possibly light structure through a profiler, at most sampling locations. E.coli will also be measured in the stations closer to shore. In addition, it is possible that some marsh sampling will be reinstated.*

7. Conclusions

In recent years, considerable work has been performed to enhance P load estimates from urban and non-farm rural non-point sources into the Great Lakes and associated inland waters. However, the validity of these estimates is uncertain since extensive knowledge gaps still exist. Based on feedback provided at the three public workshops and a review of current literature critical issues that should be addressed include:

1. A more reliable land cover/use classification scheme that is transferable across studies is critically needed to allow estimate validation across sites
2. There is a need for improved spatial and temporal estimates of atmospheric P loads, particularly the amount of atmospheric dry deposition.
3. The error in sediment transport estimates is critically important and far too large. Future work should be done to better predict the distribution of particle sizes and composition of suspended solids rather than just the total mass of solids.
4. A more complete understanding of urban hydrology, as it relates to P transport, is needed. Specifically, a better understanding of urban groundwater and hydraulic connections to land surface via tile drains and other anthropogenic infrastructure (e.g. sanitary and storm sewers) is needed. It still remains to be determined what role groundwater plays in urban P cycling.
5. There needs for better integration of existing and future datasets. To do this, all future data collected should be obtained via unbiased event-based sampling methods that will allow for direct comparison across sites.
6. Improved understanding of how structural BMPs will alter P cycling.
7. Further elucidate P biogeochemistry, particularly related to kinetics. Fundamental research is needed to describe interactions between different P species (particulate, dissolved, organic, etc.). Does dissolved and particulate P vary seasonally? Are they source dependent? Are they altered by structural BMPs? These are complex questions that when answered will greatly improve P load estimates from urban areas.

Developing a complete picture of P loads produced from urban and non-farm rural non-point sources requires a combination of research, modelling and monitoring. While addressing many of the aforementioned knowledge gaps will take time and estimates will continuously improve, immediate action should be taken to enhance current estimates. Many of the critical needs described above could be addressed relatively quickly (1-3 years) and would greatly improve our understanding of P loading from these areas. While other entities are vital, this effort must be led by a high-level government agency since coordination and long-term stewardship are critical.

8. Recommendations

With the aim of providing estimates of P loads from urban and non-farm rural non-point sources, it is recommend that the Ministry of the Environment coordinate the long-term exchange of nutrient data and conduct an immediate 5 year study to develop accurate P load estimates from these areas. This study should be broken into two phases. During the first phase, it is recommended that 3-4 projects lasting 2-3 years focus on intensively monitoring and modelling P dynamics within select urban systems. Project sites should be selected so that an array of non-rural land cover/use is evaluated. Specific objectives of this effort are to address as many of the critical needs listed above (Section 7) and develop solid estimates of P loads from urban sources. Ideally, project teams should include a mix of academic, private sector and government experts. Project teams should be encouraged to develop different models that could be transferable to other study locations. A minimum level of monitoring coordination is required to ensure all data generated is directly comparable between sites and projects. A technical review board could help to facilitate coordination and should be considered. To ensure data is comparable, it is recommended that the MOE define land cover/use classification used by all projects. Furthermore, analysis should, at a minimum, include some level of P speciation in addition to total P measurements. The primary objective of the first phase is to refine scientific understanding of fundamental processes that can be used to provide more complete estimates for regions not included in the initial study. Outcomes of this study may include, but are not limited to: refined export coefficients that more accurately describe urban land cover/use with temporal specificity (i.e. vary monthly or seasonally), characterization of transport in urban groundwater, quantitative measures of structural BMPs that can be used to modify P loads, or evaluations of modelling schemes and monitoring methods.

Phase two (1-2 years) of this effort should apply the results from Phase I to provide a complete picture of urban and non-farm rural non-point source loads of P to the Great Lakes basin, specifically lakes Erie and Ontario. This portion of the project should focus on developing and evaluating P models capable of predicting P loads from both monitored and unmonitored sites. Developing more than one model and estimate is strongly encouraged. A multi-tier modelling approach should be considered. For example, coupling mechanistic models that describe P cycling at small scales (individual lots, structural BMPs, etc.) with empirical models that describe P transport at the watershed scale. Prior to the start of this exercise an acceptable level(s) of uncertainty in P estimates must be defined by the MOE. Defining this criterion will determine the level of complexity required during modelling. References that provide insight regarding model complexity and should be reviewed prior to developing P load models include those by Zhang, et al. [15], Ramin, et al. [41], Wellen, et al. [42], and Wellen, et al. [43].

9. References

[1] Daloglu, I., K.H. Cho, and D. Scavia, *Evaluating causes of trends in long-term dissolved reactive phosphorus loads to lake erie.* Environmental Science & Technology, 2012. **46**(19): p. 10660-10666.

[2] Zhou, Y., D.R. Obenour, D. Scavia, T.H. Johengen, and A.M. Michalak, *Spatial and temporal trends in lake erie hypoxia, 1987–2007.* Environmental science & technology, 2013. **47**(2): p. 899-905.

[3] Winter, J.G. and H.C. Duthie, *Export coefficient modelling to assess phosphorus loading in an urban watershed.* Journal of the American Water Resources Association, 2000. **36**(5): p. 1053-1061.

[4] Meals, D.W. and L.F. Budd, *Lake champlain basin nonpoint source phosphorus assessment.* Journal of the American Water Resources Association, 1998. **34**(2): p. 251-265.

[5] Wellen, C., G.B. Arhonditsis, T. Labencki, and D. Boyd, *A bayesian methodological framework for accommodating interannual variability of nutrient loading with the sparrow model.* WATER RESOURCES RESEARCH, 2012. **48**(W10505): p. 1-22.

[6] Hamilton Harbour RAP Technical Team, *Contaminant loadings and concentrations to hamilton harbour: 2003-2007 update*, ed. D. Boyd, J.D. Hall, and K. O'Connor. 2010, Burlington, Ontario, Canada.

[7] Kluesener, J.W., *Nutrient transport and transformations in lake wingra, wisconsin*. 1971, University of Wisconsin-Madison: Madison, WI. p. 484.

[8] Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith, *Nonpoint pollution of surface waters with phosphorus and nitrogen.* Ecological Applications, 1998. **8**(3): p. 559-568.

[9] Baker, D.B. and R.P. Richards, *Phosphorus budgets and riverine phosphorus export in northwestern ohio watersheds.* Journal of Environmental Quality, 2002. **31**(1): p. 96-108.

[10] Lehman, J.T., D.W. Bell, and K.E. McDonald, *Reduced river phosphorus following implementation of a lawn fertilizer ordinance.* Lake and Reservoir Management, 2009. **25**(3): p. 307-312.

[11] Lehman, J.T., D.W. Bell, J.P. Doubek, and K.E. McDonald, *Reduced additions to river phosphorus for three years following implementation of a lawn fertilizer ordinance.* Lake and Reservoir Management, 2011. **27**(4): p. 390-397.

[12] Dolan, D.M. and S.C. Chapra, *Great lakes total phosphorus revisited: 1. Loading analysis and update (1994-2008).* Journal of Great Lakes Research, 2012. **38**(4): p. 730-740.

[13] Burton, G.A.J. and R. Pitt, *Stormwater effects handbook: A tool box for watershed managers, scientists, and engineers*. 2001, Boca Raton, FL: CRC Press.

[14] Novotny, V. and H. Olem, *Water quality: Prevention, identification and management of diffuse pollution.* . 1994, New York, New York, USA: Van Nostrand Reinhold.

[15] Zhang, W., D.-K. Kim, Y.R. Rao, S. Watson, S. Mugalingam, T. Labencki, M. Dittrich, A. Morley, and G.B. Arhonditsis, *Can simple phosphorus mass balance models guide management decisions? A case study in the bay of quinte, ontario, canada.* Ecological Modelling, 2013. **257**: p. 66-79.

[16] Cowen, W.F. and G.F. Lee, *Leaves as source of phosphorus.* Environmental Science & Technology, 1973. **7**(9): p. 853-854.

[17] Fissore, C., S.E. Hobbie, J.Y. King, J.P. McFadden, K.C. Nelson, and L.A. Baker, *The residential landscape: Fluxes of elements and the role of household decisions.* Urban Ecosystems, 2012. **15**(1): p. 1-18.

[18] Duxbury, J. and J. Peverly, *Nitrogen and phosphorus losses from organic soils.* Journal of Environmental Quality, 1978. **7**(4): p. 566-570.

[19] Fox, R.L. and E. Kamprath, *Adsorption and leaching of p in acid organic soils and high organic matter sand.* Soil Science Society of America Journal, 1971. **35**(1): p. 154-156.

[20] Miller, M., *Contribution of nitrogen and phosphorus to subsurface drainage water from intensively cropped mineral and organic soils in ontario.* Journal of Environmental Quality, 1979. **8**(1): p. 42-48.

[21] Gotoh, S. and W. Patrick, *Transformation of iron in a waterlogged soil as influenced by redox potential and ph.* Soil Science Society of America Journal, 1974. **38**(1): p. 66-71.

[22] Ozanne, P., D. Kirkton, and T. Shaw, *The loss of phosphorus from sandy soils.* Crop and Pasture Science, 1961. **12**(3): p. 409-423.

[23] Stamm, C., H.P. Jarvie, and T. Scott, *What's more important for managing phosphorus: Loads, concentrations or both?* Environmental Science & Technology, 2014. **48**(1): p. 23-24.

[24] Gentry, L.E., M.B. David, T.V. Royer, C.A. Mitchell, and K.M. Starks, *Phosphorus transport pathways to streams in tile-drained agricultural watersheds.* Journal of Environmental Quality, 2007. **36**(2): p. 408-415.

[25] Tan, C.S. and T.Q. Zhang, *Surface runoff and sub-surface drainage phosphorus losses under regular free drainage and controlled drainage with sub-irrigation systems in southern ontario.* Canadian Journal of Soil Science, 2011. **91**(3): p. 349-359.

[26] U.S. Geological Survey, *National water information system (nwisweb): U.S. Geological survey database at* http://water.Usgs.Gov/nawqa/data. 2001.

[27] Dolan, D.M. and K.P. McGunagle, *Lake erie total phosphorus loading analysis and update: 1996–2002.* Journal of Great Lakes Research, 2005. **31, Supplement 2**(0): p. 11-22.

[28] Long, T., *Email communication with S. McElmurry regarding research at hamilton harbour*, March 20, 2014.

[29] NVPDC and ESI, *Northern virginia bmp handbook: A guide to planning and designing best management practices in northern virginia*. 1992, Northern Virginia Planning District Commission and Engineers and Surveyors Institute: Fairfax, VA.

[30] Winer, R., *National pollutant removal performance database for stormwater treatment practices*. 2000: Center for Watershed Protection Ellicott City, MD.

[31] Stone, M. and M.C. English, *Geochemical composition, phosphorus speciation and mass-transport of fine-grained sediment in 2 lake erie tributaries.* Hydrobiologia, 1993. **253**(1-3): p. 17-29.

[32] SWAMP, *Synthesis of monitoring studies conducted under the stormwater assessment monitoring and performance program*, in *Preapared by the Stormwater Assessment Monitoring and Performance (SWAMP) Program*. 2005. p. 101.

[33] Martin, E.H., *Effectiveness of an urban runoff detention pond-wetlands system.* Journal of Environmental Engineering, 1988. **114**(4): p. 810-827.

[34] Stone, M., *Personal communiction with S. McElmurry during workshop at University of Waterloo March 14,* 2014.

[35] Baker, D.B., D.E. Ewing, L.T. Johnson, J.W. Kramer, B.J. Merryfield, R.B. Confesor Jr, R. Peter Richards, and A.A. Roerdink, *Lagrangian analysis of the transport and processing of agricultural runoff in the lower maumee river and maumee bay.* Journal of Great Lakes Research, 2014.

[36] USEPA, *Urban stormwater bmp performance monitoring: A guidance manual for meeting the national stormwater bmp database requirements. EPA 821-c-02-005.*, Office of Water, Editor. 2002, US Environmental Protection Agency: Washington, DC.

[37] Moffa and Associates, *Wet weather flow assessment protocols*. 2001, Water Environment Research Foundation,: Syracuse, New York.

[38] Borah, D.K., G. Yagow, A. Saleh, P.L. Barnes, W. Rosenthal, E.C. Krug, and L.M. Hauck, *Sediment and nutrient modelling for tmdl development and implementation.* Transactions of the Asabe, 2006. **49**(4): p. 967-986.

[39] Eimers, S.M.R.a.M.C., *Unexpected declines in stream phosphorus concentrations across southern ontario.* Canadian Journal of Fisheries and Aquatic Science, 2013. **71**.

[40] Han, H., J.D. Allan, and N.S. Bosch, *Historical pattern of phosphorus loading to lake erie watersheds.* Journal of Great Lakes Research, 2012. **38**(2): p. 289-298.

[41] Ramin, M., T. Labencki, D. Boyd, D. Trolle, and G.B. Arhonditsis, *A bayesian synthesis of predictions from different models for setting water quality criteria.* Ecological Modelling, 2012. **242**: p. 127-145.

[42] Wellen, C., G.B. Arhonditsis, T. Labencki, and D. Boyd, *A bayesian methodological framework for accommodating interannual variability of nutrient loading with the sparrow model.* Water Resources Research, 2012. **48**(10): p. W10505.

[43] Wellen, C., G.B. Arhonditsis, T. Labencki, and D. Boyd, *Application of the sparrow model in watersheds with limited information: A bayesian assessment of the model*

uncertainty and the value of additional monitoring. Hydrological Processes, 2014. **28**(3): p. 1260-1283.

[44] Whitehead, P.G., L. Jin, H.M. Baulch, D.A. Butterfield, S.K. Oni, P.J. Dillon, M. Futter, A.J. Wade, R. North, E.M. O'Connor, and H.P. Jarvie, *Modelling phosphorus dynamics in multi-branch river systems: A study of the black river, lake simcoe, ontario, canada.* Science of The Total Environment, 2011. **412–413**(0): p. 315-323.

[45] Winter, J.G., *Phosphorus inputs to lake simcoe from 1990 to 2003: Declines in tributary loads and observations on lake water quality.* Journal of Great Lake Research, 2007. **33**: p. 381-396.

[46] Alexander, R., R. Smith, and G. Schwarz, *Estimates of diffuse phosphorus sources in surface waters of the united states using a spatially referenced watershed model.* Water Science & Technology, 2004. **49**(3): p. 1-10.

[47] McMahon, G., R.B. Alexander, and S. Qian, *Support of total maximum daily load programs using spatially referenced regression models.* Journal of Water Resources Planning and Management, 2003. **129**(4): p. 315-329.

[48] Kim, D.-K., W. Zhang, V. Hiriart-Baer, C. Wellen, T. Long, D. Boyd, and G.B. Arhonditsis, *Towards the development of integrated modelling systems in aquatic biogeochemistry: A bayesian approach.* Journal of Great Lakes Research, 2014: p. in press.

[49] Michalak, A.M., E.J. Anderson, D. Beletsky, S. Boland, N.S. Bosch, T.B. Bridgeman, J.D. Chaffin, K. Cho, R. Confesor, I. Daloğlu, J.V. DePinto, M.A. Evans, G.L. Fahnenstiel, L. He, J.C. Ho, L. Jenkins, T.H. Johengen, K.C. Kuo, E. LaPorte, X. Liu, M.R. McWilliams, M.R. Moore, D.J. Posselt, R.P. Richards, D. Scavia, A.L. Steiner, E. Verhamme, D.M. Wright, and M.A. Zagorski, *Record-setting algal bloom in lake erie caused by agricultural and meteorological trends consistent with expected future conditions.* Proceedings of the National Academy of Sciences, 2013. **110**(16): p. 6448-6452.

[50] Wellen, C., G.B. Arhonditsis, T. Long, and D. Boyd, *Accommodating environmental thresholds and extreme events in hydrological models: A bayesian approach.* Journal of Great Lakes Research, in press.

[51] D'Andrea, M., W.J. Snodgrass, and P.D. Chessie, *Development of a wet weather flow management master plan for the city of toronto.* Water quality research journal of Canada, 2004. **39**(4): p. 417-431.

[52] Nixon, S.W., S.L. Granger, and B.L. Nowicki, *An assessment of the annual mass balance of carbon, nitrogen, and phosphorus in narragansett bay.* Biogeochemistry, 1995. **31**(1): p. 15-61.

[53] Ma, J., G.X. Ying, and J.J. Sansalone, *Transport and distribution of particulate matter phosphorus fractions in rainfall-runoff from roadway source areas.* Journal of Environmental Engineering-Asce, 2010. **136**(11): p. 1197-1205.

[54] Doherty, J. and J.M. Johnston, *Methodologies for calibration and predictive analysis of a watershed model1.* JAWRA Journal of the American Water Resources Association, 2003. **39**(2): p. 251-265.

[55] Long, T., C. Wellen, G.B. Arhonditsis, and D. Boyd, *Evaluation of stormwater and snowmelt inputs, landuse and seasonality on nutrient dynamics in the watersheds of hamilton harbour, ontario, canada.* Journal of Environmental Management, submitted.

[56] Baulch, H.M., M.N. Futter, L. Jin, P.G. Whitehead, D.T. Woods, P.J. Dillon, D.A. Butterfield, S.K. Oni, L.P. Aspden, and E.M. O'Connor, *Phosphorus dynamics across intensively monitored subcatchments in the beaver river.* Inland Waters, 2013. **3**(2): p. 187-206.

[57] Winter, J.G., *Email communication with S. McElmurry describing monitoring effort in Lake Simocoe watershed*, *March 18,* 2014.

[58] Milton, T.O., *-draft- terms of reference for bristol survey- phase 1*. 2001, Philips Engineering Ltd.

Appendix

List of workshop participants

Name	**Affiliation**
Barbara Anderson	Ontario Ministry of the Environment
Bill Annable	University of Waterloo
Phil Argiroff	Michigan Department of Environmental Quality
Andrew Bachman	Ohio Environmental Protection Agency
Syeda Basira Banuri	Region of Peel
Brad Bass	Environment Canada
Martin Bateson	Town of Milton
Tirupati Bolisetti	University of Windsor
Theresa Bos	Lake Simcoe Region Conservation Authority
Gary Bowen	Toronto and Region Conservation Authority
Terri Bulman	Ontario Ministry of the Environment
Jan Ciborowski	University of Windsor
Tom Copelan	City of London
Meg Corcoran	US Environmental Protection Agency
Nicola Crawhall	Great Lakes Cities Initiative
Bill Creal	Michigan Department of Environmental Quality
Pat Donnley	City of London
Jennifer Dougherty	Credit Valley Conservation Authority
Paul Drca	City of Windsor
Michael English	Wilfrid Laurier University
Bob Fields	Norfolk County, Regional Public Works Commissioners of Ontairo
Jessie Gardner Costa	University of Windsor
Paul Henshaw	University of Windsor
Eric Hodgins	Waterloo Region
Barb Hodgins	Town of Ajax
Robert Hoffman	What-If Technologies
Neil Hutchinson	Hutchinson Environmental
Phil James	Credit Valley Conservation Authority
Darko Joksimovic	Ryerson University
Karen Jones	Ministry of the Environment
Marcy Kamerath	US Environmental Protection Agency
Francine Kelly-Hooper	Kelly Hooper Environmental Inc.
Wendy Kemp	York Region
Christine Kosmowski	Michigan Water Environment Association
Kevin Kratt	Tetra Tech

Name	**Affiliation**
Berta Krichker	City of London
Aaron Law	Ministry of the Environment
David Lembcke	Lake Simcoe Region Conservation Authority/Conservation Ontario
Luis Leon	City of Burlington
Tanya Long	Ministry of the Environment
Tom Mahood	Kelly Hooper Environmental Inc.
Christina Mallin	The Kingdom of the Netherlands
Paul Mankiewicz	Gaia Institute
Audrey Manuel	Environment Canada
Michal Masior	The Corporation of Haldimand County
Keith McCormack	Hubbell, Roth and Clark
Merrin McCrae	University of Waterloo
Shawn McElmurry	Wayne State University
Denise McGoldrick	City of Waterloo
Kevin Money	Essex Region Conservation Authority
Linda Mortsch	Environment Canada/University of Waterloo
Denisa Necula	City of Markham
Richard Nesbit	Hutchinson Environmental
Weber, Nicole	AECOM
Sherrie Pagsuyoin	University of Waterloo
MaryLou Renshaw	Indiana Dept. of Environmental Management
Jordon Richie	Green Roof Consultant
Rajesh Seth	Univeristy of Windsor
Imtiaz Shah	Upper Thames River Conservation Authority
Robin Skeates	Ministry of the Environment
Bill Snodgrass	City of Toronto
Eleanor Stainsby	Ministry of the Environment
Mike Stone	University of Waterloo
Dave Szeptycki	York Region
Bill Taylor	University of Waterloo
Sabrina Ternier	Ministry of the Environment
Clara Tucker	Ministry of the Environment
Matt Uza	Ontario Ministry of the Environment
Mike Walters	Lake Simcoe Region Conservation Authority
John Watkins	Wayne State University

Name	**Affiliation**
Marcus Williams	What-If Technologies
Jenny Winter	Ministry of the Environment
Brent Wootton	Fleming College Alternative Water Treatment
Maya Zein	Wayne State University