THIS DRAFT PAPER SHOULD NOT BE CITED WITHOUT AUTHOR'S PERMISSION

Modeling Flow, Nutrient and Sediment Delivery from a Large International Watershed using a Field-Scale SWAT model

Awoke Dagnew, Donald Scavia, Yu-Chen Wang, Rebecca Muenich, Colleen Long, and Margaret Kalcic

Environmental Consulting and Technology, Inc. (Dagnew), Ann Arbor MI, 48105, USA; Graham Sustainability Institute (Dagnew (former), Wang, Long), University of Michigan, Ann Arbor MI, 48104, USA; School for Environment and Sustainability (Scavia), University of Michigan, Ann Arbor, MI 48104, USA; Sustainable Engineering and the Built Environment (Muenich), Arizona State University, Tempe, AZ 5281, USA; Food, Agriculture and Biological Engineering (Kalcic), Ohio State University, Columbus, OH 43210, USA (Correspondence to Dagnew: adagnew@ectinc.com).

Research Impact Statement: A well-calibrated and validated flow and water quality model was used to assess nutrient load, concentration, yield, and distribution for a large international watershed.

Abstract: A large international watershed, the St. Clair-Detroit River System (SCDRS), containing both extensive urban and agricultural areas, was modeled using the Soil and Water Assessment Tool (SWAT) model. The watershed, located in southeastern Michigan, US, and southwestern Ontario, Canada, encompasses the St. Clair, Clinton, Detroit, Sydenham, Upper, and Lower Thames sub-watersheds. The SWAT input data and model resolution (i.e., Hydrologic Response Units, HRUs), were established to mimic farm boundaries, the first time this has been done for a watershed of this size. The model was calibrated (2007-2015) and validated (2001-2006) with a mix of manual and automatic methods at six locations for flow and water quality at various time scales. The model was evaluated using Nash-Sutcliffe efficiency (NSe) and percent bias (PBs) and was used to explore major water quality issues. We showed the importance of allowing key parameters to vary among sub-watersheds to improve goodness of fit, and that the resulting parameters were consistent with sub-watershed characteristics. Agricultural sources in the Thames and Sydenham sub-watersheds and point sources from Detroit sub-watershed were major contributors of phosphorus. Spatial distribution of phosphorus yields at HRU and subbasin levels identified locations for potential management targeting for both point and non-point sources and revealed that in some sub-watersheds non-point sources are dominated by urban sources.

Keywords: SWAT; watershed modeling; international watershed; HRUs as fields; flow and water quality

INTRODUCTION

Watersheds are widely accepted units of analysis for water resources planning and management (McKinney et al., 1999; IJC, 2009; Sheelanere et al., 2013), and have been the focus for guiding water resource and management decisions for decades. However, their natural and anthropogenic processes and activities are often too complex and variable, both spatially and temporally, to be captured thoroughly through monitoring alone (Mirchi et al., 2009). Therefore, watershed modeling tools, especially flow and water quality models, have been used increasingly to simulate watershed processes and human use to help guide those decisions at local, national and international scales (Daniel et al., 2011; Singh and Frevert, 2010; Madani and Marino, 2009). These modeling tools are particularly valuable for developing a common understanding and framework for setting goals among nations with shared watersheds (IJC, 2009).

One of the most widely used watershed models is the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), a semi-distributed, physically based flow and water quality model that has been used in watersheds around the world with widely varying characteristics in size and composition (Gassman et al., 2007; 2014). It is designed to capture information ranging from very coarse to fine spatial scales by dividing the watershed into subbasins based on topography, and then dividing the subbasins into smaller Hydrologic Response Units (HRUs) based on unique land use, soil type, slope, and/or management combinations. While these HRUs can be at very fine scales, this increased resolution and complexity improves results only when there is an equivalent level of input information (Johnston and Smakhtin, 2014; Jakeman et al., 2006). Fortunately, in recent years, extensive data sets, such as land-use data generated from remote sensing and tile drainage systems characteristics collected by government and non-government organizations, enable relatively detailed watershed models.

However, even with detailed input data, SWAT still has a large number of parameters that cannot be measured directly and therefore need to be estimated through model calibration (Lie et al., 2010). The most frequently used calibration practice is to evaluate simulation performance at a single downstream location (Shi et al., 2013), which ignores spatial heterogeneity. This is particularly problematic for large systems where parameters estimated for some parts of the watershed may be unrealistic for other parts. For example, Leta et al. (2017) assessed the impact of calibrating at a single site, at multiple sites with constant parameter values, and at multiple sites with varying parameter values for a 1,162 km^2 watershed in Belgium. Their results indicated that using different parameter values among different regions improved calibration results. In their study for a 239 km^2 watershed in Idaho, Zhang et al. (2008)

also showed the importance of calibrating at multiple monitoring sites for better representations of regional conditions and goodness-of-fit. Hence, for large and/or spatially heterogeneous watersheds, calibration/validation processes at multiple locations is crucial to ensure accurate representations of local and regional flow, sediment, and nutrient simulations (Bai et al., 2017; Leta et al., 2017; Wang et al., 2012; Zhang et al., 2008).

A water quality agreement between the United States and Canada (GLWQA, 2016), crafted in response to Lake Erie's re-eutrophication (Scavia et al., 2014), has led to new phosphorous loading targets. Attention has logically been placed on loads from the Detroit and the Maumee rivers because they contribute about 90% of total phosphorus (TP) load to the western basin of the lake (Scavia et al., 2016). While there have been several assessments for the Maumee watershed (e.g., Scavia et al., 2017; Muenich et al., 2016; Kalcic et al., 2016), there has been no similar assessment for the nearly 20,000 km^2 international watershed that drains into Lake Erie from the Detroit River. This study was designed to begin filling that gap with a robust watershed model to allow assessing potential nutrient load reduction strategies.

The goal of this study is to calibrate the SWAT model for this very large, complex international watershed at multiple locations and investigate the spatial distribution of nutrient sources and loads. In pursuit of this goal we first assembled and harmonized into seamless model input US and Canadian data that have their own characteristics, developed with different methodologies and interpretations, and with their own formatting and naming conventions (IJC, 2015).

STUDY AREA

The St. Clair-Detroit River system (SCDRS) drains a 19,040 km^2 watershed area from parts of southeastern Michigan in the US (40% of watershed area) and southwestern Ontario in Canada (60% of watershed area) and contributes its load to Lake Erie through the Detroit River (Figure 1). It is composed of about 50% cropland, 20% urban area, 12% forest, 8% grassland, and 7% water bodies. The US portion of the watershed is dominated by the Detroit Metropolitan area, whereas the Canadian portion is dominated by tile-drained croplands growing corn, soybeans, and winter wheat. Over the 15 years study period (2001-2015), total annual precipitation and annual average temperatures vary between 740 and 1200 mm, and 7.5 and

11.0ºC, respectively, averaging at 908 mm and 9.3ºC. Elevation ranges from 422 m above sea level at the watershed boundary to 145m at the outlet, with mostly flat slopes.

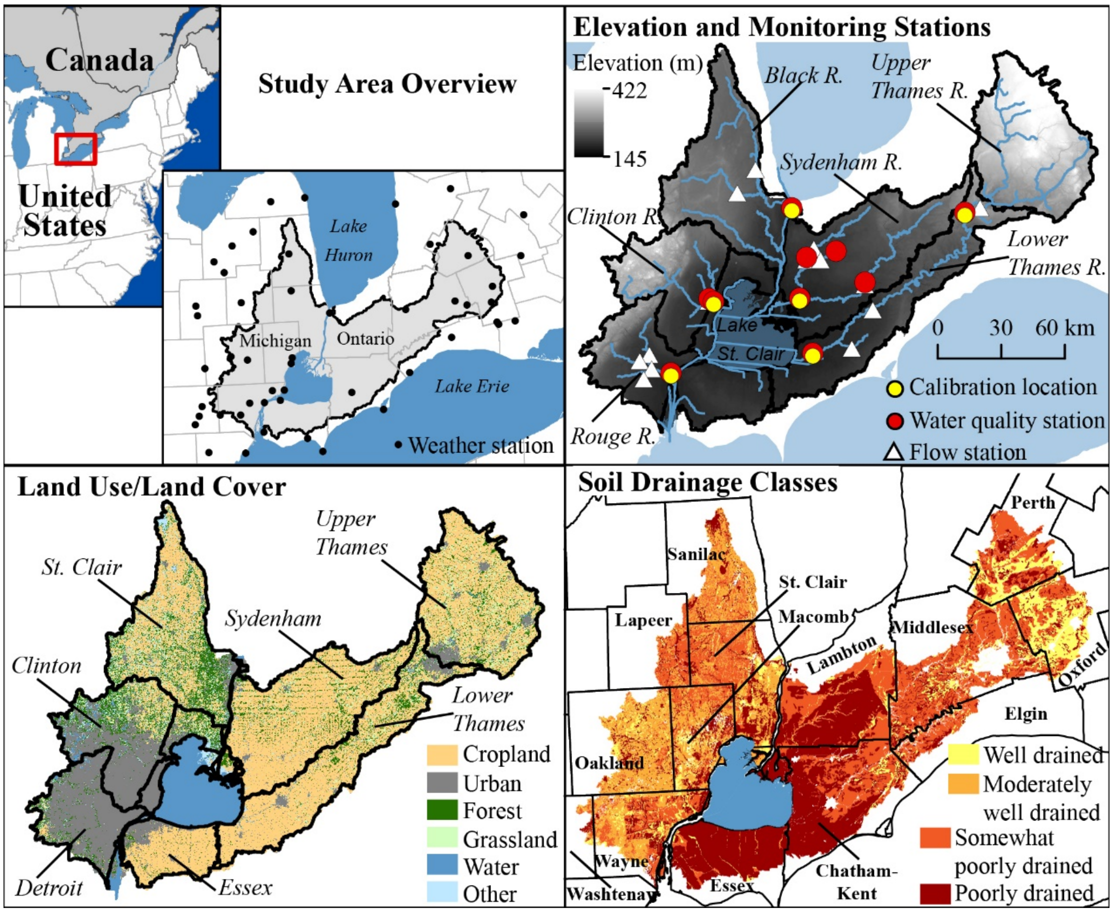

Figure 1: Study area with geographic location and weather stations (top-left), land use/land cover and sub-watershed boundaries (bottom-left), soil and county boundaries (bottom-right) and DEM and calibration locations (top-right) information. The channel which connects Lake Huron to Lake St. Clair is St. Clair River, and Lake St. Clair to Lake Erie is Detroit River. Water flows from Lake Huron to Lake Erie through Lake St. Clair.

The US portion drains three HUC8 watersheds (St. Clair [SC], Clinton [CL], and Detroit [DT] sub-watersheds) drained primarily by the Black River (BR), Clinton River (CR), and Rouge River (RR), respectively. The Canadian portion drains three tertiary watersheds (Upper Thames [UT], and Lower Thames [LT] and Sydenham [SY] sub-watersheds) through the Thames River (TR) and Sydenham River (SR). For this study, the TR includes both Upper

(UTR) and Lower Thames River (LTR) segments. The watershed includes two smaller sub-watersheds, Essex in Canada and Lake St. Clair in the US. While calibration and validation were performed at the outlet of the six major rivers (BR, CR, RR, SR, UTR and LTR), most load assessments were made for the entirety of each sub-watershed (SC, CL, DT, SY, UT and LT) that the major rivers drain. Hence, it is important to note the difference in names between the sub-watershed and river, especially for the Detroit and St. Clair sub-watersheds that are drained through the Rouge and Black rivers.

Overall, 79% of the watershed's agricultural land is in Canada and 83% of the urban land is in the US. The CL and DT sub-watersheds are heavily urbanized (about 56% and 89% of each as urban, respectively), and the SC, SY, UT, and LT sub-watersheds are dominated by agriculture (63%, 89%, and 87% agricultural, respectively). This spatial variation in land use/land cover (LULC) provides both challenges and opportunities for investigating model performance. Moreover, five of the six HUC8 (tertiary) sub-watersheds drain into the 1100 km^2 Lake St. Clair (Figure 1) that retained an average 13% of its TP input over the 1998-2016, and 21% over the 2013-2015 time period (Bocaniov and Scavia, 2018; Scavia et al., 2019).

DATA

Basic inputs

With the exception of data on elevation and weather, all model input was obtained separately for the US and Canada and then merged. DEM data with 30m x 30m resolution from the US Geological Survey–The National Map (USGS, 2016) were used for the entire watershed for elevation, slope, and subbasin delineation. Daily precipitation and maximum and minimum temperatures were obtained from the National Oceanic and Atmospheric Administration's Global Historical Climatology Network (NOAA-GHCN, 2016) for 16 US stations and 15 Canadian stations for 1999-2015 (Figure 1). LULC layers for 2011-2015 with 30m x 30m grid cells were from the US Department of Agriculture National Agricultural Statistics Service (USDA-NASS, 2016) Cropland Data Layer and the Agriculture and Agri-Food Canada Annual Crop Inventory (AAFC, 2016). The 2015 LULC data layer was used to setup the SWAT model and the 5-year data set was used to generate crop rotations. Soil data layers were from the USDA Natural Resources Conservation Service Soil Survey Geographic Database (SSURGO) (USDA-

NRCS, 2017) and from the AAFC's Soil Landscapes of Canada (version 3.2) (AAFC, 2016). Road network data was from U.S. Census Burau (U.S. Census Bureau, 2016. TIGER/Line. Accessed November 2016, https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2015&layergroup=Roads) and Ontario Road Network (Ontario Road Network, 2016. ORN. Accessed November 2016, https://www.ontario.ca/data/ontario-road-network-road-net-element).

Flow and water quality

The USGS National Water Information System (USGS-NWIS, 2016) and the Canadian National Water Data Archive hydrometric data (HYDAT, 2016) were used to obtain daily flow data for the most downstream gauging stations in each sub-watershed (Figure 1, Table S2). Any data gap of 60 days or more was filled using either the stage discharge relationship, if stage data were available, or with the unit area method using data from a nearby station along the same or adjacent stream. If a gap was less than 60 days, it was filled using structural time series (Ryberg and Vecchia, 2017).

Total suspended sediment (TSS), total nitrogen (TN), nitrate (NO3), total phosphorus (TP) and dissolved reactive phosphorus (DRP) concentration data for the US were obtained from the Water Quality Portal (WQP, 2016). Canadian data were from the Provincial Stream Water Quality Monitoring Network (PWQMN, 2016) and Environment and Climate Change Canada (ECCC, D. Burniston and A. Dove, personal communication, 2017). Average sampling frequency ranged from 3 to 17 samples per year for the US and 7 to 21 for Canada.

Because flow and water quality data were often measured at different locations (Figure 1), calibration points were generally at the most downstream water quality stations to avoid extensive interpolation of water quality concentrations and to account for most of the sub-watershed areas. Daily flow data at the calibration locations were estimated using the drainage-area method (Hirsch, 1979) from the upstream flow stations. Monthly and annual nutrient load estimates for calibration at these locations were made using the weighted regression on time, discharge and season (WRTDS) method (Hirsch et al., 2010) based on sample concentration values and daily flow.

Management data layers

Management data layers include cropping systems, fertilizer and manure application rates and placement, tillage practices, and tile drainage. County level fertilizer sales data were from the International Plant Nutrition Institute (IPNI, 2016) for the US and provincial level fertilizer sale data were from Statistics Canada (STATCAN, 2016). Unique application rates for individual crops were based on regional N and P fertilizer application rate information from USDA Economic Research Service (USDA-ERS, 2016) and Canadian Field Print Initiative (Canadian Field Print Initiative, 2017. Accessed March 2017, http://fieldprint.ca/fertilizer-use-survey/). Manure amounts were based on livestock (dairy, beef, swine, sheep, goat, chicken and turkey) counts in each county from USDA-NASS (USDA-NASS, 2016) and from the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA, 2016). Spatial distribution of manure application in Canada was provided by OMAFRA (K. McKague, personal communication, 2017) as locations (points) of animal farms and field areas that receive manure from each animal farm without explicit indication of which field (s).

Tillage practices for sub-watersheds in the US and county/sub-county level for Canada were obtained from USGS and STATCAN, respectively. The latest US tillage data were from 2004, but it detailed practices for each crop type. Canadian data were from 2011, but they did not distinguish among crop types. Data on the distribution of subsurface (tile) drainage systems in Canada were from OMAFRA (2016). Tile drainage information is not available for the US, so we assumed all cropland with poorly drained soils employed tiles (Kalcic et al., 2015). Tile drainage installation depth and spacing specification for the Canadian side of the watershed were recommended to vary by soil type (K. McKague, personal communication, 2017). As such, tile depths were set at 650 mm, 750 mm and 950 mm for clayey, silty, and sandy soils, respectively, with corresponding spacing at 8 m, 12 m, and 15 m, respectively. For the US side, a uniform 1000 mm depth and 20m spacing were used.

Three reservoirs in the upper Thames region (Fanshawe, Wildwood, and Pittock) with surface-area (ha)/volume (ha-m) controls of 262/1235, 192/796, and 142/266, respectively, were included in the model. Information about the physical features of the reservoirs, daily outflow data, and water quality samples were obtained from the Upper Thames River Conservation Authority website (UTCA, 2017) and M. Helsten (personal communication, 2017). Monthly industrial and municipal point source (Figure 2) data were collected from EPA Enforcement and Compliance History (U.S. Environmental Protection Agency, 2017. ECHO. Accessed May 2017, http://tinyurl.com/ybgda4u3) and the Great Lakes Water Authority – Water Resources Recovery Facility (GLWA-WRRF) (M. Khan, C. Willey, personal communication, 2018) for the US, and from OMECC's (Ontario Ministry of Environment and Climate Change) Effluent Monitoring and Effluent Limits (EMEL) Regulations (http://tinyurl.com/y7j9fqhq) for Canada.

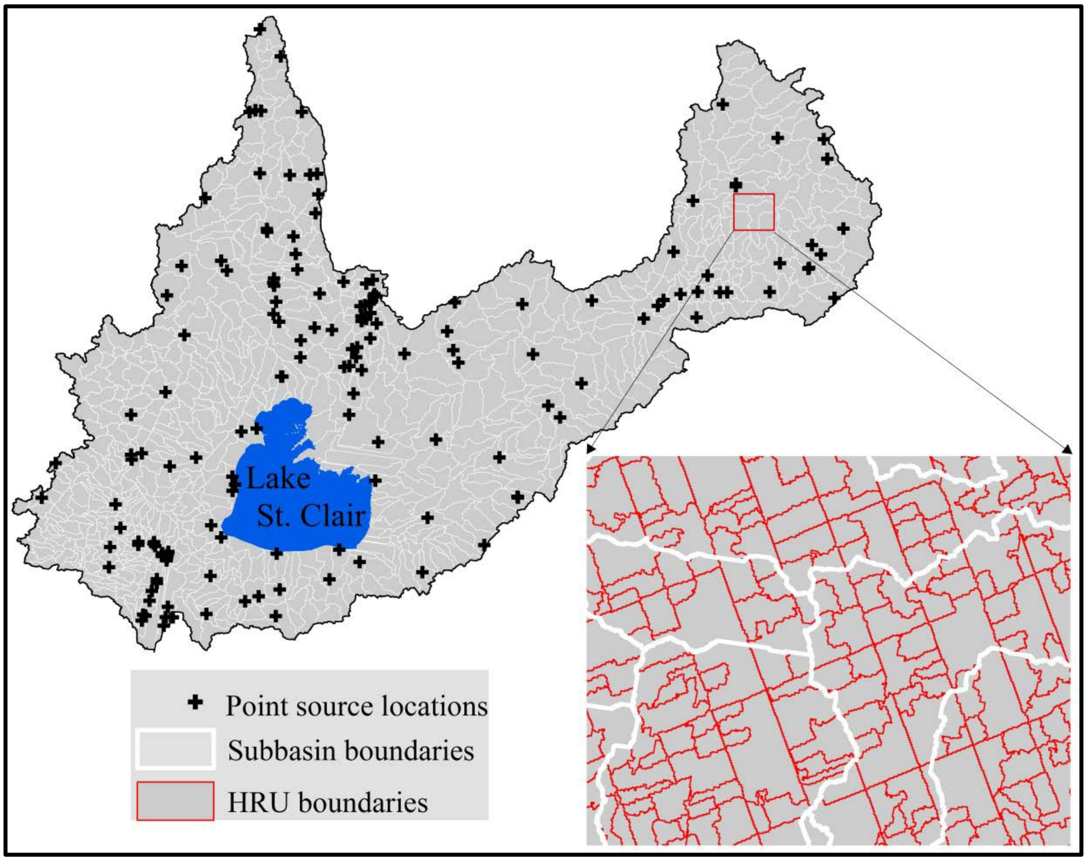

Figure 2: Subbasins and hydrologic response units (HRUs) along with point source locations in the watershed

METHODOLOGY

Data Assimilation

Because this was a binational watershed study, it was essential to ensure data from the two countries were harmonized. The US and Canadian LULC data have the same resolution but different land use type names and identification codes. Because SWAT is based on US data types, Canadian LULC type names and identification codes were converted to the US format (Figure 1). Canadian soil data required additional calculations and unit conversions to conform to US-based SWAT parameters (Table 1). Though there is some anecdotal evidence that Canadian manure production per animal may be different from the US, we used US values for both.

Table 1: Relationship between Canadian versus SWAT major soil parameter names and units, and the changes made

SWAT Soil		**Canadian Soil**		**Comments**	**Equations**
Parameter	***Unit***	***Parameter***	***Unit***		
SOL_ZMX	mm	max(LDEPTH)	cm	*converted*	Unit conversions
SOL_Z	mm	LDEPTH	cm	*converted*	
SOL_AWC	mmH2O/ mm soil	X	X	*Calculated*	SOL_AWC = KP1500-KP33
SOL_K	mm/hr	KSAT	cm/hr	*Converted*	Unit conversions
ROCK	% total weight	COFRAG	% by volume	*converted*	
SOL_ALB	fraction	X	X	*Calculated*	SOL_ALB = 0.4/(0.688*SOL_CBN)
USLE_K	0.013 (t.m2.hr)/ (m3.t.cm)	X	X	*Calculated*	Equation from SWAT I/O documentation (Arnold et al. 2012 Page 307)

Notes: X = parameter not available, SOL_ZMX=max(LDEPTH)= maximum rooting depth of soil, SOL_Z=LDEPTH=depth from soil surface, SOL_AWC=available water capacity of soil, SOL_K=KSAT=saturated hydraulic conductivity, ROCK=COFRAG=rock fragment content, SOL_ALB=moist soil albedo, USLE_K=soil equation erodibility factor, SOL_CBN=organic carbon content of soil, KP1500=water retention at 1500 kP, KP33= water retention at 33 kP

Model setup

Using an area threshold based on the DEM and identification of additional outlet locations to accommodate future comparison and/or spatial verification from smaller sub-watersheds models and/or evolving monitoring efforts, the watershed was divided into 800 subbasins (Figure 2) with an average area of 24 km^2. Smaller subbasins were created in predominantly urban areas to capture their higher variation in drainage and land use types, and to potentially test urban management scenarios in future work at finer spatial scales. Each subbasin was further divided into HRUs using predefined field boundaries as discussed below. The ArcGIS interface, ArcSWAT, version 2012.10_3.18 was used for setup and SWAT2012 rev635, as modified by Kalcic et al. (2016), was used for simulations.

Field boundaries and data processing

LULC, road network, and subbasins were used to define field boundaries using a combination of the methods described by Kalcic et al. (2015) and Teshager et al. (2016). Following Teshager et al. (2016), LULC and road network data were used as the primary sources to identify field boundaries. As such, the watershed was divided into 27,751 "fields" with an average area of about 69 ha, of which 15,219 (54.8%) are cropland. These fields were assigned unique soil type identifiers (Kalcic et al., 2015), and an ArcGIS shapefile that contains the soil identifiers and LULC for each field was created. The shapefile was then used to define HRUs in the ArcSWAT model setup with 0% thresholds for LULC, soil, and slope, and the 27,751 fields thus became the SWAT HRUs (Figure 2).

A key advantage of using field boundaries to generate HRUs is that management practices can be assigned at a more detailed spatial scale than in more traditional SWAT models. Crop rotations for each HRU were estimated by overlaying the 2011-2015 LULC data layers and extracting the major cropping systems in each cropland fields. The most dominant crop rotations involved corn, soybeans, and winter wheat. In order to maintain a manageable number of rotations, crop rotations were limited to a maximum of three years. Tile drainage data and field boundaries were overlaid to determine fields with tile drainage systems. If the majority of a field was covered by the tile drainage layer, the field was considered to have tiles. Canadian fields (HRUs) that receive manure were determined based on proximity to animal farm location and total field area receiving manure from the animal farm.

The field boundaries were also used to distribute the county level conventional (Cv), conservation (Cs), and no-till (NT) tillage practices. The type of tillage practices assigned for a crop field in a county depended on the proportions of practices (Cv:Cs:NT) in that county and the cropping system (crop rotation) in the field. Conventional tillage practices were assigned more in fields with intensive corn, single crop, or non-alternate rotations (e.g., continuous corn). On the other hand, more conservative tillage practices (Cs and NT) were assigned more in fields with alternate rotations (e.g., corn-Soybeans-Winter wheat). Given this information on field-scale crop rotations and regional application rates of mineral N and P for different crops, a similar approach was used to allocate county/provincial level fertilizer applications across agricultural HRUs. Corn fields generally received N and P fertilizer at higher application rates than winter wheat or soybeans. Corn in continuous-corn rotation received more mineral fertilizer than corn in any other alternate rotations (Table S1).

The field boundaries were also designed for analysis and display of input and output information (e.g., distribution of fertilizer/manure application, flow, phosphorus load, etc.), and to model infield best management practices (BMPs) (e.g., filter strips, grassed waterways, drainage management, etc.) at finer scales.

Calibration and validation

Calibration and validation were performed at the outlets of the three US sub-watersheds and the three Canadian sub-watersheds (Figure 1). The model simulated 1999-2015, using the first two years as the warm-up period. Flow was calibrated for 2007-2015 and validated for 2001-2006 at daily, monthly, and annual time scales. Upon successful flow calibration, the model was calibrated for total suspended sediment loads, followed by nutrients (TN, NO3, TP, and DRP) at daily time steps. Since monthly and annual scales were more relevant for management application and policy advice, water quality parameters were further adjusted to also match WRTDS's monthly and annual water quality loads.

The significant variation in LULC and land management across such large watershed was expected to result in different controlling dynamics, especially physical drivers. Therefore, during calibration, certain subbasin and HRU parameters were allowed to vary across the six major sub-watersheds (Table S3, S4).We used both manual calibration and SWATCUP's SUFI2 (Abbaspour, 2015) auto-calibration procedures. Watershed level parameters were initially adjusted manually based on experience and information about local conditions. For example, parameters that control snow cover were estimated based on comparisons of observed and simulated snowfall frequency and snow depth values for the area. Then, SUFI2 was used to estimate HRU and subbasin parameter values and to understand their general direction of change in each major sub-watershed. Finally, manual calibration was used for all parameters to improve fit.

Model performance was evaluated by comparing observed and simulated values using three commonly used statistics for watershed modeling: coefficient of determination (R^2), Nash-Sutcliffe efficiency coefficient (NSe), and percent bias (PBs).

The NSe is used to assess how good simulated values fit observations. The NSe values range from 1 to -∞ with 1 being a perfect 1:1 fit between simulated and observed values. PBs provides insights on the tendency of simulations in under- or over-estimating values, and ranges from -∞ to +∞. A PBs value of 0.0% indicates a perfect match between average simulated and observed values, and negative and positive values show under- and over-estimation, respectively. The R^2 values examine how well simulated values are correlated with observations, i.e., follow similar trends; 0.0 indicates no correlation and 1.0 a perfect correlation. According to Moriasi et al. (2007), monthly simulations with NSe > 0.75 are considered "very good", > 0.65 and ≤ 0.75 are "good", > 0.50 and ≤ 0.65 are "satisfactory", and values ≤ 0.50 are "unsatisfactory" for watershed models. Similarly, values of |PBs| < 10%, 10% - 15%, 15% - 25%, and ≥ 25% fall into those same categories for flow simulations. The same categories apply for sediment if |PBs| < 15%, 15% - 30%, 30% - 55%, and ≥ 55% and for nutrients |PBs| < 25%, 25% - 40%, 40% - 70%, and ≥ 70%.

Finally, to evaluate the significance of allowing parameters to vary among sub-watersheds, the final calibrated flow parameter set for each sub-watershed was assigned uniformly across the entire watershed and NSe and PBs were compared to those for the varying parameter case. As a result, six sets of statistics for each sub-watershed were compared.

RESULTS AND DISCUSSION

Input Characterization

Using the spatial allocation scheme (HRU boundaries), we distributed crop rotations, fertilizer/manure applications, tile drainage, and tillage practices for each HRU explicitly (Figure 3) to better represent actual conditions. With respect to cropping systems, three-year rotations involving corn (C), soybeans (S), and winter wheat (W) covered about 43% of the cropland area. Distribution of crop rotation types was similar within each country, with CSW dominating, followed by CS and then SS (Table 2). However, corn-only or soybeans-only cropping systems were more abundant in Canada than the US (Figure 3), and 40% of the Canadian soybean intensive fields were in the Essex region. Crop rotations for each county and HUC8/tertiary sub-watershed are detailed in Figure S1 and S2.

Table 2: Percentages of cropland area covered with the different types of crop rotations divided between US and Canada (C=corn, S=soybeans, W=winter wheat)

Crop rotation	% cropland area		
	Canada	US	Overall
CC	8.4	1.6	7.1
CS*	25.4	35.5	27.3
SS	13.5	13.1	13.4
CSW**	42.8	45.4	43.3
SW	0.4	0.3	0.4
SSW	9.5	4.1	8.5
Total	100.0	100.0	100.0

*Includes both CS and SC rotations
**Includes CSW or SWC or WCS rotations

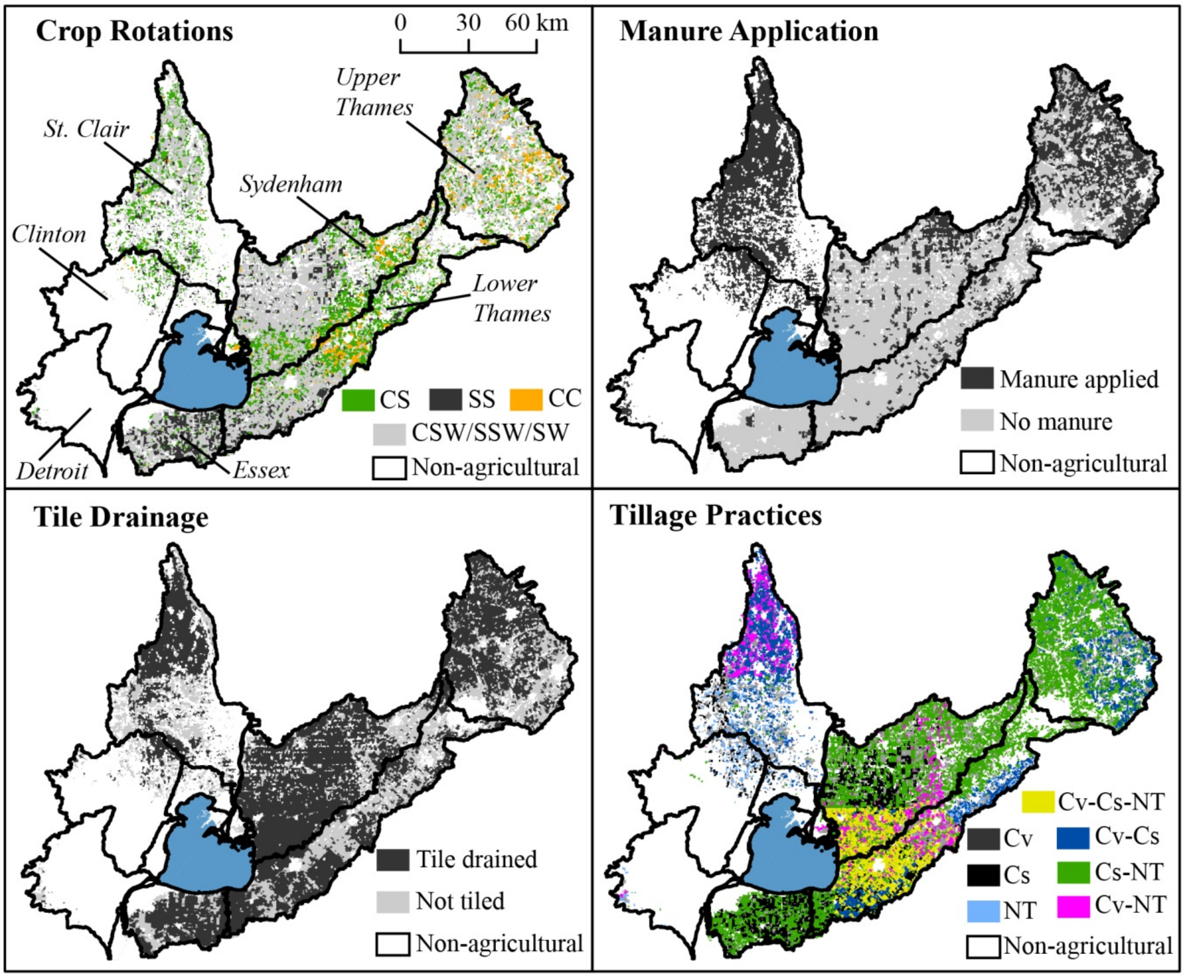

Figure 3: HRU-level agricultural management practice model inputs (C=Corn, S=Soybeans, W=Winter wheat, Cv=Conventional tillage, Cs=Conservation tillage, NT=No-till)

Allocation of conventional (Cv), conservation (Cs), and no-till (NT) tillage practices (Figure 3) resulted in about 70% of cropland receiving alternating practices with either two or three tillage types (Figure 4). The most dominant tillage practice was Cs-NT (39.4%) and was mainly in Canada. US croplands were dominated by Cv-Cs tillage. While cropping systems that alternate corn-soybeans-winter wheat in a three-year rotation received all three tillage practices, most of the continuous conventional tillage practices were assigned for single crop rotations (Figure 5).

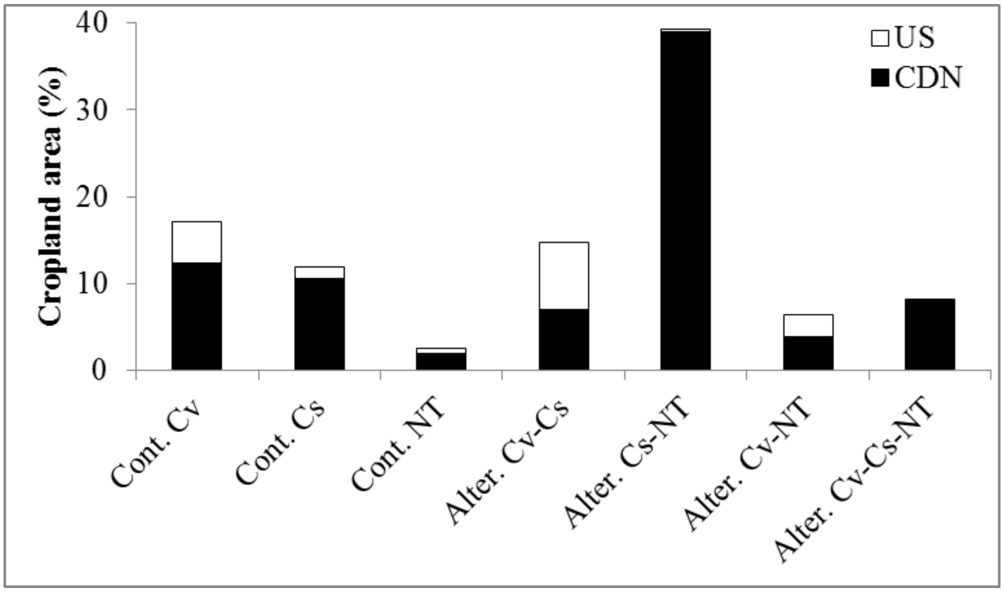

Figure 4: Estimated distribution of tillage practices in US and Canadian parts of the SCDRS watershed (Cont. Cv=Continuous conventional, Cont. Cs=Continuous conservation, Cont. NT= Continuous No-Till, Alter. = Alternating)

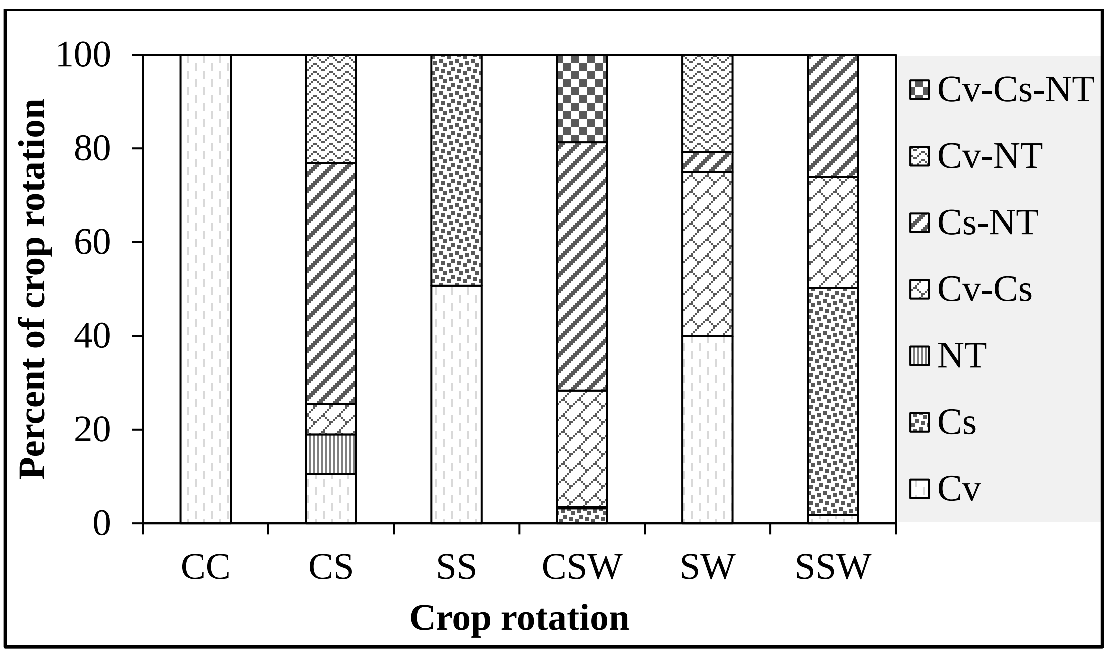

Figure 5: Estimated relationship between tillage practices and crop rotations (C=Corn, S=Soybeans, W=Winter wheat, Cont. Cv=Continuous conventional, Cont. Cs=Continuous conservation, Cont. NT= Continuous No-Till, Alter. = Alternating)

Tile drainage was denser in Essex region, lower parts of SY and LT, and upper parts of SC and UT sub-watersheds (Figure 3). About 67% of Canadian and 55% of US agricultural areas were considered tiled (Table 3). Most of the UT and upper parts of SY agricultural fields receive manure generated in their respective counties while few fields in LT and Essex area received manure. In the US, manure was assumed to be distributed across all agricultural fields, and because of this and fewer livestock, solid manure application rates in the US were lower (85-670 kg/ha for dairy, 8-50 kg/ha for Beef and 1-35 kg/ha for swine) than in Canada (345-1082 kg/ha for dairy, 261-695 kg/ha for Beef and 667-1556 kg/ha for swine).

Table 3: Percentages of agricultural area with tile drainage systems divided between US and Canada at sub-watershed level

HUC8/Tertiary name	Tiled area	
	% total area	% agricultural area
St. Clair (SC)	37	59
Clinton (CL)	8	46
Detroit (DR)	1	16
Lake St. Clair	5	29
U.S. total	**18**	**55**
Upper Thames (UT)	54	62
Lower Thames (LT)	49	55
Thames total	51	59
Sydenham (SY)	69	77
Essex	58	72
Canada total	**58**	**67**
Watershed total	**42**	**64**

Calibration and Validation

Flow. The model reproduced observed flow hydrographs fairly well (Figure 6). Using Moriasi et al. (2007) performance criteria, the monthly flow calibration NSe (Table 4) were judged "very good" for the ULT, LTR, and SR sub-watersheds; "good" for BR and RR; and "satisfactory" for CR. PBs during calibration and both NSe and PBs during validation for all six locations were rated as "very good". The model also performed well at daily (NSe > 0.5 except BR, and |PBs|<10%) and annual (NSe >0.65 and |PBs| < 10%) time scales (Table S5).

Table 4: Monthly flow estimation performance statistics for calibration (2007-2015) and validation (2001-2006) years (R^2 = coefficient of determination, NSe = Nash-Sutcliffe efficiency, PBs = percent bias)

Statistics	***Monthly statistics for flow calibration(validation) period***					
	Upper Thames River (UTR)	**Black River (BR)**	**Sydenham River (SR)**	**Clinton River (CR)**	**Lower Thames River (LTR)**	**Rouge Rover (RR)**
R^2	0.84(0.93)	0.72(0.76)	0.85(0.87)	0.63(0.80)	0.87(0.92)	0.71(0.78)
NSe	0.84(0.93)	0.72(0.76)	0.85(0.86)	0.53(0.75)	0.87(0.91)	0.70(0.75)
PBs	0.1(3.2)	9.2(-2.9)	-1.2(8.4)	-2.7(1.9)	-2.7(5.4)	-1.1(-8.5)

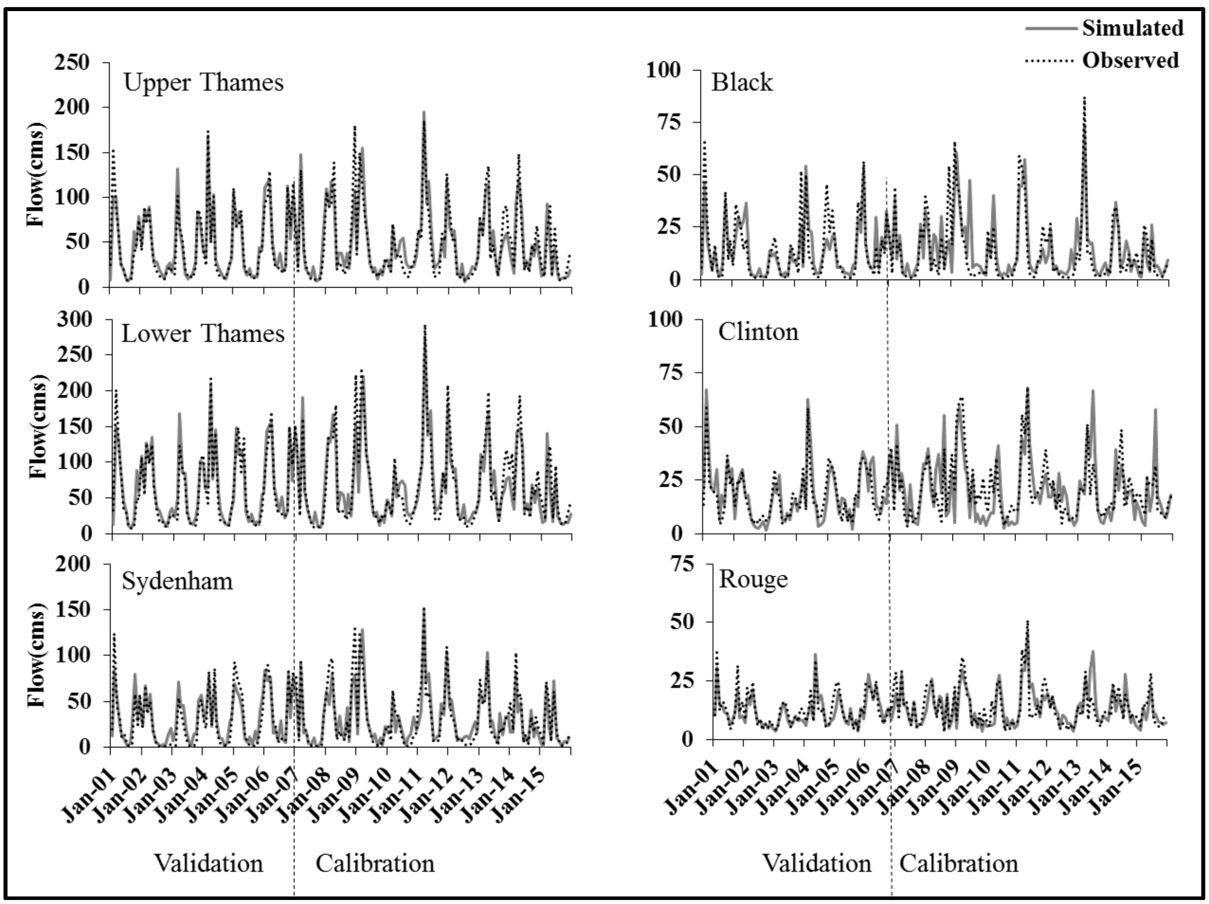

Figure 6: Monthly observed and estimated flow time series at each major sub-watershed outlet locations for both calibration (2007-2015) and validation years (2001-2006)

As expected, allowing parameters to vary among sub-watersheds provided a better representation of regional conditions and improved model performance (Tables S2 and S3). During calibration, some flow parameter values varied substantially across the watershed, especially between agricultural- and urban-dominated sub-watersheds (Tables S4). Flow was particularly affected by changes in parameters for main channel average width (CH_W2) and/or depth (CH_D) and average slope (CH_S2) in both of the highly urbanized streams (CR and RR). This adjustment for urban streams is consistent with the fact that urbanization not only increases runoff but also alters routing of flow downstream through changes in channel dimensions (Booth, 1990; Baker et al., 2008).

The calibration also resulted in substantially lower soil water capacity parameter values (SOL_AWC) in urbanized areas, consistent with the fact that urbanization reduces soil permeability, infiltration, and water holding capacity through soil disturbance, displacement, pore space reduction, low organic matter, and high surface traffic (Craul, 1985; Jim, 1998; Yang and Zhang, 2015; Wiesner et al. 2016). For example, the European Commission Bio Intelligence Serve (2014) reported that changing forest land to urban land could decrease the maximum soil water content by up to 25%.

Differences in other parameter values, such as increasing the runoff curve number from the SWAT default value for moisture condition II (CNII) for the UT by 10% and the LT by 4% reflected the differences in slopes between the two regions (~0.12% and ~0.03%, respectively, along the main stream course). These two regions also have different soil drainage class distributions. While the UT has more well drained soils, the LT is dominated by poorly drained soils. As such, SOL_AWC was increased by 10% above the default value and the soil evaporation compensation factor (ESCO) was set at 0.90 for the LT, compared to an ESCO value of 0.30, and the default value for SOL_AWC for the UT. The increase in SOL_AWC for the LT reflected the higher water holding capacity of the poorly drained soils. Moreover, the higher ESCO value for the UT was consistent with its higher water holding capacity of the soil that compensated for evaporation.

Overall, comparison of the final flow calibration statistics (Table 4) against statistics from uniform parameters across the entire watershed (Table S6) showed the strength of varying parameter values. If, for example, parameters which were best for UTR flow conditions were used across the watershed, the NSe values for CR, BR and RR would have dropped by 62%, 11% and 6%, respectively, and the |PBs| values for CR, BR and SR would have increased by 34.3%, 29.2%, and 12.7%, respectively. Similarly, if best parameter sets for CR flow conditions were used across the watershed, |PBs| values would have increased by 25.4%, 19.6%, 13.6%, 12.5%, and 11.9%, for RR, BR, LTR, UTR and SR, respectively, and the NSe values for RR and BR would have dropped by 34% and 14%.

A closer look at the effects of parameter values from one sub-watershed applied to another indicated that even exchanging parameter sets between urbanized sub-watershed (CR, RR) reduced fit. For example, using the CR optimal parameter values for the RR reduced its NSe and increased its PBs values by 34.3% and 25.4%, respectively. The RR parameter values had similar effects for the CR. Interestingly, while parameter values from the agricultural sub-watershed (SY) reduced fit for the urbanized river (CR), the urbanized sub-watershed (CL) parameters had less impact on the agricultural one (SR).

Water quality. Measured nutrients and sediment dynamics were also replicated sufficiently (Figure 7, Table 5, Figure S4-S7). Monthly water quality calibration and validation statistics were better for TP than DRP and better for TN than NO3. All calibrations and validations were rated as “good” or better for PBs. Most calibration and validation NSe values were rated as “good” or “satisfactory”. However, the phosphorus-related NSe values for UTR calibration were unsatisfactory, as was the RR validation, and both calibration and validation for the BR. Similar to flow, ratings for the major rivers in agricultural sub-watersheds (SR, LTR and UTR) were better than river in urbanized sub-watersheds (CR and RR).

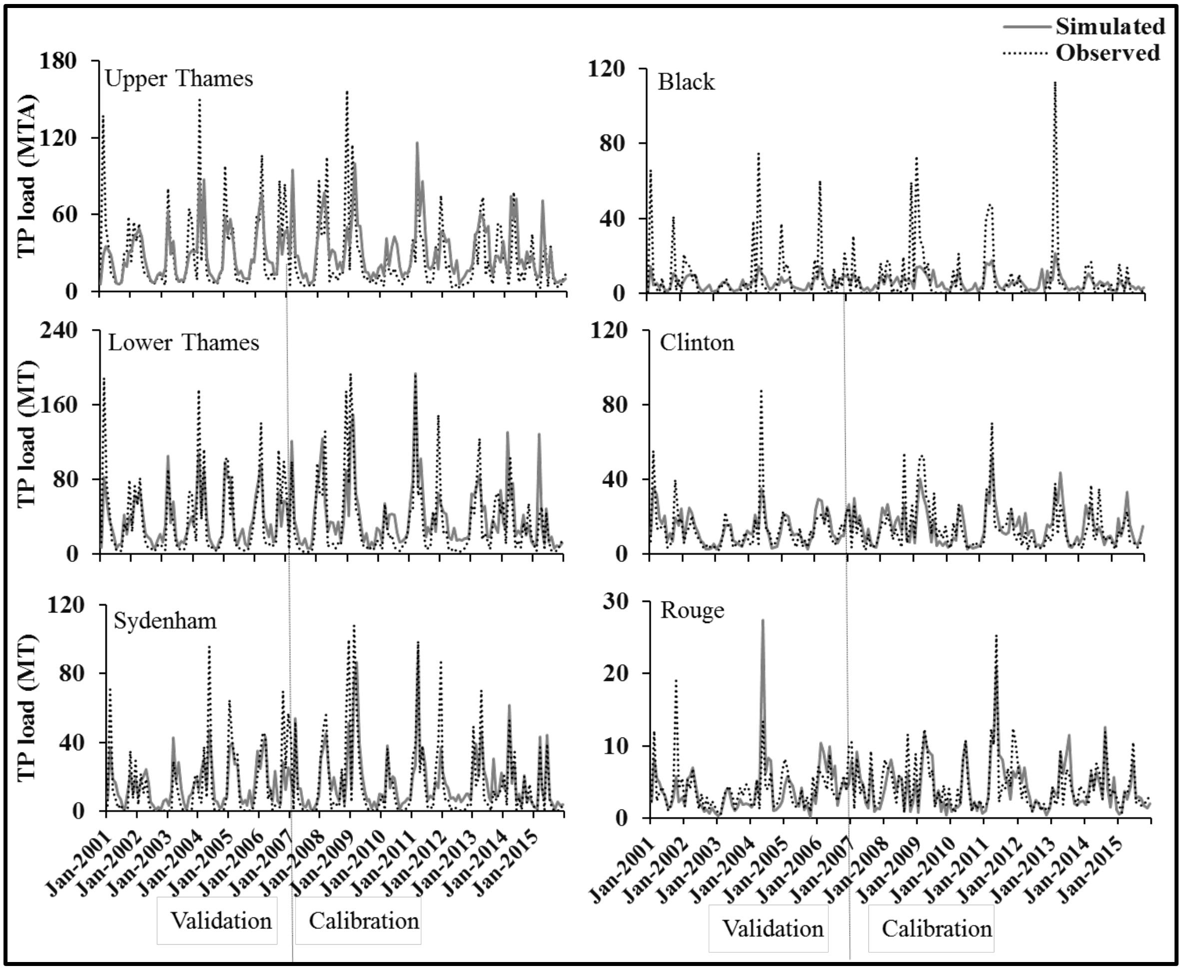

Figure 7: Monthly observed and estimated total phosphorus (TP) time series at the six major sub-watershed outlet locations for both calibration (2007-2015) and validation (2001-2006) periods

Table 5: Monthly water quality model performance statistics for calibration (2007-2015) and validation (2001-2006) years. PBs and NSe ratings: **bold** = "unsatisfactory".

	Statistics	Monthly statistics for water quality calibration(validation)					
		Upper Thames	**Black**	**Sydenham**	**Clinton**	**Lower Thames**	**Rouge**
TP	**R^2**	0.54(0.63)	0.54(0.59)	0.75(0.68)	0.64(0.55)	0.62(0.75)	0.73(0.42)
	NSe	**0.48**(0.59)	**0.29**(**0.25**)	0.73(0.62)	0.64(0.54)	0.59(0.70)	0.71(**0.10**)
	PBs	22.6(9.7)	-25.6(-29.1)	5.9(6.3)	5.6(4.8)	18.0(9.6)	-5.0(-4.8)
DRP	**R^2**	0.44(0.59)	0.48(0.50)	0.64(0.57)	0.57(0.51)	0.55(0.65)	0.71(0.49)
	NSe	**0.42**(0.52)	**0.26**(**0.21**)	0.53(0.52)	0.51(**0.46**)	0.52(0.58)	0.70(**0.05**)
	PBs	27.8(12.1)	-28.7(-35.2)	-6.3(-8.2)	9.6(7.8)	21.5(10.9)	25.1(14.8)
TN	**R^2**	0.61(0.65)	0.52(0.55)	0.72(0.65)	0.55(0.54)	0.59(0.66)	0.64(0.53)
	NSe	0.54(0.57)	**0.27**(**0.32**)	0.70(0.61)	0.54(0.52)	0.57(0.62)	0.61(**0.40**)
	PBs	7.8(13.9)	36.4(42.9)	17.9(23.4)	-15.8(-14.6)	-8.0(8.6)	-5.2(-11.4)

NO_3	**R^2**	0.55(0.52)	0.49(0.47)	0.56(0.52)	0.48(0.48)	0.58(0.66)	0.63(0.42)
	NSe	0.53(**0.49**)	**0.25**(**0.27**)	0.54(**0.47**)	**0.44**(**0.42**)	0.53(0.55)	**0.44**(**0.21**)
	PBs	15.6(14.2)	-24.7(-31.1)	5.9(6.3)	-27.3(-23.4)	-3.0(13.6)	-15.1(-24.8)
TSS	**R^2**	0.66(0.77)	0.61(0.62)	0.73(0.67)	0.57(0.63)	0.67(0.70)	0.61(0.68)
	NSe	0.59(0.62)	**0.49**(0.52)	0.57(0.55)	**0.47**(0.57)	0.60(0.65)	0.58(0.60)
	PBs	-7.5(-2.9)	-15.6(-9.9)	14.3(11.6)	-16.5(-12.4)	-12.0(-7.9)	-14.0(-18.4)

Note: TP = total phosphorus, DRP = dissolved reactive phosphorus, TN = total nitrogen, NO_3 = nitrate, TSS = total suspended sediment, R^2 = coefficient of determination, NSe = Nash-Sutcliffe efficiency, PBs = percent bias)

Similar to flow, some water quality parameters vary considerably across sub-watersheds (Table S4). For example, values of initial nitrate concentration in the soil layer (SOL_NO3) were set to 100 mg N/kg-soil for UT and SY, whereas values for CL and DT were 25 and 0 mg N/kg-soil, respectively, perhaps reflecting differences in soil fertility. The rate constant for in-stream mineralization of organic phosphorus to dissolved phosphorus (BC4) was higher for Canadian rivers (0.28 day^{-1}, 0.25 day^{-1} and 0.16 day^{-1} for SR, UTR and LTR, respectively) than for US rivers (0.018 day^{-1} for all BR, CR, RR), suggesting potentially higher concentrations of DRP in Canadian streams. There are also distinct differences in parameter values between UT and LT sub-watersheds. Almost all nutrient parameter values were higher for UT than LT, implying higher initial soil nutrient content and increased nutrient yields in the UT compared to LT.

Nutrient load assessments

Because phosphorus is the primary driver of interest in Lake Erie (Scavia et al., 2014; 2016), we focus primarily on phosphorus loading.

Annual average loads. The DT and the Thames (UT and LT) sub-watershed loads were similar and together contribute >60% of the TP and >70% of the DRP loads on an average annual basis (Table 6). However, about 90% of TP and DRP load from the DT sub-watershed came from point sources, mainly one waste water treatment plant, whereas about 90% of the load from the Thames comes from agriculture. Despite being mainly urban, the CL sub-watershed load came primarily from non-point source runoff, with combined urban and agricultural non-point sources accounting for 83% and 68% of Clinton's TP and DRP loads, respectively. Moreover, urban non-point source accounts for about 68% and 75% of CL's total non-point source TP and DRP loads, respectively. Phosphorus loads from the SY, the most agriculturally intense sub-watershed, accounted for 13% of the overall watershed's TP and DRP loads. Among the six sub-watersheds, the SC delivered the lowest loads (10% and 5% of TP and DRP, respectively). The smaller sub-watersheds (Essex and Lake St. Clair; Figure 1) contributed 4.4% and 0.8% of TP, and 2.5% and 0.5% of DRP loads, respectively. Even though the Essex region sub-watershed area was about twice that of the Lake St. Clair sub-watershed, it delivered about five times the phosphorus load due to extensive agriculture and densely tiled soils.

Table 6: Average annual total phosphorus (TP) and dissolved reactive phosphorus (DRP) loads in MTA (metric ton per annum) from both point sources (PS) and non-point sources (NPS) for each sub-watershed

HUC8/Tertiary watershed name	Total PS		Total NPS		Total Load		Drainage Area (km^2)
	TP	DRP	TP	DRP	TP	DRP	
St. Clair	28	15	150	21	177	36	3025
Clinton	33	18	158	39	191	57	1969
Detroit	492	257	55	30	547	287	1594
Lake St. Clair	5	3	9	1	14	4	575
U.S. Total	**558**	**293**	**372**	**91**	**929**	**384**	**7163**
Sydenham	26	12	201	83	227	95	3508
Thames	51	24	472	224	523	248	5827
Essex	6	3	71	16	77	19	1098
Canada Total	**83**	**39**	**744**	**323**	**827**	**362**	**10433**
Watershed Total*	**641**	**332**	**1116**	**414**	**1756**	**746**	**17596**

*This does not include Lake St. Clair and other small unaccounted areas along St. Clair and Detroit connecting channels

DRP represented 42% of the TP load overall; however, it was 52% of the point sources and 37% of the non-point source TP load. While this variation in the DRP/TP ratio did not seem to be correlated with the composition of LULC, there were clear differences among different sources. The DRP fraction from US non-point sources was much lower than from Canadian non-point sources, likely due to extensive tile drainage in the Canadian portion. In contrast, US point sources had higher DRP fractions.

Our annual average TP load estimates were similar to the WRTDS-based averages reported by Scavia et al. (2019) because our model was calibrated to WRTDS estimates (Figure 8). Our estimates were also similar to Maccoux et al. (2016) for the CR and BR, somewhat higher for the SR and TR, but considerably lower for the RR. Maccoux et al. (2016) and we used the same water quality monitoring station for the Rouge River (Figure 1), but Maccoux et al. considered the drainage area for the station to be 565 km^2 whereas the actual drainage area for the station was 1,200 km^2 (USGS, https://waterdata.usgs.gov/nwis/nwismap/?site_no=04168550&agency_cd=USGS). Hence Maccoux et al.'s TP estimations for RR were overestimated because they overestimated unmonitored loads. Our annual average DRP load estimates showed similar discrepancies with Maccoux et al. (2016). Our estimate was much lower for the RR and much higher for the TR (Figure 11). Other discrepancies among the three studies could be due to the lack of more frequent water quality sample data, inherent differences in structure and assumptions of different estimation techniques, and span of years considered for the studies. For example, Maccoux et al. (2016) estimates for 2003-2013 used the Stratified Beale's Ratio Estimator (Beale, 1962; Dolan et al., 1981), Scavia et al (2019) estimates for 1998-2016 used WRTDS, and our estimates for 2001-2015 used SWAT.

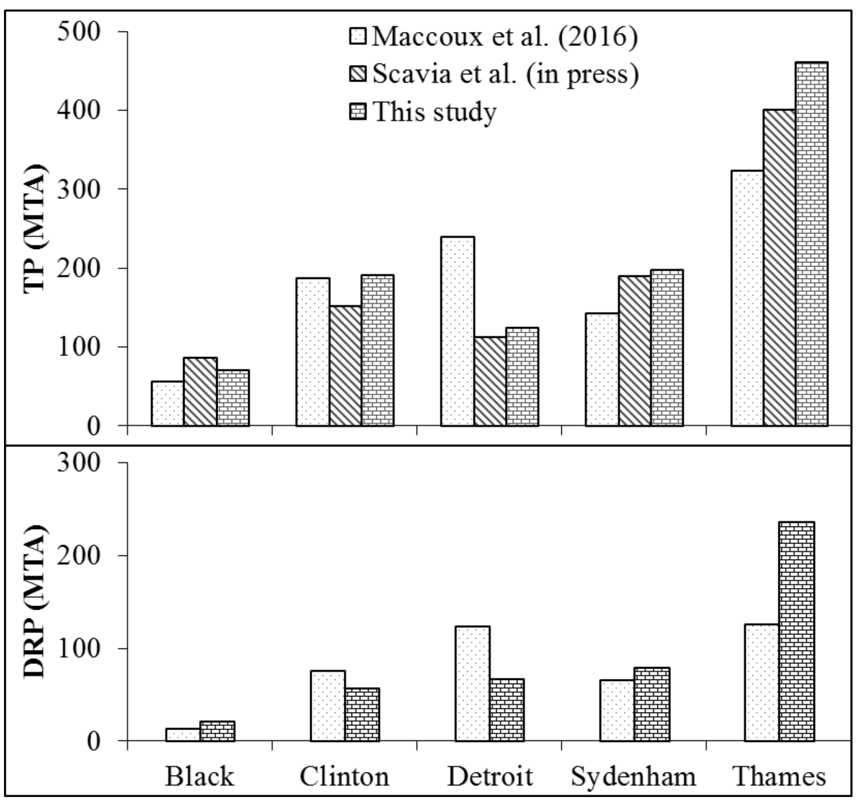

Figure 8: Comparisons of average annual phosphorus load estimations of total phosphorus (TP, Top), and dissolved reactive phosphorus (DRP, bottom), for each major sub-watershed. The

Detroit sub-watershed loads in this figure do not include the GLWA's (Great Lakes Water Authority) waste water treatment point source loads.

In our analysis, annual TP loads increased slightly for all but CR between 2001 and 2009 and then decreased through 2015, with the trends more obvious for rivers in the agriculture dominated areas: SR, TR, and BR (Figure S3). On average between 2001 and 2009, TP increased by 24.7 MTA, 14.8 MTA, 4.1 MTA, and 1.6 MTA for TR, SR, Black, and RR, respectively. The decreases in TP between 2010 and 2015 were of 42.2 MTA, 23.7 MTA, 8.9 MTA, and 4.0 MTA, respectively. DRP followed similar trends, especially for the three rivers in agricultural sub-watersheds, but to a lesser degree than TP, with DRP increases of 8.6 MTA, 4.4 MTA, 1.1 MTA and 0.8 MTA, and decreases of 20.0 MTA, 9.7 MTA, 2.5 MTA, and 1.1 MTA for the same time intervals and river orders. Similar trends have been reported for the Maumee River (Baker et al. 2014), another major P contributor to Lake Erie. In most cases, these trends were reflecting changes in flow (Figure S3) but flow alone could not explain the trend for the TR and SR where flow was relatively constant between 2001 and 2005. It appears that, in those cases, agricultural practices that provide access to more nutrient (e.g., high fertilizer applications) and facilitate nutrient movement into streams (e.g., tile drainage systems) are also responsible for these trends.

Spatial distribution of yields - Sub-watershed scale. Examining sub-watershed and HRU yields provide information potentially useful for targeting management actions to the highest source areas. While the average annual TP loads from the DT and Thames sub-watersheds were similar (Table 6), TP yields (3.43 kg /ha and 0.90 kg /ha, respectively) and DRP yields (1.80 kg /ha and 0.43 kg /ha, respectively) differ considerably due to the difference in drainage areas. In addition, the Thames delivered much more phosphorus from non-point sources (0.81 kg TP/ha and 0.38 kg DRP/ha) than the DT sub-watershed (0.35 kg TP/ha and 0.19 kg DRP/ha) (Figure 9). The Thames and CL sub-watersheds had similar overall TP yields; however, DRP yield was higher for the Thames. The SY and SC sub-watersheds had comparable TP yields but the SY produces much higher DRP per hectare. Overall, the TP yield from the US was about 60% higher than that from Canada. However, Canadian non-point source TP and DRP yields were 40% and 140% higher than the US, and the US point source yields were 9 times and 10 times higher than Canada for TP and DRP, respectively.

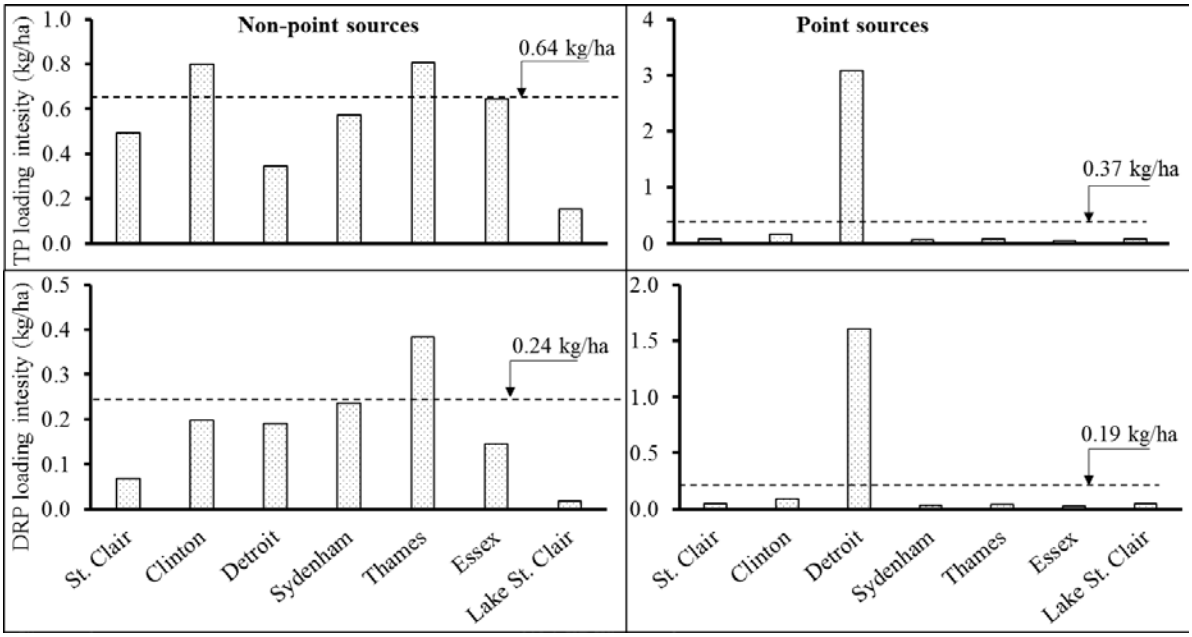

Figure 9: Average non-point (left) and point source (right) total phosphorus (TP) and dissolved reactive phosphorus (DRP) yields at the outlet of each sub-watershed (dashed horizontal line shows watershed average values).

These sub-watershed-specific yields of total, point, and non-point sources (Figure 9) can be useful for developing load reduction strategies. For example, while the overall TP yield from DT sub-watershed was about four times that of Thames; most of the yield from the DT sub-watershed was from point sources. Comparing non-point source yields, on the other hand, showed that the Thames sub-watershed yield was about twice that of the DT. Thus, in exploring management options at this scale, more attention should be placed on point sources in the DT sub-watershed and non-point source for agricultural areas of Thames sub-watershed.

Spatial distribution of non-point source yields – sub-basin and HRU scales. While evaluating yields at the sub-watershed scale was useful for higher-level strategies, assessments at sub-basin (24 km^2) and HRU (field) scales enabled the potential targeting of management practices. Average HRU-level TP yields were 1.38, 1.10, 0.78, 0.53, 0.96, and 0.63 kg/ha for UT, LT, SY, DT, CL and SC sub-watersheds respectively. Average DRP yields are 0.69, 0.50, 0.33, 0.36, 0.32, and 0.12 kg/ha, respectively. The median HRU-level yields for TP and DRP were lower than the average values (Figure 10). This indicated that regional average values were skewed by very high yielding areas across the watershed which in turn implied the presence of a good opportunity to focus management practices on certain areas to reduce the majority of nutrient loading from the watershed.

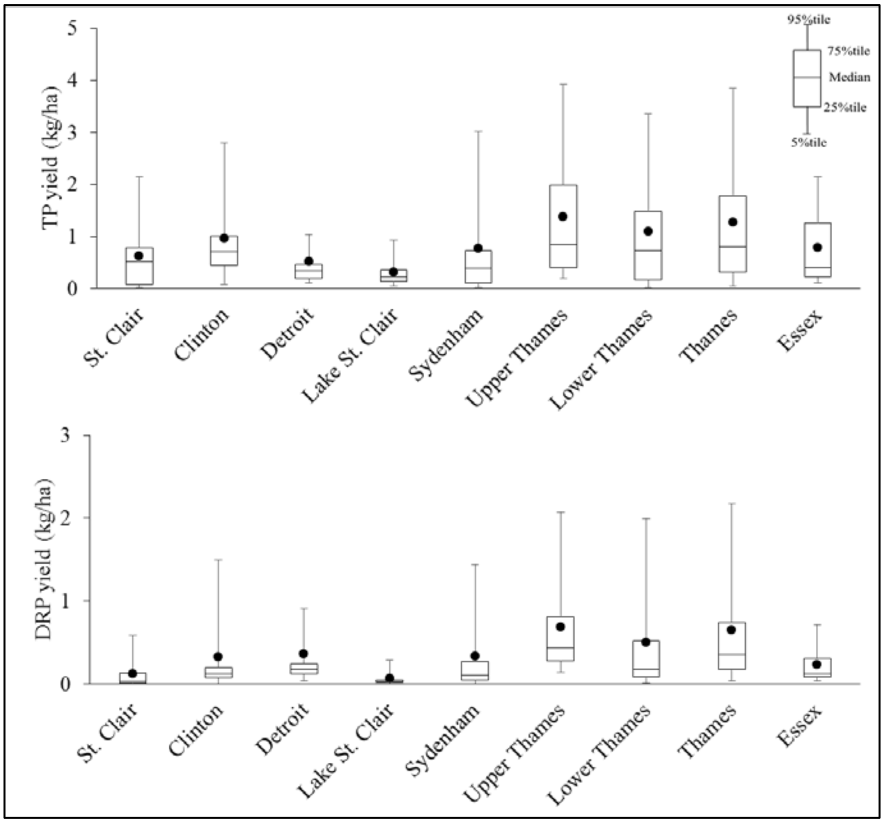

Figure 10: Distributions of HRU-level non-point source total phosphorus (TP) and dissolved reactive phosphorus (DRP) yields for each sub-watershed. ***Dots*** indicate average yield values.

Spatial patterns of non-point P yields at the HRU (field) and subbasin levels (Figure 11) provided further insight into potential areas of focus for non-point source reduction. High non-point source DRP yields spread relatively evenly across the Canadian watershed; whereas some of the highest TP yields were found in the upper parts of SY and Thames sub-watersheds. DRP yields from the US sub-watersheds were distinctly lower than the Canadian counterparts; however, certain non-agricultural areas in the US (lower parts of SC, upper parts of CL and some places in Detroit sub-watershed) appeared to have high yields as well. The higher DRP yields from Canadian sub-watersheds could be attributed to higher tile drainage density, higher proportion of cropland, and higher fertilizer application rates. For example, inorganic P application rates ranged from 22.8 to 44.8 kg/ha, 7.8 to 24.4 kg/ha, and 7.4 to 13.7 kg/ha for corn, winter wheat and soybeans, respectively, in Canada. These values were 5.9 to 10.9 kg/ha, 5.7 to 10.1 kg/ha, and 4.8 to 7.8 kg/ha in the US. Similarly, manure application rates were higher in Canadian agricultural areas (see "Input Characterization" section). The Canadian tile drainage system was also about twice as dense as in the US (see "Management data layers" section). As a result, Canadian portions of the watershed had higher sources of DRP (inorganic fertilizer or manure) and a system that facilitates its movement (denser drainage tile system).

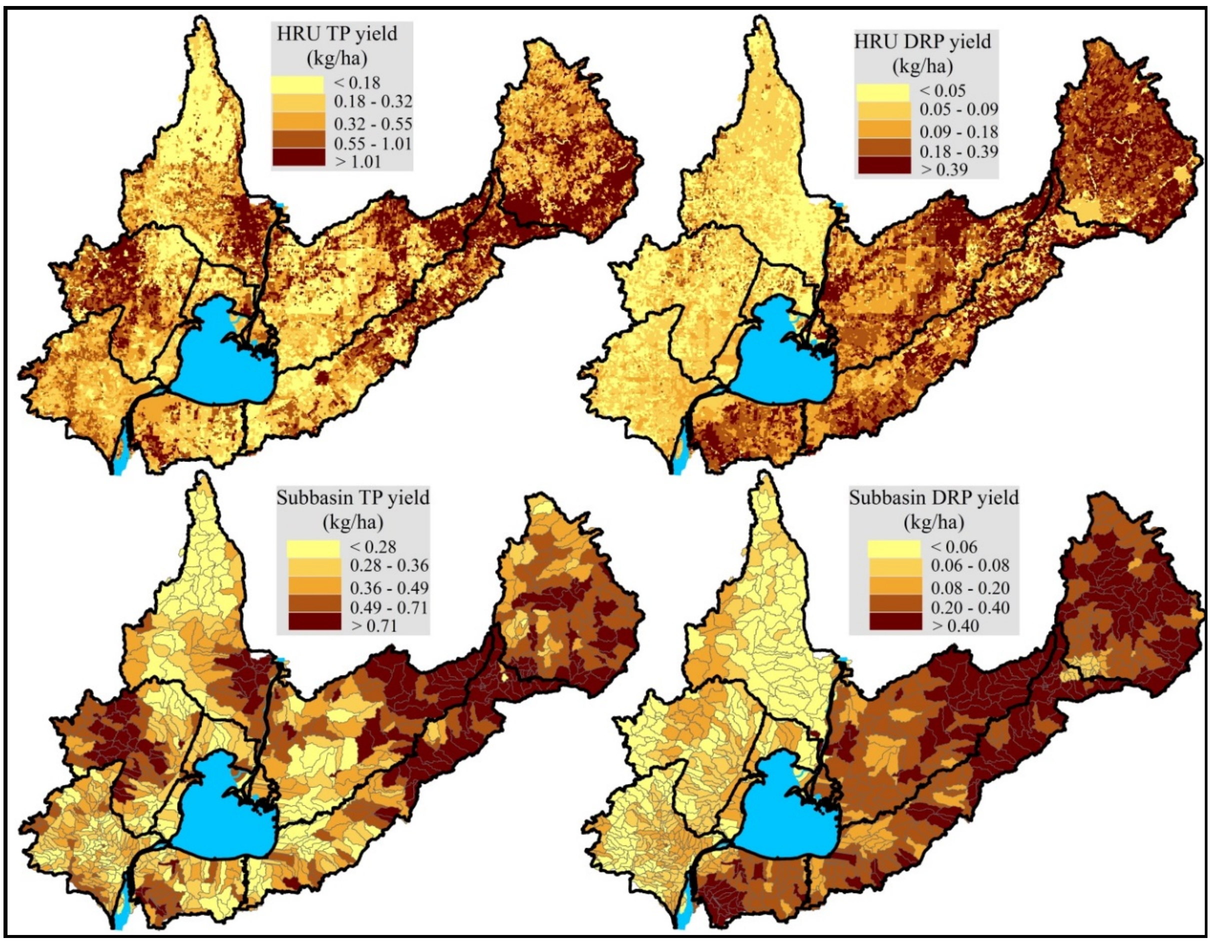

Figure 11: a) HRU-level (top) and subbasin-level (bottom) distributions of non-point source total phosphorus (TP, left) and dissolved reactive phosphorus (DRP, right) yields.

The distribution of P yields suggested that US agricultural areas had relatively low TP and DRP yields. For example, while the northern part of the CL sub-watershed was agricultural, the higher P yields from that sub-watershed were actually from non-agricultural areas in the central and west portions of the sub-watershed. Similarly, yields from the agricultural areas in the northern part of the SC sub-watershed were smaller than those from the non-agricultural areas. Most of the high phosphorus yielding areas in CL, for example, were urban areas located in a relatively higher slope region of the sub-watershed. Moreover, the major point source contribution of the watershed came from the DT sub-watershed (Table 6). These underscored the need to focus on Canadian agricultural runoff reduction strategies and both US point source management and urban runoff reduction strategies.

CONCLUSION

We integrated and harmonized US and Canadian datasets, including crop rotations, fertilizer/manure applications, tillage practices, and tile drainage systems; structured a SWAT model at finer resolution (field-scale) than ever done before for a 19,000 km^2 watershed; and calibrated and validated it at daily, monthly, and yearly time scales at six locations. While some input data (e.g., crop rotations) were constructed from a 30mx30m grid cell data, others (e.g., fertilizer application, tillage practice, manure generated, etc.) were available at county or provincial level. Hence, a great deal effort was invested in allocating model inputs from the lower spatial resolution to the field scale. Such distribution of model inputs not only improved model estimates at stream mouths but also provided more confidence in assessing flow and nutrient estimates at field level.

In most cases, a very good fit to flow measurements and good fit to water quality load estimates were achieved using manual and automatic calibration techniques at monthly time scales. It was evident from the calibration and validation processes that allowing some key parameters to vary across sub-watersheds improved model performance and that the variations were consistent with different sub-watershed characteristics.

Annual phosphorus loads increased between 2001 and 2009 and decreased afterwards, with the trend strongest in agricultural areas. Phosphorus yields were highest in Canadian agricultural areas and the US watershed was dominated by point sources, primarily from Great Lakes Water Authority treatment facility (Table 6 and Figure 8). Field-scale analysis used to identify areas within the Canadian agricultural and US urban landscapes with relatively high P yield from non-point sources point to where agricultural and urban management practices should be focused.

The main limitations of this study are the lack of some input data at the modeled scale and the relatively low number of water quality observations for calibration and validation. These limitations increased uncertainties in water quality calibration and validation results, and outputs at the field scale. More spatially explicit input data for nutrient inputs (fertilizer and manure application rates, soil nutrient content, etc.), agricultural practices (tillage, tile drainage, cover crop, filter strip in agricultural fields), and water quality observations would increase confidence of representations of nutrient and sediment estimates at both the field scale and stream mouths.

SUPPORTING INFORMATION

Additional supporting information may be found online under the Supporting Information tab for this article: Tables and Figures showing detail model input characterizations, parameter estimations and result evaluations.

ACKNOWLEDGEMENTS

This work was funded by the Fred A and Barbara M Erb Family Foundation grant number 903. We appreciate the insights and advice offered by Lynn Vaccaro and Jennifer Read from the University of Michigan, and Kevin McKague from Ontario Ministry of Agriculture, Food and Rural Affairs. We also gratefully acknowledge Debbie Burniston and Alice Dove from Environment and Climate Change Canada, Kevin McKague from Ontario Ministry of Agriculture, Food and Rural Affairs, Katie Stammler from the Essex Region Conservation Authority, Majid Khan and Catherine Willey from the Great Lakes Water Authority, Wanhong Yang from University of Guelph, Pamela Joose from Agriculture and Agri-Food Canada, and Karen Maaskant from Upper Thames River Conservation Authority for providing data.

LITERATURE CITED

AAFC (Agriculture and Agri-Food Canada), 2016. Annual Crop Inventory. URL https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9, (accessed 7.8.2016).

Abbaspour, K.C., 2015. SWATCalibration and Uncertainty Programs, A User manual, Eawag, Dubendorf, Switzerland.

Ahmadi, M., Arabi, M., Ascough, J.C., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environmental Modelling & Software 59, 135–145. https://doi.org/10.1016/j.envsoft.2014.05.012

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N., Jha, M. K., 2012. SWAT: Model Use, Calibration, and Validation. Transactions of the ASABE 55, 1491–1508. https://doi.org/10.13031/2013.42256

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large Area Hydrologic Modeling and Assessment Part I: Model Development1. JAWRA Journal of the American Water Resources Association 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Bai, J., Shen, Z., Yan, T., 2017. A comparison of single- and multi-site calibration and validation: a case study of SWAT in the Miyun Reservoir watershed, China. Frontiers of Earth Science 11, 592–600. https://doi.org/10.1007/s11707-017-0656-x

Baker, D.B., Confesor, R., Ewing, D.E., Johnson, L.T., Kramer, J.W., Merryfield, B.J., 2014. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: The importance of bioavailability. Journal of Great Lakes Research 40, 502–517. https://doi.org/10.1016/j.jglr.2014.05.001

Beale, E., 1962. Some uses of computers in operational research, Ind. Organ. 31(1), 27–28.

BIO Intelligence Service. 2014. Soil and water in a changing environment, Final Report prepared for European Commission (DG ENV), with support from HydroLogic.

Bocaniov, S.A., Scavia, D., 2018. Nutrient Loss Rates in Relation to Transport Time Scales in a Large Shallow Lake (Lake St. Clair, USA—Canada): Insights From a Three-Dimensional Model. Water Resources Research 54, 3825–3840. https://doi.org/10.1029/2017WR021876

Cao, W., Bowden, W.B., Davie, T., Fenemor, A., 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. Hydrological Processes 20, 1057–1073. https://doi.org/10.1002/hyp.5933

Chaibou Begou, J., Jomaa, S., Benabdallah, S., Bazie, P., Afouda, A., Rode, M., 2016. Multi-Site Validation of the SWAT Model on the Bani Catchment: Model Performance and Predictive Uncertainty. Water 8, 178. https://doi.org/10.3390/w8050178

Daniel, E.B., 2011. Watershed Modeling and its Applications: A State-of-the-Art Review. The Open Hydrology Journal 5, 26–50. https://doi.org/10.2174/1874378101105010026

Daniel, E.B., Camp, J.V., LeBoeuf, E.J., Penrod, J.R., Abkowitz, M.D., Dobbins, J.P., 2010. Journal of Spatial Hydrology Vol.10, No.2, Fall 2010 16.

Dolan, D.M., Yui, A.K., Geist, R.D., 1981. Evaluation of River Load Estimation Methods for Total Phosphorus. Journal of Great Lakes Research 7, 207–214. https://doi.org/10.1016/S0380-1330(81)72047-1

Gassman, P. W., Reyes, M. R., Green, C. H., Arnold, J. G., 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. Transactions of the ASABE 50, 1211–1250. https://doi.org/10.13031/2013.23637

Gassman, P. W., Balmer, C., Siemers, M., and Srinivasan, R., 2014. The SWAT Literature Database: Overview of database structure and key SWAT literature trends. Proceedings of the 2014 International SWAT Conference, July 28–1 August, Pernambuco, Brazil, Texas Water Resources Institute Technical Report – TR472.

GLWQA (Great Lakes Water Quality Agreement), 2016. The United States and Canada adopt phosphorus load reduction targets to combat Lake Erie algal blooms. URL https://binational.net/2016/02/22/finalptargets-ciblesfinalesdep/, (accessed 6.22.2016).

Hasan, M.A., Pradhanang, S.M., 2017. Estimation of flow regime for a spatially varied Himalayan watershed using improved multi-site calibration of the Soil and Water Assessment Tool (SWAT) model. Environmental Earth Sciences 76. https://doi.org/10.1007/s12665-017-7134-3

Hirsch, R.M., 1979. An evaluation of some record reconstruction techniques. Water Resources Research 15, 1781–1790. https://doi.org/10.1029/WR015i006p01781

Hirsch, R.M., Moyer, D.L., Archfield, S.A., 2010. Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs. J Am Water Resour Assoc 46, 857–880. https://doi.org/10.1111/j.1752-1688.2010.00482.x

HYDAT, 2016. Hydrometric Data, National Water Data Archive, Water Survey of Canada. URL https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html, (accessed 11.8.2016).

IJC (International Join Commission), 2015. The International Watershed Initiative. From Concept to Cornerstone of the International Joint Commission. A Watershed approach for Coordinated Stewardship of Shared Canada-U.S. Waters. URL http://ijc.org/files/tinymce/uploaded/IWI/IJC-IWI-EN-WEB.pdf.

IJC (International Join Commission), 2009. The International Watershed Initiative. Implementing a New Paradigm for Transboundary Basins. URL http://www.ijc.org/files/publications/ID1627.pdf.

IPNI (International Pant Nutrition Institute), 2016. Nutrient Use Geographic Information System. URL http://nugis.ipni.net/About%20NuGIS/, (accessed 8.30.2016).

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21, 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Johnston, R., Smakhtin, V., 2014. Hydrological Modeling of Large river Basins: How Much is Enough? Water Resources Management 28, 2695–2730. https://doi.org/10.1007/s11269-014-0637-8

Kalcic, M.M., Chaubey, I., Frankenberger, J., 2015. Defining Soil and Water Assessment Tool (SWAT) hydrologic response units (HRUs) by field boundaries. Biol Eng 8, 12.

Kalcic, M.M., Kirchhoff, C., Bosch, N., Muenich, R.L., Murray, M., Griffith Gardner, J., Scavia, D., 2016. Engaging Stakeholders To Define Feasible and Desirable Agricultural Conservation in Western Lake Erie Watersheds. Environmental Science & Technology 50, 8135–8145. https://doi.org/10.1021/acs.est.6b01420

Leta, O.T., van Griensven, A., Bauwens, W., 2017. Effect of Single and Multisite Calibration Techniques on the Parameter Estimation, Performance, and Output of a SWAT Model of a

Spatially Heterogeneous Catchment. Journal of Hydrologic Engineering 22, 05016036. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001471

Li, X., Weller, D.E., Jordan, T.E., 2010. Watershed model calibration using multi-objective optimization and multi-site averaging. Journal of Hydrology 380, 277–288. https://doi.org/10.1016/j.jhydrol.2009.11.003

Li, Z., Shao, Q., Xu, Z., Cai, X., 2010. Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. Journal of Hydrology 385, 76–83. https://doi.org/10.1016/j.jhydrol.2010.01.025

Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie: A detailed accounting by year, basin, country, and tributary. Journal of Great Lakes Research 42, 1151–1165. https://doi.org/10.1016/j.jglr.2016.08.005

Madani, K., Mariño, M.A., 2009. System Dynamics Analysis for Managing Iran's Zayandeh-Rud River Basin. Water Resour Manage 23, 2163–2187. https://doi.org/10.1007/s11269-008-9376-z

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources 26, 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1

Mckinney, D.C., Cai, X., Lasdon, L.S., Agency, U.S., Development, I., Mckinney, D.C., Cai, X., Lasdon, L.S., 1999. INTEGRATED WATER RESOURCES MANAGEMENT MODEL FOR THE SYR DARYA BASIN Prepared by:

Medema, W., Furber, A., Adamowski, J., Zhou, Q., Mayer, I., 2016. Exploring the Potential Impact of Serious Games on Social Learning and Stakeholder Collaborations for Transboundary Watershed Management of the St. Lawrence River Basin. Water 8, 175. https://doi.org/10.3390/w8050175

Mekonnen, B.A., Mazurek, K.A., Putz, G., 2017. Modeling of nutrient export and effects of management practices in a cold-climate prairie watershed: Assiniboine River watershed, Canada. Agricultural Water Management 180, 235–251. https://doi.org/10.1016/j.agwat.2016.06.023

Mirchi, A., Watkins, D., Madani, K., 2009. Modeling for Watershed Planning, Management, and Decision Making. Watersheds: management, restoration and environmental Impact, Chapter 6. ISBN: 978-1-61668-667-3.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE 50, 885–900. https://doi.org/10.13031/2013.23153

NWQMC-WQP (National Water Quality Monitoring Council – Water Quality Portal), 2016. URL https://www.waterqualitydata.us/portal/(accessed 11.8.2016).

Niraula, R., Norman, L.M., Meixner, T., Callegary, J.B., 2012. Multi-gauge Calibration for Modeling the Semi-Arid Santa Cruz Watershed in Arizona-Mexico Border Area Using SWAT. Air, Soil and Water Research 5, ASWR.S9410. https://doi.org/10.4137/ASWR.S9410

NOAA-GHCN (National Oceanic and Atmospheric Administration – Global Historical Climatology Network), 2016. URL ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/,(accessed 7.28.2016).

OMAFRA (Ontario Ministry of Agriculture, Food and Rural Affairs), 2016. URL http://www.omafra.gov.on.ca/english/stats/welcome.html, (accessed 10.25.2016).

PWQMN (Provincial (Stream) Water Quality Monitoring Network), 2016. Ontario. URL https://www.ontario.ca/data/provincial-stream-water-quality-monitoring-network, (accessed 11.8.2016).

Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. Journal of Hydrology 198, 69–97. https://doi.org/10.1016/S0022-1694(96)03329-X

Ryberg, K.R., Vecchia, A.V., n.d. Vignette for waterData—An R Package for Retrieval, Analysis, and Anomaly Calculation of Daily Hydrologic Time Series Data 19.

Scavia, D., Baconiov, S., Dagnew, A., Long, C., and Wang, Y., 2019. Interaction of climate change and monitoring protocols can influence approaches to nutrient load management: The St. Clair – Detroit River System and Lake Erie. Journal of Great Lakes Research 45, 40–49. https://doi.org/10.1016/j.jglr.2018.11.008.

Scavia, D., David Allan, J., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, T.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Peter Richards, R., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T.M., Zhang, H., Zhou, Y., 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. Journal of Great Lakes Research 40, 226–246. https://doi.org/10.1016/j.jglr.2014.02.004

Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. Journal of Great Lakes Research 42, 1139–1150. https://doi.org/10.1016/j.jglr.2016.09.007

Serrat-Capdevila, A., Valdés, J.B., Pérez, J.G., Baird, K., Mata, L.J., Maddock, T., 2007. Modeling climate change impacts – and uncertainty – on the hydrology of a riparian system: The San Pedro Basin (Arizona/Sonora). Journal of Hydrology 347, 48–66. https://doi.org/10.1016/j.jhydrol.2007.08.028

Sheelanere, P., Noble, B.F., Patrick, R.J., 2013. Institutional requirements for watershed cumulative effects assessment and management: Lessons from a Canadian trans-boundary watershed. Land Use Policy 30, 67–75. https://doi.org/10.1016/j.landusepol.2012.03.001

Shi, P., Ma, X., Hou, Y., Li, Q., Zhang, Z., Qu, S., Chen, C., Cai, T., Fang, X., 2013. Effects of Land-Use and Climate Change on Hydrological Processes in the Upstream of Huai River, China. Water Resources Management 27, 1263–1278. https://doi.org/10.1007/s11269-012-0237-4

Shrestha, M.K., Recknagel, F., Frizenschaf, J., Meyer, W., 2016. Assessing SWAT models based on single and multi-site calibration for the simulation of flow and nutrient loads in the semi-arid Onkaparinga catchment in South Australia. Agricultural Water Management 175, 61–71. https://doi.org/10.1016/j.agwat.2016.02.009

Singh, V.P., Frevert, D.K., 2006. Watershed models. Taylor & Francis, Boca Raton. CRC Press.

STATCAN (Statistics Canada), 2016. Canadian Fertilizer Shipments Survey. URL https://www150.statcan.gc.ca/n1/en/type/data?text=001-0066..001-0069,(accessed 10.24.2016).

Teshager, A.D., Gassman, P.W., Secchi, S., Schoof, J.T., Misgna, G., 2016. Modeling Agricultural Watersheds with the Soil and Water Assessment Tool (SWAT): Calibration and Validation with a Novel Procedure for Spatially Explicit HRUs. Environmental Management 57, 894–911. https://doi.org/10.1007/s00267-015-0636-4

USDA-ERS (U.S. Department of Agriculture – Economic Research Service) URL https://data.ers.usda.gov/reports.aspx?ID=17882 (accessed 8.8.2016).

USDA-NASS (U.S. Department of Agriculture – National agricultural Statistics Services), 2016. CropScape - Cropland Data Layer. URL https://nassgeodata.gmu.edu/CropScape/ (6.3.2016).

USDA-NRCS (U.S. Department of Agriculture – Natural Resources Conservation Service), 2017. Web Soil Survey. URL https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx (accessed 5.10.2017).

USGS (U.S. Geological Survey), 2016. The National Map. URL https://viewer.nationalmap.gov/basic/#startUp (accessed 6.9.2016).

USGS-NWIS (U.S. Geological Survey – National Water Information System), 2016. Water Data for the Nation. URL https://waterdata.usgs.gov/nwis, (accessed 11.8.2016).

Vaughn, J.C. (Ed.), 2010. Watersheds: management, restoration, and environmental impact. Nova Science Publishers, New York.

UTCA (Upper Thames River Conservation Authority), 2017. Surface Water and Groundwater Studies. URL http://thamesriver.on.ca/watershed-health/surfacewater-groundwater-studies/, (accessed 2.20.2017).

Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y., Yao, A., 2012. Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China. Hydrology and Earth System Sciences 16, 4621–4632. https://doi.org/10.5194/hess-16-4621-2012

White, K.L., Chaubey, I., n.d. SENSITIVITY ANALYSIS, CALIBRATION, AND VALIDATIONS FOR A MULTISITE AND MULTIVARIABLE SWAT MODEL. JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION 13.

Yang, J.-L., Zhang, G.-L., 2015. Formation, characteristics and eco-environmental implications of urban soils – A review. Soil Science and Plant Nutrition 61, 30–46. https://doi.org/10.1080/00380768.2015.1035622

Zhang, X., Srinivasan, R., Van Liew, M., 2008. Multi-Site Calibration of the SWAT Model for Hydrologic Modeling. Transactions of the ASABE 51, 2039–2049. https://doi.org/10.13031/2013.25407

Zhang, X., Srinivasan, R., Liew, M.V., 2010. On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. Hydrological Processes 24, 955–969. https://doi.org/10.1002/hyp.7528

Supplementary Information

Table S1: Estimated nitrogen and phosphorus mineral fertilizer application rates in the watershed (C=Corn, S=Soybeans, W=Winter wheat)

Fertilizer Type	Country	County	*Mineral fertilize application rates for crops in each rotation per county (kg/ha)*																		
			Corn						**Soybeans**								**Winter Wheat**				
			CC	**CS**	**SC**	**CSW**	**SWC**	**WCS**	**CS**	**SC**	**SS**	**CSW**	**SSW**	**SW**	**SWC**	**WCS**	**CSW**	**SSW**	**SW**	**SWC**	**WCS**
Mineral nitrogen application rate (kg/ha)	**USA**	Macomb	107.8	88.2	90.5	90.3	95.0	92.1	9.8	10.1	10.0	10.0	9.5	9.8	10.6	10.2	85.3	80.8	83.1	89.7	87.0
		St. Clair	95.5	84.0	88.2	90.8	85.4	86.1	9.9	10.4	10.9	10.7	9.6	10.6	10.0	10.1	69.4	62.2	68.7	65.3	65.8
		Lapeer	112.7	92.8	92.8	92.2	92.3	92.5	14.6	14.6	14.6	14.6	−	14.7	14.6	14.6	68.0	−	68.6	68.0	68.2
		Oakland	98.0	83.3	83.3	83.3	82.5	82.9	9.8	9.8		9.8	−	−	9.7	9.8	63.7	−	−	63.1	63.4
		Sanilac	125.9	106.3	107.4	108.5	109.2	107.7	15.2	15.3	15.3	15.5	15.6	16.4	15.6	15.4	72.3	72.8	76.3	72.8	71.8
		Washtenaw	−	122.5	121.8	122.5	122.5	122.5	29.4	29.2	29.3	29.4	−	−	29.4	29.4	93.1	−	−	93.1	93.1
		Wayne	−	−	121.3	122.5	118.8	122.5		24.3	24.5	24.5	−	−	23.8	24.5	93.1	−	−	90.3	93.1
	Canada	Kent	156.5	138.3	134.4	134.0	132.0	133.9	4.1	4.0	3.9	4.0	3.9	4.0	3.9	4.0	84.4	82.3	84.5	83.1	84.3
		Elgin	173.8	146.7	145.8	142.5	148.4	147.2	4.5	4.5	4.9	4.4	3.8	4.9	4.6	4.5	87.7	76.0	97.6	91.3	90.6
		Essex	154.3	128.3	130.9	128.4	132.8	131.5	3.8	3.9	3.8	3.8	4.0	3.9	3.9	3.9	80.9	85.0	82.9	83.6	82.8
		Huron	141.1	127.4	127.4	127.4	127.4	127.4	3.6	3.6	3.6	3.6	−	−	3.6	3.6	77.4	−	−	77.4	77.4
		Lambton	158.9	134.3	143.1	155.8	154.1	150.9	3.7	3.9	4.1	4.3	4.0	4.2	4.3	4.2	91.3	85.8	88.5	90.3	88.5
		Middlesex	168.1	158.0	148.6	147.4	144.6	146.9	4.5	4.2	4.1	4.2	3.8	4.4	4.1	4.2	89.5	80.8	93.5	87.8	89.2
		Oxford	151.5	132.7	134.7	135.8	133.5	136.3	3.8	3.8	3.8	3.9	−	3.7	3.8	3.9	82.5	−	79.1	81.0	82.7
		Perth	140.5	128.1	127.8	128.1	128.1	128.1	3.7	3.7	3.7	3.7	3.7	3.6	3.7	3.7	82.3	82.8	81.9	82.4	82.4
Mineral phosphorus application rate (kg/ha)	**USA**	Macomb	7.8	6.9	7.0	7.0	7.4	7.2	5.9	6.0	6.0	6.0	5.7	5.9	6.3	6.1	7.0	6.7	6.8	7.4	7.2
		St. Clair	7.6	5.9	6.2	6.4	6.0	6.1	4.9	5.2	5.5	5.3	4.8	5.3	5.0	5.1	6.4	5.7	6.3	6.0	6.1
		Lapeer	7.8	6.8	6.8	6.8	6.8	6.8	5.9	5.9	5.9	5.8	−	5.9	5.8	5.8	7.8	−	7.8	7.8	7.8
		Oakland	8.8	7.8	7.8	7.8	7.8	7.8	5.9	5.9		5.9	−	−	5.8	5.9	7.8	−	−	7.8	7.8
		Sanilac	10.1	8.1	8.2	8.3	8.3	8.2	6.1	6.1	6.1	6.2	6.2	6.5	6.2	6.2	10.3	10.4	10.9	10.4	10.3
		Washtenaw	−	9.8	9.7	9.8	9.8	9.8	7.8	7.8	7.8	7.8	−	−	7.8	7.8	9.8	−	−	9.8	9.8
		Wayne	−	−	7.8	7.8	7.6	7.8		6.8	6.9	6.9	−	−	6.7	6.9	7.8	−	−	7.6	7.8
	Canada	Kent	36.5	25.6	24.9	24.8	24.4	24.8	10.2	10.0	9.9	9.9	9.7	9.9	9.8	9.9	19.9	19.4	19.9	19.6	19.8
		Elgin	42.0	28.2	28.0	27.4	28.5	28.3	11.3	11.2	12.2	11.0	9.5	12.2	11.4	11.3	21.9	19.0	24.4	22.8	22.6
		Essex	33.9	23.8	24.2	23.8	24.6	24.4	7.6	7.8	7.6	7.6	8.0	7.8	7.9	7.8	17.1	18.0	7.8	17.7	17.5
		Huron	31.9	22.8	22.8	22.8	22.8	22.8	13.7	13.7	13.7	13.7	−	−	13.7	13.7	20.0	−	−	20.0	20.0
		Lambton	34.8	23.2	24.7	26.9	26.6	26.0	7.4	7.9	8.1	8.6	8.1	8.3	8.5	8.3	21.5	20.2	20.8	21.3	20.8
		Middlesex	44.8	33.9	31.9	31.6	31.0	31.5	11.3	10.6	10.3	10.5	9.5	11.0	10.3	10.5	21.1	19.0	22.0	20.7	21.0
		Oxford	37.9	28.4	28.9	29.1	28.6	29.2	9.5	9.6	9.6	9.7	−	9.3	9.5	9.7	24.3	−	23.3	23.8	24.3
		Perth	31.7	24.7	24.6	24.7	24.7	24.7	9.1	9.1	9.2	9.1	9.2	9.1	9.2	9.2	18.3	18.4	18.2	18.3	18.3

Table S2. Tributary name, drainage area of calibration locations, flow and water quality gauging stations, and number of point source (PS) facilities and combined sewer overflow (CSO) outfalls.

River Name	**Drainage Area at calibration location (km^2)**	**As % of entire watershed***	**Number of PS Facilities / CSO Outfalls**	**Flow gaging Station**	**Water quality station**
Black	1843	10.3	18 / 9	USGS-04159492, USGS-04159900	USEPA-740267 and USGS-04160055 (Same location), USGS-04160075
Clinton	1916	10.7	8 / 2	USGS-04165500	USEPA-500233 and USGS-04165553 (Same location), USGS-04165500
Sydenham	2959	16.6	11 / -	HYDAT-02GG009, HYDAT-02GG013, HYDAT-02GG003	ECCC-ON02GG1000, PWQMN-04002701602, PWQMN-04002701702
Thames	4989	27.9	29 / -	HYDAT-02GE003, HYDAT-02GE007	ECCC-ON02GE1000, PWQMN-04001305802, PWQMN-04001308202, PWQMN-04001300782
Rouge	1230	6.9	4 / 79	USGS-04166500, USGS-04167000, USGS-04167150, USGS-04168000, USGS-04168400	EPA-820070 and USGS-04168550 (Same location)

*This excludes the area of Lake St. Clair

Table S3: Flow parameters altered during calibration. CNII=Runoff curve number at moisture condition II, OV_N=Manning's "n" value for overland flow, HRU_SLP=Average slope of HRU, SLSUBBSN=Average slope length, SOL_AWC=Available water capacity of soil, ESCO=Soil evaporation compensation factor, GW_DELAY=Groundwater delay time, GWQMN=Threshold depth of water in aquifer for return flow to occur, GW_REVAP=Groundwater evap coefficient, Alpha_Bf=Baseflow alpha factor, CH_W2=Average width of main channel at top of bank, CH_S2=Average slope of main channel along the channel length, CH_D=Depth of main channel from top of bank to bottom, SFTMP=Snowfall temperature, SMTMP=Snow melt base temperature, SMFMX=Melt factor of snow on June 21, SMFMN=melt factor of snow on December 21, TIMP=Snow pack temperature lag factor, SNOCOVMX=Minimum snow water content that corresponds to 100% snow cover, SNO50COV=Fraction of snow volume represented by SNOCOVMX that corresponds to 50% snow cover, SURLAG=Surface runoff lag coefficient (relative = default value is multiplied by 1+ fitted values, replace= default value is replaced by fitted values)

Parameter	Unit	Default value	Change type	Scale	Fitted values					
					Upper Thames	Black	Sydenham	Clinton	Lower Thames	Rouge
CNII	-	varies (35 - 94)	relative	hru	0.100	-0.050	0.110	0.085	0.040	-0.010
OV_N	-	varies (0.03 - 0.90)	relative	hru	3.00	3.00	1.55	1.50	3.00	2.70
HRU_SLP	m/m	varies (0-0.327)	relative	hru	-0.70	-0.50	-0.47	-0.60	-0.70	-0.60
SLSUBBSN	m	varies (9.15 - 121.95)	relative	hru	0.70	-0.40	0.48	0.60	0.70	0.20
SOL_AWC	mm H2O/mm soil	varies (0.214-586.6)	relative	hru	0.0	0.1	-0.3	-0.5	0.1	-0.2
ESCO	-	0.95	replace	hru	0.30	0.90	0.51	0.80	0.90	0.50
GW_DELAY	days	31	replace	hru	31	31	25	16	31	30
GWQMN	mm H2O	1000	replace	hru	50	400	100	0	20	350
GW_REVAP	-	0.02	replace	hru	0.05	0.10	0.04	0.02	0.02	0.10
ALPHA_BF	1/days	0.048	replace	hru	0.10	0.50	0.06	0.60	0.10	0.20
CH_W2	m	varies (0.04-476.9)	relative	subbasin	0.0	0.0	0.0	0.0	0.0	0.5
CH_S2	m/m	varies (0.00-0.083)	relative	subbasin	0.0	0.0	0.0	-0.6	0.0	-0.7
CH_D	m	varies (0.01 - 6.7)	relative	subbasin	0.0	0.0	0.0	1.2	0.0	-0.2
SFTMP	oC	1	replace	watershed	-0.1					
SMTMP	oC	0.5	replace	watershed	-0.1					
SMFMX	mmH2O/oC-day	4.5	replace	watershed	4.9					
SMFMN	mmH2O/oC-day	4.5	replace	watershed	3.7					
TIMP	-	1	replace	watershed	0.28					
SNOCOVMX	mm H2O	1	replace	watershed	50.7					
SNO50COV	-	0.5	replace	watershed	0.41					
SURLAG	-	4	replace	watershed	0.27					

Table S4: Water quality parameters altered during calibration, ANION_EXCL=Fraction of porosity from which anions are excluded, SOL_NO3= Initial nitrate concentration in the soil layer, SOL_ORGN= Initial organic nitrogen concentration in the soil layer, SHALLST_N= Initial concentration of nitrate in shallow aquifer, LAT_ORGN= Organic nitrogen in base flow, ERORGN= Organic nitrogen enrichment ratio for loading with sediment, SOL_SOLP = Initial soluble phosphorus concentration in the soil layer, SOL_ORGP=Initial organic phosphorus concentration in the soil layer, GWSOLP=Concentration of soluble phosphorus in groundwater contribution of streamflow, LAT_ORGP=Organic phosphorus in base flow, ERORGP=Phosphorus enrichment ratio for loading with sediment, BC4=Rate constant for decay of organic phosphorus to dissolved, NPERCO=Nitrate percolation coefficient, PHOSKD=Phosphorus soil partitioning coefficient, P_UPDIS=Phosphorus uptake distribution parameter, N_UPDIS=Nitrogen uptake distribution parameter, CDN=Denitrification exponential rate coefficient, PSP=Phosphorus availability index, BIOMIX=Biological mixing coefficient (relative = default value is multiplied by 1+ fitted values, replace= default value is replaced by fitted values).

Parameter	Unit	Default value	Change type	Scale	Fitted values					
					Upper Thames	Black	Sydenham	Clinton	Lower Thames	Rouge
ANION_EXCL		0.5	replace	hru	0.95	0.05	0.15	0.53	0.20	0.06
SOL_NO3	mg N/kg soil	0	replace	hru	100	0	100	25	40	0
SOL_ORGN	mg N/kg soil	0	replace	hru	5	10	10	5	1	1
SHALLST_N	mg N/L	0	replace	hru	200	0	170	120	100	0
LAT_ORGN	mg/L	0	replace	hru	10	10	15	5	1	1
ERORGN		0	replace	hru	0.90	1.30	0.80	0.92	0.25	0.45
SOL_SOLP	mg P/kg soil	5	replace	hru	0.67	0.05	0.10	0.10	0.20	0.15
SOL_ORGP	mg P/kg soil	0	replace	hru	0.25	0.35	0.30	0.42	0.25	0.30
GWSOLP	mg P/L	0	replace	hru	0.67	0.05	0.10	0.05	0.15	0.10
LAT_ORGP	mg/L	0	replace	hru	0.25	0.35	0.30	0.45	0.20	0.30
ERORGP		0	replace	hru	0.50	0.65	0.55	1.00	0.35	0.40
BC4	1/day	0.35	relative	subbasin	-0.30	-0.95	-0.20	-0.95	-0.55	-0.95
NPERCO		0.2	replace	watershed	0.25					
PHOSKD	m3/Mg	175	replace	watershed	200					
P_UPDIS		20	replace	watershed	100					
N_UPDIS		20	replace	watershed	40					
CDN		0.5	replace	watershed	1.4					
PSP		0.4	replace	watershed	0.01					
BIOMIX		0.2	replace	hru	0.01					

Table S5: Daily and annual flow calibration and validation performance statistics (R^2 = coefficient of determination, NSe = Nash-Sutcliffe efficiency, PBs = percent bias)

Time step	Statistics	*Performance values for calibration(validation) period*					
		Upper Thames	**Black**	**Sydenham**	**Clinton**	**Lower Thames**	**Rouge**
Daily	R^2	0.69(0.80)	0.51(0.53)	0.69(0.65)	0.63(0.80)	0.87(0.92)	0.65(0.64)
	NSe	0.68(0.80)	0.43(0.52)	0.66(0.61)	0.53(0.75)	0.87(0.91)	0.64(0.64)
	PBs	0.1(3.2)	9.4(-2.7)	-1.2(8.7)	-2.7(1.9)	-2.7(5.4)	-1.2(-8.5)
Annual	R^2	0.91(0.97)	0.88(0.78)	0.88(0.89)	0.59(0.92)	0.92(0.94)	0.73(0.94)
	NSe	0.91(0.93)	0.81(0.69)	0.88(0.76)	0.58(0.70)	0.91(0.85)	0.68(0.67)
	PBs	0.1(3.2)	9.4(-2.7)	-1.2(8.7)	-2.8(1.8)	-2.4(5.6)	-1.2(-8.5)

Table S6: Monthly flow statistics for each sub-watershed if uniform alteration of parameters across the watershed were used (red=significant, orange=moderate, and green=slight or no changes. grey=final calibration/validation statistics, R^2 = coefficient of determination, NSe = Nash-Sutcliffe efficiency, PBs = percent bias)

		Statistics	**Flow calibration/validation statistics values**					
			Upper Thames	**Black**	**Sydenham**	**Clinton**	**Lower Thames**	**Rouge**
Flow parameter sets used from	Upper Thames	R^2	0.84(0.93)	0.69(0.71)	0.84(0.85)	0.52(0.59)	0.86(0.92)	0.67(0.74)
		NSe	0.84(0.93)	0.64(0.54)	0.82(0.83)	0.20(0.29)	0.86(0.91)	0.66(0.67)
		PBs	0.1(3.2)	-20(-35.4)	-13.9(-6.9)	-37.0(-34.5)	-5.4(2.3)	-6.1(-14.1)
	Black	R^2	0.73(0.62)	0.72(0.76)	0.81(0.87)	0.57(0.64)	0.86(0.92)	0.64(0.72)
		NSe	0.72(0.61)	0.72(0.76)	0.81(0.87)	0.47(0.55)	0.86(0.91)	0.60(0.72)
		PBs	9.5(13.1)	9.2(-2.9)	5.2(10.2)	-18.7(-18.8)	-7.1(-1.4)	5.7(-4.1)
	Sydenham	R^2	0.80(0.86)	0.68(0.71)	0.85(0.87)	0.54(0.66	0.85(0.89)	0.68(0.76)
		NSe	0.78(0.81)	0.68(0.69)	0.85(0.86)	0.41(0.57)	0.85(0.87)	0.62(0.75)
		PBs	7.6(12.7)	1.5(-10.3)	-1.2(8.4)	-23.0(-19.2)	3.2(12.7)	7.7(0.2)
	Clinton	R2	0.84(0.89)	0.69(0.74)	0.82(0.83)	0.63(0.80)	0.85(0.88)	0.77(0.85)
		NSe	0.81(0.85)	0.62(0.70)	0.80(0.79)	0.53(0.75)	0.84(0.81)	0.46(0.65)
		PBs	12.6(16.5)	28.8(20.5)	10.7(21.5)	-2.7(1.9)	10.9(21.4)	24.3(18.0)
	Lower Thames	R^2	0.82(0.92)	0.71(0.76)	0.83(0.89)	0.52(0.59)	0.87(0.92)	0.60(0.69)
		NSe	0.81(0.91)	0.70(0.76)	0.83(0.88)	0.42(0.50)	0.87(0.91)	0.55(0.68)
		PBs	7.5(8.0)	12.9(-0.6)	-3.3(1.7)	-18.1(-18.7)	-2.7(5.4)	6.7(-3.3)
	Rouge	R^2	0.85(0.92)	0.74(0.75)	0.84(0.86)	0.59(0.65)	0.88(0.92)	0.71(0.78)
		NSe	0.84(0.92)	0.71(0.62)	0.82(0.85)	0.35(0.44)	0.86(0.92)	0.70(0.75)
		PBs	-11.2(-5.2)	-16.8(-30.4)	-15.1(-5.5)	-32.4(-29.1)	-9.7(-0.9)	-1.1(-8.5)

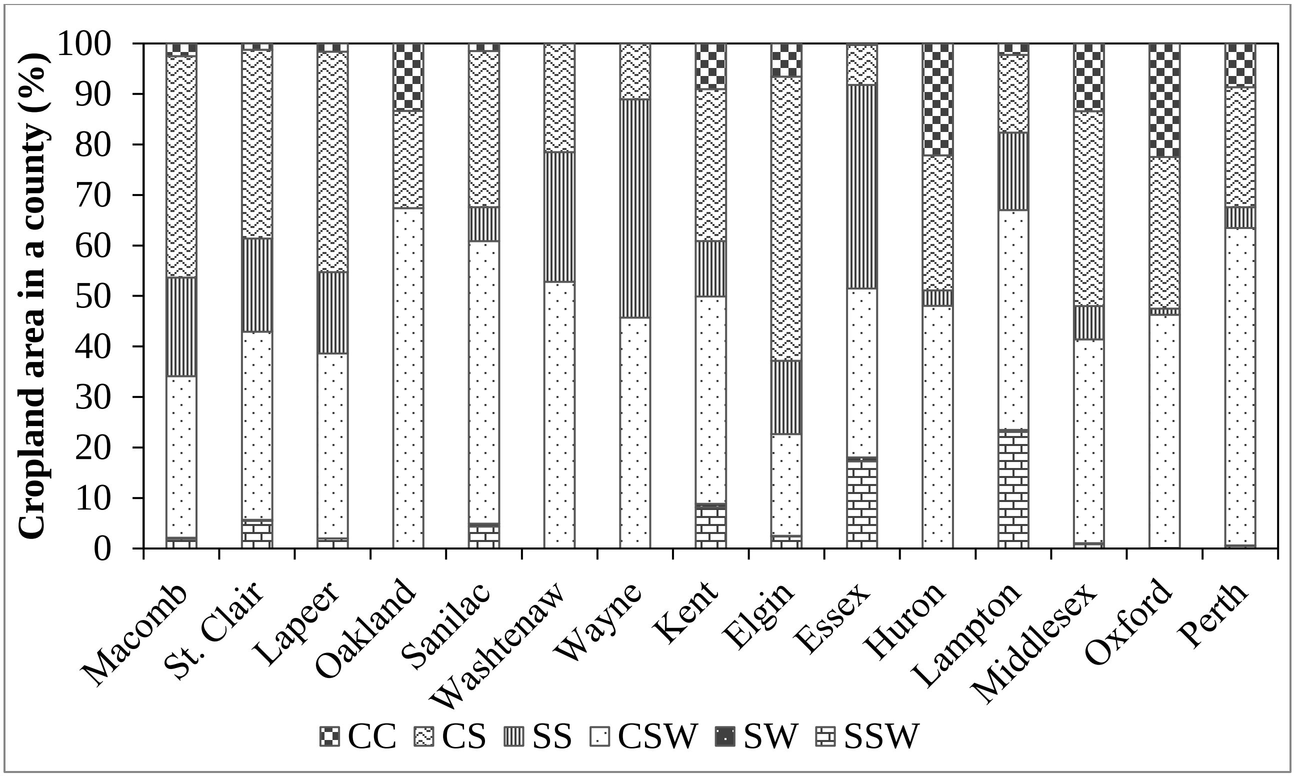

Figure S1: Crop rotation distribution in each county with in the watershed (C = Corn, S=Soybeans, W=Winter wheat)

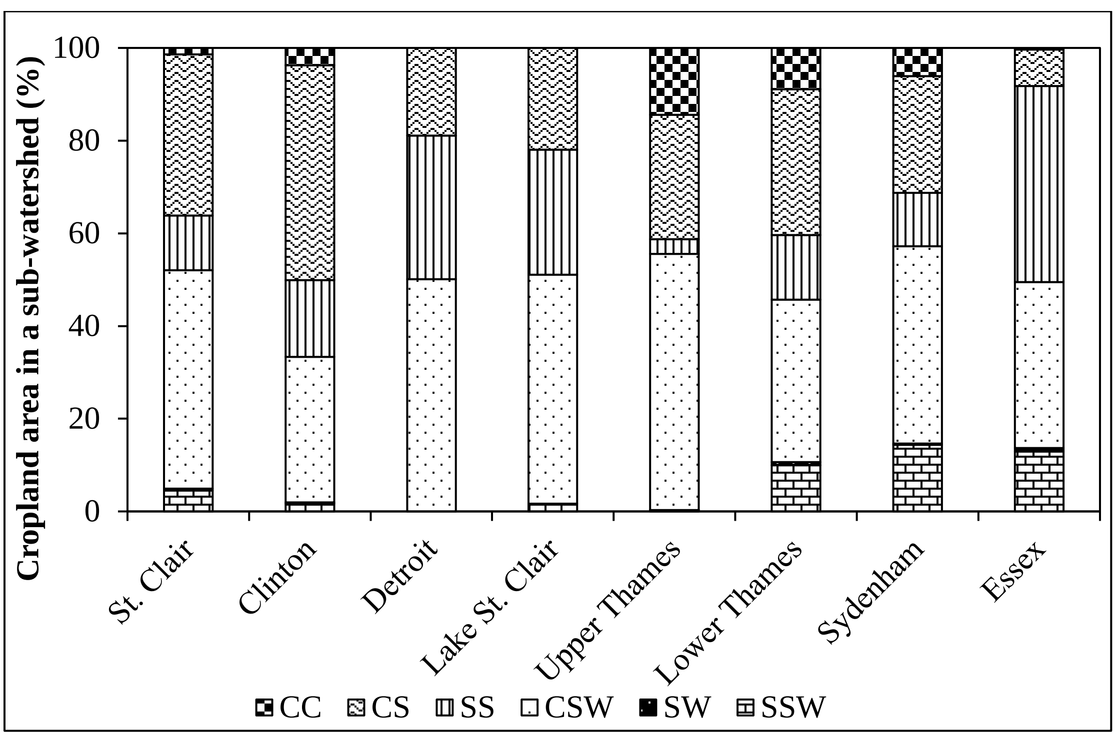

Figure S2: Crop rotation distribution in each sub-watershed (C = Corn, S=Soybeans, W=Winter wheat)

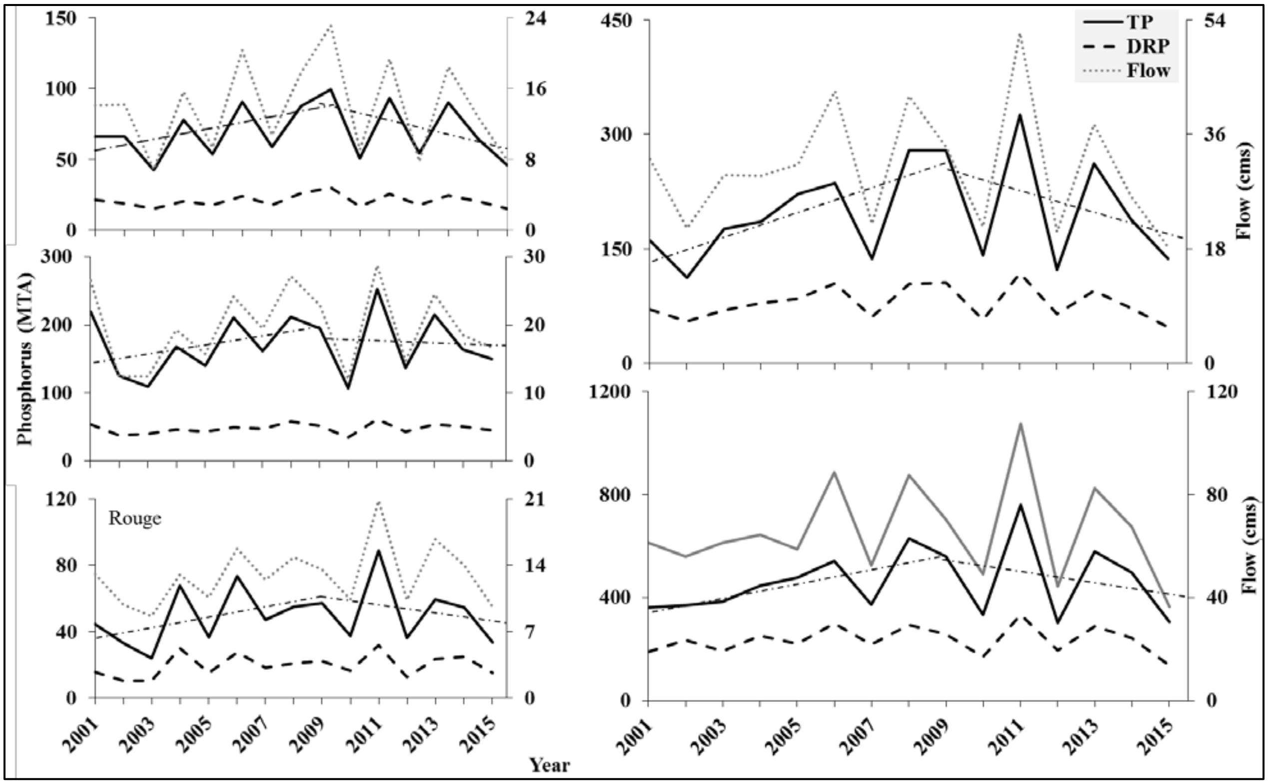

Figure S3: Annual total phosphorus (TP,solid black line), dissolved reactive phosphorus (DRP, broken black line) and flow (broken grey line) annual time series for each major sub-watershed. Unlabeled black center line indicates general trend (regression) line for TP for 2001-2009 and 2009-2015.