JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 49, No. 5

AMERICAN WATER RESOURCES ASSOCIATION

October 2013

AN OPTIMIZATION METHOD FOR ESTIMATING CONSTITUENT MEAN CONCENTRATIONS IN BASE FLOW-DOMINATED FLOW[1]

Laurent Ahiablame, Bernard Engel, and Indrajeet Chaubey[2]

ABSTRACT: Pollutant coefficients have been widely used to assess runoff nonpoint source pollution from individual land uses (e.g., agricultural, residential) of a watershed. Pollutant coefficients, known as event mean concentrations (EMCs), were developed by the U.S. Environmental Protection Agency's Nationwide Urban Runoff Program (NURP) to serve as a national measure for characterizing pollutant loading in a receiving water body. The term "baseflow pollutant coefficient (BPC)" is used in this study as a surrogate for EMC to describe mean concentration of pollutants in base flow-dominated flow. A method for characterizing base flow quantity and quality for different land uses was explored using inverse modeling with two optimization techniques (a least square method and a genetic algorithm [GA] optimization), land use information, and streamflow quantity and quality data. The inverse model was formulated as a constrained minimization problem and demonstrated with data for 15 watersheds in Indiana. Results showed that estimated pollutant coefficients are comparable to the published literature. This indicates that the proposed method has the potential to effectively estimate constituent mean concentrations for pollutant load determination in gauged and ungauged watersheds, albeit more analysis with larger and more robust datasets is desirable to further refine and validate the accuracy of the approach.

(KEY TERMS: base flow; least square method; land use; nonpoint source pollution; genetic algorithm.)

Ahiablame, Laurent, Bernard Engel, and Indrajeet Chaubey, 2013. An Optimization Method for Estimating Constituent Mean Concentrations in Base Flow-Dominated Flow. *Journal of the American Water Resources Association* (JAWRA) 49(5): 1167-1178. DOI: 10.1111/jawr.12076

INTRODUCTION

Quantification of pollutant losses to fluvial systems has attracted a great deal of interest in watershed planning and management (Cohn, 1995; NRC, 2001; GAO, 2003; Cohn, 2005). Downstream transport of pollutants, including nutrients, can lead to the deterioration of water quality conditions in lakes, reservoirs, and estuaries (USEPA, 2000), with substantial effects on human health. Typical examples include hypoxia in the Gulf of Mexico and eutrophication in the Great Lakes (Richards *et al.*, 2002; Petrolia and Gowda, 2006; Alexander *et al.*, 2008).

In the United States, target loads have been established through total maximum daily loads (TMDL),

[1]Paper No. JAWRA-12-0195-P of the *Journal of the American Water Resources Association* (JAWRA). Received August 28, 2012; accepted February 12, 2013. **Discussions are open until six months from print publication.**

[2]Respectively, Post-Doctoral Research Assistant (Ahiablame) and Professor (Engel), Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana 47907; and Professor (Chaubey), Department of Agricultural and Biological Engineering, Department of Earth, Atmospheric and Planetary Sciences, and Division of Environmental and Ecological Engineering, Purdue University, 225 South University Street, West Lafayette, Indiana 47907-2093 (E-Mail/Chaubey: ichaubey@purdue.edu).

under the provisions of Section 303(d) of the Clean Water Act, to diminish problems associated with the transport of pollutants in streams, and facilitate the protection and restoration of downstream water bodies (Saad *et al.*, 2011). Evaluation of the effectiveness of TMDLs for water quality improvement can be achieved through pollutant monitoring with field studies or through the use of mathematical algorithms to represent field conditions (Shirmohammadi *et al.*, 2006). Even though pollutant monitoring in streams is necessary for characterizing and identifying changes or trends in water quality over time, it is often limited to specific conditions (temporal and spatial), primarily because such monitoring studies require substantial time and financial resources. Simulation modeling is an alternative to monitoring efforts for broad scale assessment over varying geologic, topographic, vegetation, soil, and weather conditions.

To improve the accuracy of modeling results and support planning, regulatory, and management efforts, simple yet reliable methodologies such as development of event mean concentrations (EMCs) are needed for quantifying levels of pollutant concentrations resulting from different land uses at the watershed scale. Event mean concentrations were developed by the U.S. Environmental Protection Agency's Nationwide Urban Runoff Program (NURP) to serve as a national measure for characterizing pollutant loading in a receiving water body based on runoff events from individual land uses (USEPA, 1983; Huber, 2006). The concept of EMC can also be used to gain understanding of the effectiveness of best management practices during runoff events (USEPA, 1983). In this study, "baseflow pollutant coefficient" (BPC) is used as a surrogate for EMC and refers to the long-term flow-weighted mean concentration of pollutants in base flow-dominated flow. Base flow-dominated flow is defined as daily flow in which the proportion of base flow is equal or greater than 90% of daily total streamflow.

Base flow and storm flow are the two basic components of streamflow. Changes in the amount and quality of one will generally affect total streamflow and its quality. Base flow is important for its role in sustaining streamflow, especially during low-flow periods that occur on the majority of days in any given year. Contaminated base flow can adversely affect streamflow quality characteristics and associated ecosystem health. Knowledge of pollutant concentrations in base flow is needed to accurately estimate and predict levels of pollution resulting from different land use and land management activities. This will allow development of mitigation efforts at the watershed scale for effective planning and management of water resources.

An approach for estimation of BPC for total phosphorus (TP) and total nitrogen (TN) generated from various land use types was explored with two optimization techniques, consisting of a least square and a genetic algorithm (GA) optimization techniques. GAs have been extensively used in a variety of water quality studies such as selection and placement of best management practices for water quality improvement, cost optimization for water quality management, determination of pollutant removal levels from river systems, and calibration of watershed and water quality models (e.g., Wang, 1991; Franchini, 1996; Veith *et al.*, 2003; Liu *et al.*, 2007; Preis and Ostfeld, 2008; Maringanti *et al.*, 2009, 2011). However, little information is available on the use of GA to predict nonpoint source (NPS) pollutant concentration in streams and rivers.

The most common technique reported in the literature for estimating pollutant concentrations and loads in streams and rivers is the least square regression. For example, Driver and Tasker (1990) used multiple regression analysis to develop a series of models with total contributing drainage area, impervious area, land use, and mean annual climatic characteristics to estimate stormwater constituent loads, stormwater volumes, and mean constituent concentrations for urban watersheds in the United States. Spahr *et al.* (2010) also focused on the conterminous United States to develop regression equations for estimating flow-proportional TP and TN concentrations in streams. The Spahr *et al.*'s models predicted concentrations were less than 5 mg/l for the majority of streams in the nation, but more than 5 mg/l for the Midwest region (from Ohio to Eastern Nebraska), especially for TN (due to increased application of fertilizer and manure in this region). Similar findings were reported by Smith *et al.* (2003) who also applied regression analysis using data from 63 minimally impacted U.S. Geological Survey (USGS) reference basins to predict natural background concentrations of nutrients in the conterminous United States streams/rivers.

Even though previous studies investigated water quality constituent concentrations in streams and rivers (e.g., Driver and Tasker, 1990; Smith *et al.*, 2003; Schilling and Wolter, 2005), the relationship between land use type and water quality in base flow conditions has not been amply addressed in the literature. This study used an inverse modeling approach to estimate mean pollutant concentrations in base flow-dominated streamflow from known pollutant loads (calculated from measured concentrations), stream discharge, and watershed land uses. The overall goal of our study was to determine if a simplified procedure to approximate pollutant delivery proportions of individual land use types in a watershed using known/measured loads is feasible. The specific objectives were to estimate the following: (1) base flow-dominated flow quantity and (2) base flow-dominated flow pollutant coefficients for different land uses. This study evaluated the feasibility of the proposed methodology with two optimization techniques: a least square method and a GA optimization technique. The proposed methodology was demonstrated with a group of watersheds in Indiana, but can be easily replicated and applied in other locations and regions.

MATERIALS AND METHODS

Data

The study was conducted with 15 Indiana watersheds (Figure 1). The climate in Indiana is continental and temperate with warm summers and cold winters. Average daily air temperature varies between −10 and 29°C in the north and −6 to 32°C in the south. Average annual precipitation varies from north to south between 890 and 1,100 mm (ISCO, 2011), with nearly 69% (710 mm or 28 in) returning to the atmosphere in the form of evapotranspiration (Clark, 1980; Fowler and Wilson, 1996).

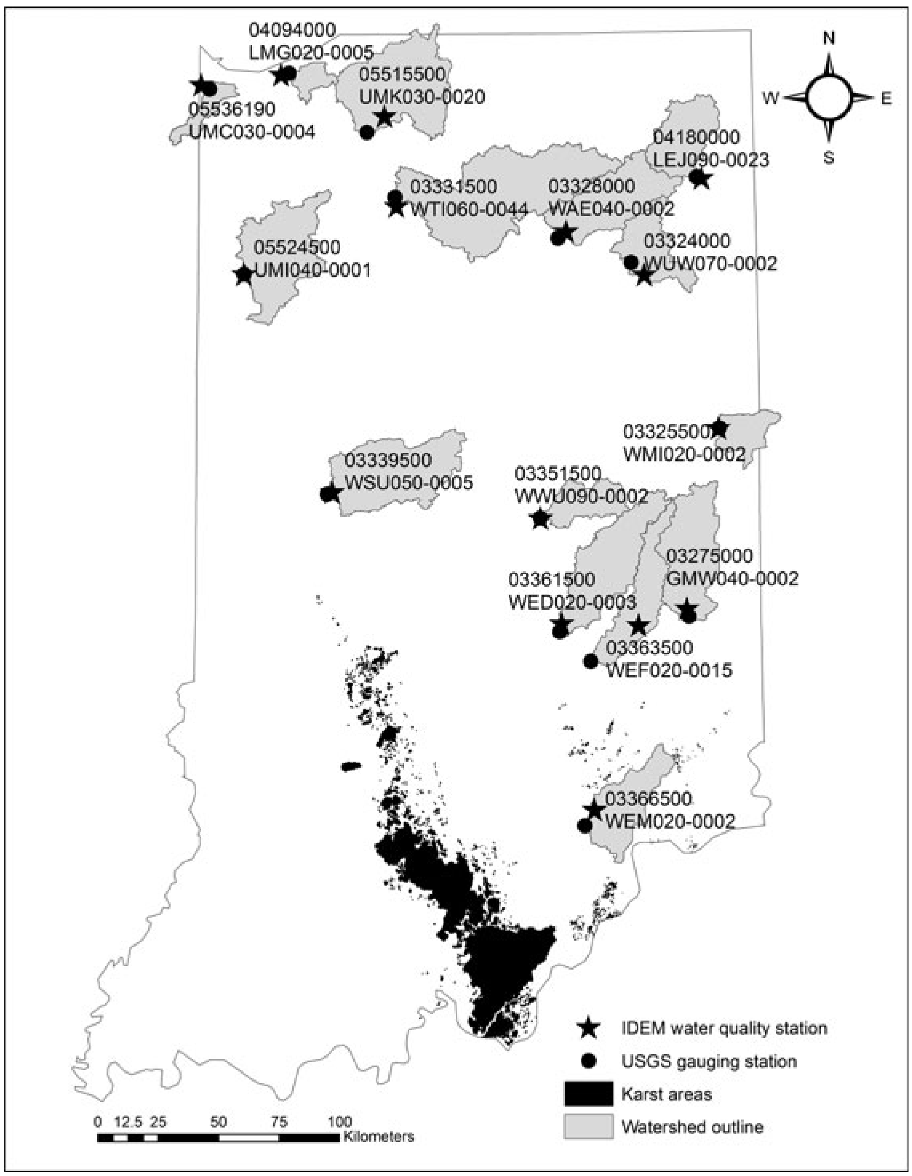

FIGURE 1. Location of U.S. Geological Survey (USGS) Gauging Stations (first label), Indiana Department of Environmental Management (IDEM) Water Quality Fixed Stations (second label), and 15 Delineated Watersheds in Indiana as Used for Nutrient Concentration Estimation in Base Flow-Dominated Flow.

The selection of the 15 watersheds was driven by the availability and quality of streamflow and water quality data. This analysis was part of a larger study (Ahiablame *et al.*, 2013), which identified candidate watersheds for base flow estimation. Criteria used for watershed selection include availability of long-term streamflow data free from intense flow regulation and diversion, availability of water quality stations close to streamflow gauge stations (as discussed below), exclusion of watersheds with karst landscapes, and exclusion of streams receiving discharge from wastewater treatment plants. The 15 watersheds were manually delineated with ArcGIS software to determine the contributing areas and land uses draining into the gauging stations used for the analysis. The areas of the delineated watersheds range from 170 to 2,200 km^2 (Table 1). These watersheds are predominantly agricultural watersheds, except one watershed, which is highly urbanized (Tables 1 and 2).

The land use datasets were derived from the 2001 National Land Cover Database (Homer *et al.*, 2004). A description of the land uses in each watershed is shown in Table 2. Because Indiana is undergoing rapid land use change (Rizkalla and Swihart, 2009), the 2001 National Land Cover Data (NLCD) was utilized as midpoint to balance the effects of changes in watershed responses that may affect estimates of water quality constituents for the period of 1991-2010. The 2001 NLCD categories were reclassified to have more general land use classes (Tables 1 and 2). Deciduous forest, evergreen forest, and mixed forest were aggregated into forest land use. Developed open space, developed low intensity, developed medium density, developed high density, and commercial and industrial land uses were combined into urban land use. Cultivated crops were categorized as agricultural land use. Bare soil was the combination of barren land, rock, sand, and clay with little or no green. However, bare soil (which is 1%, on average, of the total land use area in the watersheds) is negligible compared with the other land uses; so bare soil was merged in equal proportions with agricultural and urban land use after checking the distribution of bare soil in aerial photographs. This was carried out to minimize the effect of outliers in the data for the statistical analysis. Grass land cover consisted of shrub, scrub, grassland, pasture, and hay. Open water, woody wetlands, and emergent herbaceous wetlands were combined into water land use, and equally distributed on the other four land use types.

Twenty years (1991-2010) of daily streamflow records were obtained from the USGS National Water Information System (NWIS) of the 15 gauging stations (Table 1; Figure 1). Base flow was determined from streamflow data using the Base flow Filter Program for base flow separation (BFLOW) (Arnold *et al.*, 1995). The BFLOW program has been widely used and validated by previous researchers with data for Indiana (Lim *et al.*, 2005). Both streamflow and filtered base flow datasets for each watershed were further processed to retain only daily base flow-dominated flows, that is, daily flow in which the proportion of base flow is equal or greater than 90% of daily total streamflow. The 90% level was used to ensure that the majority of flow in the streams consists of base flow. Smakhtin (2001) reported that streamflow discharge may be composed entirely of base flow, especially in dry weather conditions. It should be noted that base flow constitutes approximately 60% of total streamflow, on average, in Indiana (Ahiablame *et al.*, 2013).

Water quality data from 1991 to 2010 were compiled, and LOADEST (Runkel *et al.*, 2004) was used

TABLE 1. Watershed Area and Percent Land Use Based on 2001 NLCD in 15 Indiana Watersheds.

USGS ID	IDEM ID	Watershed Area (km^2)	FLC	ULC	GLC	ALC
03275000	GMW040-0002	1272.7	13.9	8.1	13.4	64.5
03324000	WUW070-0002	657.3	8.3	13.3	4.3	74.2
03325500	WMI020-0002	342.7	5.1	6.7	3.6	84.7
03328000	WAE040-0002	1030.4	11.0	8.2	4.2	76.7
03331500	WTI060-0044	2227.0	11.9	9.9	6.2	72.0
03339500	WSU050-0005	1299.1	4.0	7.7	3.5	84.7
03351500	WWU090-0002	447.8	6.7	14.4	6.9	72.2
03361500	WED020-0003	1087.0	8.4	9.5	7.4	74.6
03363500	WEF020-0015	772.0	7.4	6.8	4.2	81.8
03366500	WEM020-0002	745.5	51.7	4.5	14.4	29.4
04094000	LMG020-0005	171.4	8.3	4.8	3.7	83.1
04180000	LEJ090-0023	671.6	14.8	10.9	9.2	65.0
05515500	UMK030-0020	1416.5	17.8	10.6	8.5	63.0
05524500	UMI040-0001	1168.4	8.8	6.2	3.8	81.3
05536190	UMC030-0004	218.3	16.4	41.8	15.0	26.9

Notes: FLC, Forest land cover; ULC, urban land cover; GLC, grass land cover; ALC, agricultural land cover.

TABLE 2. Classification of 2001 NLCD of 15 Indiana Watersheds into Four Categories.

Symbol	Land Use Description
FLC	Percent deciduous, evergreen, and mixed forests
ULC	Percent developed open space, developed low intensity, developed medium density, and developed high density
GLC	Percent grassland, herbaceous, pasture, hay, and shrub or scrub
ALC	Percent cultivated crops

Notes: FLC, Forest land cover; ULC, urban land cover; GLC, grass land cover; ALC, agricultural land cover.

to compute daily loads in these streams. The water quality data were obtained from the Indiana Department of Environmental Management (IDEM) database (Bell, 2011) (Tables A1 and A2). As water quality data are generally not collected on a daily basis, preliminary screenings were conducted to check the availability and quality of the datasets. The IDEM fixed stations deemed suitable for this study were collocated with streamflow stations. Saad *et al.* (2011) defined collocation as measurement of streamflow and water quality with observation stations that are located on the same river and drainage areas within 5%. This approach is one of the techniques used for integrating water quality and streamflow data to develop SPARROW (SPAtially Referenced Regressions On Watershed attributes) models as discussed by Schwarz *et al.* (2006) and Saad *et al.* (2011). The daily loads corresponding to daily base flow-dominated flows were computed and used for TP and TN loads.

Optimization Approach Description

The proposed methodology entails two steps. The first step is the partitioning of base flow-dominated flow for individual land uses (Table 1). The second step is the approximation of BPC for each land use type. A series of equations, developed with known land use areas, nutrient loads, and base flow-dominated flow, were solved as a constrained minimization problem to estimate base flow-dominated flow quantity and BPC values using (1) a least square method and (2) a GA optimization in MATLAB 7.14 (R2012b). The MATLAB GA toolbox is part of the MATLAB Global Optimization Toolbox for solving problems with multiple maxima or minima, and nonsmooth optimization problems (Chipperfield *et al.*, 1994; Chipperfield and Fleming, 1995). The MATLAB Global Optimization Toolbox supports various GA options (http://www.mathworks.com/help/gads/genetic-algorithm-options.html#f6593). The default selection function, stochastic uniform, was used in this study to randomly sample an initial population from the brackets of bounds (see Table 3) given for base flow-dominated flows (or pollutant concentrations) to evaluate their fitness with the objective function such that the sum of squares of the differences between observed and estimated base flow-dominated flows (or pollutant coefficients) was minimized. This process is repeated and continues through selection, mutation, and crossover of individuals for a given number of generations (iterations) until optimal solutions (i.e., optimum base flow-dominated flow quantity per land use or pollutant concentrations) are found. Known as robust and efficient, the GA can be used to solve both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution (Holland, 1975; Goldberg, 1989; Doherty *et al.*, 2004).

The least square method is a widely used approach to solving overdetermined systems of equations to provide approximate solutions (Seal, 1967; Plackett, 1972). The least square method is one of the oldest techniques of modern statistics to minimize the sum of the squared deviations between measured and predicted values (Abdi, 2003). It varies from simple ordinary to more sophisticated nonlinear least squares (Abdi, 2003).

The proposed approach is based on the following basic assumptions:

1. Base flow originated from land use types is based on the following order: base flow from forest

TABLE 3. Lower and Upper Bounds as Used for Estimation of Base Flow-Dominated Flow Quantity and Base Flow Pollutant Coefficient (BPC) for the Genetic Algorithm Optimization.

	Base Flow-Dominated Flow Quantity		BPC (Total Phosphorus)		BPC[2] (Total Nitrogen)	
Land Use	Lower Bound (m^3/yr)	Upper Bound (m^3/yr)	Lower Bound (mg/l)	Upper Bound (mg/l)	Lower Bound (mg/l)	Upper Bound (mg/l)
Forest	300	2,000	-	0.0075	-	-
Urban	70	250	-	0.06	-	-
Grass	150	2,000	-	0.05	-	-
Agricultural	70	4,000	-	0.061	-	-

areas > base flow from grass and agricultural areas > base flow from urban areas.
2. Base flow originates from areas within watershed boundaries.
3. Only land use explains trends in pollutant concentrations and transport in streams and rivers.

Partition of Base Flow-Dominated Flow

The first step consisted of determining base flow contribution from individual land uses in the study watersheds. Average annual base flow-dominated flow and individual land use areas in the 15 study watersheds for the first 10 years of the study period (1991-2000) were used to estimate the volume of base flow-dominated flow per unit area for each of the land uses as expressed by the following:

$$\sum_{i=1}^{n} Q_{\mathrm{BF}} = \sum_{i=1}^{n} \left(q_{\mathrm{BF(FLC)}} A_{\mathrm{FLC}}\right) + \left(q_{\mathrm{BF(ULC)}} A_{\mathrm{ULC}}\right) + \left(q_{\mathrm{BF(GLC)}} A_{\mathrm{GLC}}\right) + \left(q_{\mathrm{BF(ALC)}} A_{\mathrm{ALC}}\right) \tag{1}$$

where, n is the number of study watersheds; Q_{BF} is the average annual base flow-dominated flow volume; A is the area of a given land use in the watershed; $q_{(\mathrm{BF})}$ is the base flow-dominated flow per unit area; FLC is forest land cover; ULC is urban land cover; GLC is grass land cover; and ALC is agricultural land cover in the watersheds.

Solving for $q_{(\mathrm{BF})}$ for individual land uses in Equation (1) resulted in a set of base flow-dominated flow values for each of the land uses. An initial screening analysis using the least square method (with positive constraints) provided a general idea about the initial starting points for setting the lower and upper bounds for the optimization techniques. The constraints characterizing differences in base flow coming from individual land uses were determined based on the published literature such as Rothacher (1970), Hornbeck *et al.* (1993), Robinson *et al.* (2003), Schilling and Libra (2003), Zhang and Schilling (2006), and Price (2011). Following these studies, assumptions (as described earlier) were made to reason that forest produces more base flow than urban land use, but less or same base flow as grass land use, and agricultural land use (for Indiana conditions) produces more base flow than grass.

The least square method and GA optimization were then used, independently, to approximate base flow-dominated flow per unit area for the individual land uses based on the initial starting constraint values (Table 3). The estimated base flow-dominated flow values for the individual land uses were used as regression coefficients and evaluated in Equation (1) for each watershed. The regression coefficients are q_{BF} values (base flow-dominated flow per unit area) describing the contribution of individual land uses to the total base flow-dominated flow. The data of the remaining 10 years of the study period (2001-2010) were used to validate the estimated base flow-dominated flow using Equation (1).

Estimation of Base Flow Pollutant Coefficients

The second step of the analysis was the estimation of BPCs, surrogates for EMCs in base flow conditions. Event mean concentration, in mathematical terms, is defined as the mass of pollutant contained in runoff event over the total volume of flow in the event (USEPA, 1983; Huber, 2006). The BPC values for various land uses were determined using known base flow-dominated flow and associated pollutant loads (estimated with LOADEST) as follows:

$$\sum (\mathrm{Load})_i = \sum \left(Q_{\mathrm{v(LC)}i} [\mathrm{BPC}]_{(\mathrm{LC})i}\right)_i \tag{2}$$

where $(\mathrm{Load})_i$ is the nutrient load in watershed i; $Q_{\mathrm{v(LC)}i}$ is the discharge of base flow-dominated flow for land use i in watershed i; and $[\mathrm{BPC}]_{(\mathrm{LC})i}$ is the base flow pollutant coefficient associated with land use i. The extended form of Equation (2) in a single watershed can be expressed as follows:

$$\mathrm{Load} = \left(Q_{\mathrm{v(FLC)}} \times \mathrm{BPC}_{\mathrm{FLC}}\right) + \left(Q_{\mathrm{v(ULC)}} \times \mathrm{BPC}_{\mathrm{ULC}}\right) + \left(Q_{\mathrm{v(GLC)}} \times \mathrm{BPC}_{\mathrm{GLC}}\right) + \left(Q_{\mathrm{v(ALC)}} \times \mathrm{BPC}_{\mathrm{ALC}}\right) \tag{3}$$

Similar to the estimation of base flow-dominated flow quantity, the BPC values were determined as constrained minimization of the objective function (formulated based on Equation 2) using nonnegative inequality constraints and bounds for both the least square method and the GA. The GA optimization was run under two conditions. First, base flow-dominated flows per unit area obtained from Step 1 ("Partition of Base Flow-Dominated Flow") were used as weighting factors to constrain base flow-dominated flow discharge (see Equation 3). Pollutant coefficients in the Spreadsheet Tool for Estimating Pollutant Load (STEPL) (USEPA and Tetra Tech, Inc., 2011) model were also used as guides to set constraints for BPCs across land use types. Under the second condition, the base flow-dominated flows were not constrained, but pollutant coefficients in the STEPL model were also used as guides to set constraints for BPCs across land use types. For the two conditions, there were no bounds (upper and lower) set for BPC estimates of TN (Table 3). However, upper bounds were set for BPC estimates of TP for the land use types to satisfy basic feasibility requirements (Table 3).

RESULTS AND DISCUSSION

Partition of Base Flow-Dominated Flow

This section is divided into two parts. The first part provides a general background on base flow and base flow-dominated flow in the study watersheds. The second part presents the contribution of base flow-dominated flow amount from land use types.

Average annual total streamflow and base flow vary between 350 and 450 mm/yr, and 150 and 310 mm/yr, respectively, in the 15 watersheds during the 20-year study period (Table 4). Base flow-dominated flow ranges from 75 to 190 mm/yr for the 15 watersheds (Table 4). Base flow-dominated days (i.e., days on which the proportion of base flow is 90% or more of daily streamflow) are approximately 200 days on average in the 15 watersheds (Table 4). There is no particular trend in base flow and base flow-dominated flow with respect to watershed size. Average annual filtered base flow is comparable to annual base flow reported for streams in the Midwest (Schilling and Libra, 2003; Gebert *et al.*, 2007; Schilling and Helmers, 2008), and constitutes about 60% of streamflow. The 60% streamflow as base flow is consistent with a previous study conducted by Ahiablame *et al.* (2013) for Indiana. The relatively high base flow observed in Indiana streams could be explained by intense agricultural activities (93% of the watershed used in this study are highly agricultural; see Tables 1 and 2), such as row crops, a common practice in Indiana (Smith *et al.*, 2008), and tile drainage impacts as demonstrated by Schilling and Libra (2003) and Schilling and Helmers (2008).

Estimated base flow-dominated flow quantity for each land use with the least square technique did not lead to reasonable results (based on the scientific literature and expert opinion). Thus, the least square results were not herein reported, and the limitations of the approach are discussed in a section below.

Estimated base flow-dominated flow quantity for each land use with the GA technique shows a good fit with the observed (i.e., filtered) base flow-dominated during the calibration period (R^2 = 0.90; Figure 2) and the validation period (R^2 = 0.84; Figure 2), indicating that the model has a reasonable prediction capability. Based on the analysis, it appears that changes in agricultural areas would influence base flow more than forest and grass land uses in the study watersheds. The base flow-dominated flow volume for a watershed can be estimated as follows:

$$Q_{\mathrm{BF}} = 1217.1A_{\mathrm{FLC}} + 249.9A_{\mathrm{ULC}} + 798.1A_{\mathrm{GLC}} + 1475.6A_{\mathrm{ALC}} \tag{4}$$

This analysis suggests that average annual base flow-dominated flow corresponding to each hectare of agricultural land use is expected to be 1,500 m^3, and 1,200 m^3 for each hectare of forest area, 800 m^3 for each hectare of grass area, and 250 m^3 for each hectare of urban land area (Equation 4). From a practical standpoint, the analysis indicates that agricultural areas would contribute more base flow quantity than forested areas, followed by grass areas, and urban areas, where the conversion of natural landscapes into buildings, roads/streets, and parking lots, increase impervious surfaces. This pattern could be expected for the United States Midwest region due to

TABLE 4. Observed Total Flow, Filtered Base Flow, and Estimated Base Flow-Dominated Flow and Days in 15 Indiana Watersheds.

USGS ID	IDEM ID	Total Flow (mm/yr)	Base Flow (mm/yr)	Base Flow-Dominated Flow (mm/yr)	Base Flow-Dominated Days (days)
03275000	GMW040-0002	417	271	169	252
03324000	WUW070-0002	368	173	87	230
03325500	WMI020-0002	388	151	75	225
03328000	WAE040-0002	376	232	134	244
03331500	WTI060-0044	395	316	188	232
03339500	WSU050-0005	350	193	99	217
03351500	WWU090-0002	435	281	167	237
03361500	WED020-0003	435	281	165	241
03363500	WEF020-0015	428	242	127	212
03366500	WEM020-0002	448	281	109	216
04094000	LMG020-0005	457	313	195	214
04180000	LEJ090-0023	362	218	123	233
05515500	UMK030-0020	359	317	138	259
05524500	UMI040-0001	369	243	118	206
05536190	UMC030-0004	420	255	131	218

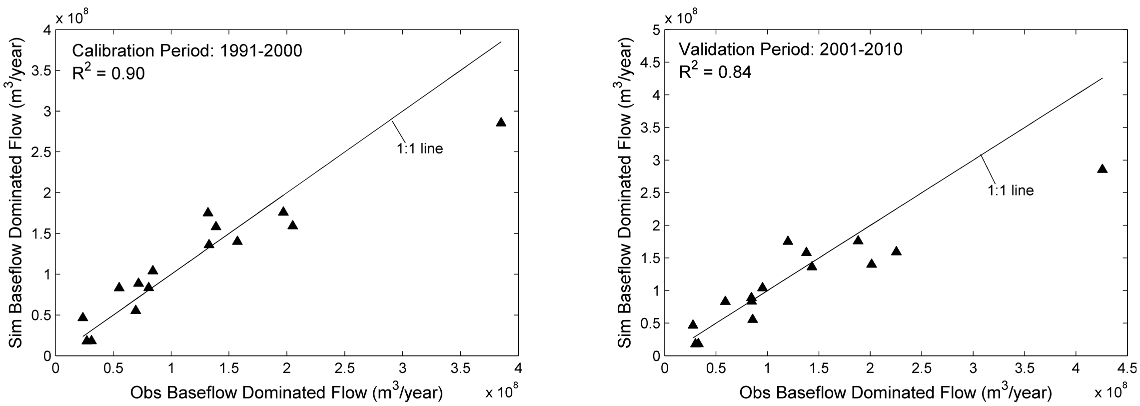

FIGURE 2. Simulated and Filtered Base Flow-Dominated Flow in 15 Indiana Watersheds During the Calibration and Validation Periods.

the extensive tile drainage networks (Schilling and Helmers, 2008).

Decreases in base flow due to urbanization have been widely reported; however, increases in impervious surface may not necessary lead to base flow reduction at the watershed scale (Brandes *et al.*, 2005), suggesting that variability in urban land area would not affect base flow to the same extent as agricultural land uses (especially in Indiana where tile drained agricultural landscape is common). Urbanization may also affect base flow in watersheds developed with water features which promote base flow increase such as detention basins and ponds.

The similarity in base flow contribution between forest and grass land uses is debatable, yet not unrealistic. Although forests are often credited with a variety of environmental benefits, they may have increased interception of precipitation and increased evapotranspiration rates compared with grass areas, reducing base flow recharge (Keppeler and Ziemer, 1990; McCulloch and Robinson, 1993; Price, 2011). Robinson *et al.* (2003), commenting on a study conducted by Engler (1919) who compared base flow from a forested watershed and base flow from grass dominated watershed, reported that Engler's (1919) findings resulted in low peak flow and high base flow yields for the forested watershed. However, Robinson *et al.* (2003) explained that the forested watershed had deeper soils so it was not clear to what extent base flow increase was the effect of vegetation or soil.

Estimated Base Flow Pollutant Coefficients

Base flow pollutant coefficients characterize proportion of pollutants delivered by individual land uses in a watershed. Attempts to isolate pollutant coefficients for each of the land uses using regression analysis (i.e., the least square method) resulted in limited success. This could be due to the lack of variability in land use type across the study watersheds as more than 85% of the watersheds are highly agricultural. The lack of variability in land uses may not provide enough coverage to represent individual land use types for deriving good BPC estimates. However, the GA optimization technique provided reasonable results.

The estimated BPCs for TP and TN with the GA optimization method are shown in Table 5. Differentiation of the individual land use contribution does not vary noticeably across the land uses, except for forest which has the lowest BPC for both nutrients (Table 5). The BPC values vary from 0.008 to 0.060 mg/l for TP, and from 0.26 to 4.71 mg/l for TN (Table 5). These values are consistent with the published literature (discussed below), and BPCs used in simple pollutant estimation models such as STEPL. The STEPL model uses 0.063 mg/l TP for urban, cropland, and pasture, 0.009 mg/l for forest, and 0.070 mg/l for feedlot in shallow groundwater (Table 5). TN concentration in STEPL varies from 0.11 mg/l for forest to 6.00 mg/l for feedlots in shallow groundwater (Table 5). The estimated BPC values with the GA approach are also comparable to findings from previous studies conducted on base flow chemistry in the Midwest United States, for which the authors reported 0-0.1 mg/l for TP and 0-12 mg/l for TN (Schilling and Wolter, 2001; V3 Companies, 2006). It should be kept in mind that the base flow data used in this analysis were base flow-dominated flow, indicating that the estimated concentrations may contain small fractions of runoff pollutants. There was no statistically significant difference between means of BPC values obtained by partitioning base flow-dominated flow and means of BPC

TABLE 5. Modeled Concentrations Using Genetic Algorithm Optimization for Total Phosphorus and Total Nitrogen in 15 Indiana Watersheds.

	Total Phosphorus (mg/l)		Total Nitrogen (mg/l)	
Land Use	**GA**	**STEPL**	**GA**	**STEPL**
FLC	0.008	0.009	0.26	0.11
ULC	0.060	0.063	4.71	1.50
GLC	0.050	0.063	2.67	1.44
ALC	0.060	0.063	2.91	1.44

Notes: FLC, Forest land cover; ULC, urban land cover; GLC, grass land cover; ALC, agricultural land cover.

values obtained without partitioning base flow-dominated flow.

The presence of nutrients in base flow could be the result of interactions of direct runoff (total flow component) and base flow, or discharge of groundwater that receives nutrient inputs through percolation. Nutrients in base flow may primarily come from agricultural activities, especially in the Midwest, as the majority of the watersheds used for the analysis have high proportions of agricultural land uses (Tables 1 and 2). Previous studies showed increased nutrient losses from agricultural landscapes in this region (Alexander *et al.*, 2008; Smith *et al.*, 2008; Ahiablame *et al.*, 2011).

Limitations of the Method and Lessons Learned

The biggest limitations observed in this study arise from the use of the least square method, which did not lead to reasonable results for base flow-dominated flow quantity and base flow pollutant coefficients comparable to results obtained with the use of the GA optimization technique. The lack of range in variability in land use types across the study watersheds was a contributing factor to the limitations of the proposed method. Although the study watersheds have different land uses, they are all dominated by agricultural land use (more than 85% of the watersheds have each more than 60% of agricultural land use). Variability refers to the spread of the proportion of a specific land use (e.g., different proportions of urban land use) from one watershed to another. Variability in the proportion of land uses is essential for capturing (balancing) the contribution of pollutants from individual land uses.

Another limitation resides in the assumptions made during this study. The analysis assumed that only land uses explain trends in pollutant concentrations in streams and rivers. The study also assumed that base flow originates only from areas within study watershed boundaries. Historically, upland activities strongly influence nutrient losses to downstream rivers and lakes (Alexander *et al.*, 2008; Smith *et al.*, 2008). Besides land use, other factors such as climate, geology, soils, anthropogenic related terrestrial processes, management practices, among others, may also affect pollutant transport and delivery to streams and rivers. Moreover, base flow may come from areas that do not match with the study watershed boundaries.

The proposed method is sensitive to the quality of water quality datasets, suggesting that measurement errors would lead to uncertainty in estimated concentrations. This study did not explicitly take into consideration measurement uncertainties. Base flow and pollutant concentrations are inherent stochastic processes that are not easily optimizable with inverse modeling. Even with a series of constraints, a small uncertainty in the data may result in increased difficulties to find good estimates. Difficulties also appeared when setting the bounds for the GA realizations for BPC estimates for TP, as a small change in the bounds for one specific land use type strongly influenced BPC estimates for other land uses.

The limitations discussed above were even more accentuated with the use of least square regression. Estimation of BPC values with the least square regression resulted in unrealistic results due likely to the lack of variability in land use type (across watersheds) as mentioned above. The least square method was also sensitive to the quality of water quality datasets. Sensitivity to the presence of unusual data points and poor extrapolation properties intrinsic to the least square method (Seal, 1967; Plackett, 1972; Abdi, 2003) may also influence the accuracy of the estimates of base flow pollutant coefficients.

In general, careful selection of water quality datasets, study watersheds (keeping in mind the variability in the proportion of land use types), and extending the study to a larger number of watersheds may improve results. The water quality data used for this study were public datasets collected at irregular time frames over the study period. These measured data were used to estimate continuous pollutant loads with LOADEST (Runkel *et al.*, 2004). Using long-term observed water quality data (to compute observed long-term loads instead of estimated loads) could also provide better estimates of base flow-dominated flow pollutant coefficients.

CONCLUSIONS

An inverse modeling method was developed to estimate constituent concentrations for base flow conditions from various land uses using data for 15 Indiana watersheds. The method consists of base flow-dominated flow partitioning for different land use types and estimation of pollutant mean concentrations in base flow-dominated flow using two different techniques: a least square method and a GA optimization approach. The least square regression did not lead to realistic results, whereas the GA optimization technique appeared to be a robust method for partitioning base flow-dominated flow per land uses. Results for the GA approach show that variability in base flow-dominated flow is strongly influenced by the proportion of agricultural land use, followed by forest, grass, and urban land uses.

Differentiation of pollutant delivery from individual land uses with the least square method resulted in limited success, likely due to data limitations. However, the GA optimization technique yielded better base flow pollutant coefficient estimates which vary from 0.008 to 0.060 mg/l for TP, and from 0.26 to 4.71 mg/l for TN for the individual land uses. This indicates that the proposed methods have potential to inversely estimate constituent mean concentrations for use in watershed models. Extension of this study should consider using a larger number of watersheds and extensive long-term and good water quality datasets across states and regions. Research is needed to improve the method with watersheds of similar size and similar land use proportions (within a watershed) to reduce sensitivity of the least square method to extreme data points. Based on the lessons learned in this study, the use of the simplest technique, that is, the constrained least square method, on robust and large datasets is recommended to avoid complex and sophisticated techniques, such as, GAs. However, when the available dataset is limited, GAs may provide superior results in estimating constituent mean concentrations.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Kossi Nouwakpo, Dr. Keith Cherkauer, Dr. Venkatesh Merwade, Dr. K.P. Sudheer, and Cibin Raj for their useful comments and help in this study. The authors acknowledge the financial support of the Illinois-Indiana Sea Grant (IISG) — College Program Research and Outreach Development and Capacity Building Projects. Comments provided by three anonymous reviewers to improve the quality of the manuscript are greatly appreciated.

APPENDIX

TABLE A1. Phosphorus and Nitrogen Concentrations, and Their Standard Deviation in 15 Indiana Watersheds for Base Flow-Dominated Flow.

	Total P (mg/l)		Total N (mg/l)	
USGS ID	Mean (STD)	Number of Samples	Mean (STD)	Number of Samples
03275000	0.084 (0.239)	38	3.07 (0.948)	46
03324000	0.038 (0.295)	221	2.21 (1.02)	221
03325500	0.078 (0.323)	139	3.79 (3.65)	179
03328000	0.061 (0.184)	40	1.02 (0.635)	40
03331500	0.032 (0.269)	173	1.07 (1.51)	145
03339500	0.040 (0.276)	62	2.10 (0.406)	55
03351500	0.035 (0.090)	102	4.11 (0.442)	213
03361500	0.042 (0.248)	146	3.67 (3.20)	170
03363500	0.053 (0.086)	129	3.66 (2.49)	156
03366500	0.041 (0.053)	49	4.33 (2.38)	49
04094000	0.075 (0.008)	143	0.75 (0.270)	143
04180000	0.052 (0.121)	136	1.81 (1.97)	148
05515500	0.033 (0.076)	124	1.41 (1.67)	180
05524500	0.057 (0.057)	54	3.97 (6.14)	54
05536190	0.046 (0.207)	168	2.38 (27.4)	168

TABLE A2. Average Annual Nutrient Loads Computed with LOADEST in 15 Indiana Watersheds

	Total P		Total N	
USGS ID	Total Loads (kg/ha/yr)	Base Flow[1] Loads (kg/ha/yr)	Total Loads (kg/ha/yr)	Base Flow[1] Loads (kg/ha/yr)
03275000	1.75	0.15	8.21	6.66
03324000	1.89	0.03	16.95	4.57
03325500	1.78	0.06	15.93	4.93
03328000	1.63	0.28	11.02	4.08
03331500	0.40	0.21	9.41	3.94
03339500	1.07	0.13	17.72	8.33
03351500	1.14	0.04	15.97	8.31
03361500	0.89	0.37	18.22	7.19
03363500	1.37	0.16	8.38	5.60
03366500	0.42	0.25	4.50	2.86
04094000	0.31	0.05	3.13	0.40
04180000	0.81	0.04	7.08	0.52
05515500	0.46	0.01	7.16	0.44
05524500	1.37	0.02	8.12	0.24
05536190	2.22	0.01	9.25	0.49

[1]Base flow: Base flow-dominated flow (i.e., daily flow in which the proportion of base flow is equal or greater than 90% of daily total streamflow).

LITERATURE CITED

Abdi, H., 2003. Least Squares. *In*: Encyclopedia of Social Science Research Methods, M. Lewis-Beck, A. Bryman, and T. Futing (Editors). Sage, Thousand Oaks, California, pp. 559-561.

Ahiablame, L., I. Chaubey, B. Engel, K. Cherkauer, and V. Merwade, 2013. Estimation of Annual Baseflow at Ungauged Sites in Indiana, USA. Journal of Hydrology 476:13-27.

Ahiablame, L., I. Chaubey, D.R. Smith, and B.A. Engel, 2011. Effect of Tile Effluent on Nutrient Concentration and Retention Efficiency in Agricultural Drainage Ditches. Agricultural Water Management 98(8):1271-1279.

Alexander, R.B., R.A. Smith, G.E. Schwarz, E.W. Boyer, J.V. Nolan, and J.W. Brakebill, 2008. Differences in Phosphorus and Nitrogen Delivery to the Gulf of Mexico From the Mississippi River Basin. Environmental Science & Technology 42(3):822-830.

Arnold, J.G., P.M. Allen, R. Muttiah, and G. Bernhardt, 1995. Automated Baseflow Separation and Recession Analysis Techniques. Groundwater 33:1010-1018.

Bell, C.M., 2011. Water Quality Assessment Information Management System (AIMSII) Database. Indiana Department of Environmental Management (IDEM), Indianapolis, Indiana, *accessed* February 24, 2011.

Brandes, D., G.J. Cavallo, and M.L. Nilson, 2005. Base Flow Trends in Urbanizing Watersheds of the Delaware River Basin. Journal of the American Water Resources Association 41:1377-1391.

Chipperfield, A.J. and P.J. Fleming, 1995. The MATLAB Genetic Algorithm Toolbox. IEE Colloqium on Applied Control Techniques Using MATLAB, Digest No. 1995/014, 26 January.

Chipperfield, A.J., P.J. Fleming, H. Pohlheim, and C.M. Fonseca, 1994. A Genetic Algorithm Toolbox for MATLAB. *In*: Proceedings of the International Conference on Systems Engineering, Coventry, UK, 6-8 September, pp. 200-207.

Clark, G.D., 1980. The Indiana Water Resource. Indiana Department of Natural Resources, Indianapolis, 508 pp.

Cohn, T.A., 1995. Recent Advances in Statistical Methods for the Estimation of Sediment and Nutrient Transport in Rivers. Reviews of Geophysics 33:1117-1124.

Cohn, T.A., 2005. Estimating Contaminant Loads in Rivers: An Application of Adjusted Maximum Likelihood to Type 1 Censored Data. Water Resources Research 41:W07003, doi: 10.1029/2004WR003833.

Doherty, D., M.A. Freeman, and R. Kumar, 2004. Optimization with MATLAB and the Genetic Algorithm and Direct Search Toolbox. http://www.mathworks.com/company/newsletters/digest/sept04/optimization.html, *accessed* May 28, 2012.

Driver, N.E. and G.D. Tasker, 1990. Techniques for Estimation of Storm-Runoff Loads, Volumes, and Selected Constituent Concentrations in Urban Watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2363, U.S. Geological Survey, 44 pp.

Engler, A., 1919. Untersuchungen Uber den Einfluss DesWaldes auf den Stand der Gewasser. Mitt. Schweiz. Anst. Forst. Versuchswes 12:636.

Fowler, K.K. and J.T. Wilson, 1996. Low Flow Characteristics of Indiana Streams. USGS Water Resources Investigations Report 96-4128.

Franchini, M., 1996. Use of a Genetic Algorithm Combined With a Local Search Method for the Automatic Calibration of Conceptual Rainfall-Runoff Models. Hydrological Sciences Journal 41 (1):21-39.

GAO (Government Accountability Office), 2003. Water Quality: Improved EPA Guidance and Support Can Help States Develop Standards That Better Target Cleanup Efforts, Tech. Rep. GAO-03-308, U.S. Government Accountability Office, Washington, D.C.

Gebert, W.A., M.J. Radloff, E.J. Considine, and J.L. Kennedy, 2007. Use of Streamflow Data to Estimate Baseflow/Ground-Water Recharge for Wisconsin. Journal of the American Water Resources Association 43(1):220-236.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, Massachusetts.

Holland, J.H., 1975. Adoption in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, Michigan.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 National Landcover Database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840.

Hornbeck, J.W., M.B. Adams, E.S. Corbett, E.S. Verry, and J.A. Lynch, 1993. Long-Term Impacts of Forest Treatment on Water Yield: A Summary for Northeastern USA. Journal of Hydrology 150:323-343.

Huber, W.C., 2006. BMP Modeling Concepts and Simulation. EPA Contract No. 68-C-01-020 Water Supply and Water Resources Division. National Risk Management Research Laboratory, Edison, New Jersey.

ISCO (Indiana State Climate Office), 2011. Indiana Climate Atlas. http://www.iclimate.org/facts.asp, *accessed* April 25, 2011.

Keppeler, E.T. and R.R. Ziemer, 1990. Logging Effects and Streamflow: Water Yield and Summer Low Flows at Caspar Creek in Northwestern California. Water Resources Research 26(7):1669-1679.

Lim, K.J., B.A. Engel, Z. Tang, J. Choi, K. Kim, S. Muthukrishnan, and D. Tripathy, 2005. Automated Web GIS-Based Hydrograph Analysis Tool, WHAT. Journal of the American Water Resources Association 41(6):1407-1416.

Liu, S., D. Butler, R. Brazier, L. Heathwaite, and S.-T. Khu, 2007. Using Genetic Algorithms to Calibrate a Water Quality Model. Science of the Total Environment 374:260-272.

Maringanti, C., I. Chaubey, M. Arabi, and B. Engel, 2011. Application of a Multi-Objective Optimization Method to Provide Least Cost Alternatives for NPS Pollution Control. Environmental Management 48(3):448-461.

Maringanti, C., I. Chaubey, and J. Popp, 2009. Development of a Multiobjective Optimization Tool for the Selection and Placement of Best Management Practices for Nonpoint Source Pollution Control. Water Resources Research 45(W06406):1-15. doi: 10.1029/2008WR007094

McCulloch, J.S. and M. Robinson, 1993. History of Forest Hydrology. Journal of Hydrology 150:189-216.

NRC (National Research Council), 2001. Assessing the TMDL Approach to Water Quality Management. Natl. Acad. Press, Washington, D.C., 109 pp.

Petrolia, D.R. and P.H. Gowda, 2006. Missing the Boat: Midwest Farm Drainage and Gulf of Mexico Hypoxia. Review of Agricultural Economics 28(2):240-253.

Plackett, R.L., 1972. The Discovery of the Method of Least Squares. Biometrika 59:239-251.

Preis, A. and A. Ostfeld, 2008. A Coupled Model Tree-Genetic Algorithm Scheme for Flow and Water Quality Predictions in Watersheds. Journal of Hydrology 349(3-4):364-375.

Price, K., 2011. Effects of Watershed Topography, Soils, Land Use, and Climate on Baseflow Hydrology in Humid Regions: A Review. Progress in Physical Geography 35(4):465-492.

Richards, R.P., F.G. Calhoun, and G. Matisoff, 2002. The Lake Erie Agricultural Systems for Environmental Quality Project: An Introduction. Journal of Environmental Quality 31(1):6-16.

Rizkalla, C.E. and R.K. Swihart, 2009. Forecasting the Effects of Land-Use Change on Forest Rodents in Indiana. Environmental Management 44:899-908.

Robinson, M., A.L. Cognard-Plancq, C. Cosandey, J. David, P. Durand, H.-W. Fuhrer, R. Hall, M.O. Hendriques, V. Marc, R.

McCarthy, M. McDonnell, C. Martin, T. Nisbet, P. O'Dea, M. Rodgers, and A. Zollner, 2003. Studies of the Impact of Forests on Peak Flows and Baseflows: A European Perspective. Forest Ecology and Management 186:85-97.

Rothacher, J., 1970. Increases in Water Yield Following Clear-cut Logging in the Pacific Northwest. Water Resources Research 6:653-658.

Runkel, R., C.G. Crawford, and T.A. Cohn, 2004. Load Estimator (LOADEST): A Fortran Program for Estimating Constituent Loads in Streams and Rivers, U.S. Geological Survey Techniques Methods, Book 4, Chap. A5.

Saad, D.A., G.E. Schwarz, D.M. Robertson, and N.L. Booth, 2011. A Multi-Agency Nutrient Dataset Used to Estimate Loads, Improve Monitoring Design, and Calibrate Regional Nutrient SPARROW Models. Journal of the American Water Resources Association 47(5):933-949, doi: 10.1111/j.1752-1688.

Schilling, K.E. and M. Helmers, 2008. Effects of Subsurface Drainage Tiles on Streamflow in Iowa Agricultural Watersheds: Exploratory Hydrograph Analysis. Journal of Soil and Water Conservation 62(4):252-261.

Schilling, K.E. and R.D. Libra, 2003. Increased Baseflow in Iowa Over the Second Half of the 20th Century. Journal of the American Water Resources Association 39:851-860.

Schilling, K.E. and C.F. Wolter, 2001. Contribution of Baseflow to Nonpoint Source Pollution Loads in an Agricultural Watershed. Groundwater 39(1):49-58.

Schilling, K.E. and C.F. Wolter, 2005. Estimation of Streamflow, Baseflow and Nitrate-Nitrogen Loads in Iowa Using Multiple Linear Regression Models. Journal of the American Water Resources Association 41:1333-1346.

Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith, 2006. The SPARROW Surface Water- Quality Model-Theory, Applications and User Documentation: U.S. Geological Survey, Techniques and Methods 6-B3, 248 pp. and CD-ROM.

Seal, H.L., 1967. The Historical Development of the Gauss Linear Model. Biometrika 54:1-23.

Shirmohammadi, A., I. Chaubey, R.D. Harmel, D.D. Bosch, R. Muñoz-Carpena, C. Dharmasri, A. Sexton, M. Arabi, M.L. Wolfe, J. Frankenberger, C. Graff, and T.M. Sohrabi, 2006. Uncertainty in TMDL Models. Transactions of the ASABE 49 (4):1033-1049.

Smakhtin, V.Y., 2001. Low Flow Hydrology: A Review. Journal of Hydrology 240:147-186.

Smith, D.R., S.J. Linvingston, B.W. Zuercher, M. Larose, G.C. Heathman, and C. Huang, 2008. Nutrient Losses From Row Crop Agriculture in Indiana. Journal of Soil and Water Conservation 63(6):396-409.

Smith, R.A., R.B. Alexander, and G.E. Schwarz, 2003. Natural Background Concentrations of Nutrients in Streams and Rivers of the Conterminous United States. Environmental Science and Technology 37(14):3039-3047.

Spahr, N.E., D.K. Mueller, and D.W. Wolock, 2010. Development and Application of Regression Models for Estimating Nutrient Concentrations in Streams of the Conterminous United States, 1992-2001: U.S. Geological Survey Scientific Investigations Report 2009-5199, 22 pp.

USEPA (U.S. Environmental Protection Agency), 1983. Results of the Nationwide Urban Runoff Program, Vol. 1, Final Report, Water Planning Division, U.S. EPA, Washington, D.C.

USEPA (U.S. Environmental Protection Agency), 2000. Nutrient Criteria Technical Guidance Manual-Lakes and Reservoirs: Office of Water, Office of Science and Technology, EPA- 822-B-00-001, Washington, D.C., 238 pp.

USEPA (U.S. Environmental Protection Agency) and Tetra Tech, Inc., 2011. Spreadsheet Tool for the Estimation of Pollutant Load (STEPL). User's Guide, Version 4.1., Tetra Tech, Inc., Fairfax, Virginia, 57 pp.

V3 Companies, 2006. Lost River Watershed. Final Water Quality Monitoring Report Lost River-04010 September-2006, Woodridge, Illinois.

Veith, T.L., M.L. Wolfe, and C.D. Heatwole, 2003. Optimization Procedure for Cost Effective BMP Placement at a Watershed Scale. Journal of the American Water Resources Association 39 (6):1331-1343.

Wang, Q.J., 1991. The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resources Research 27(9):2467-2471.

Zhang, Y.K. and K.E. Schilling, 2006. Increasing Streamflow and Baseflow in the Mississippi River Since 1940: Effect of Land Use Change. Journal of Hydrology 324:412-422.