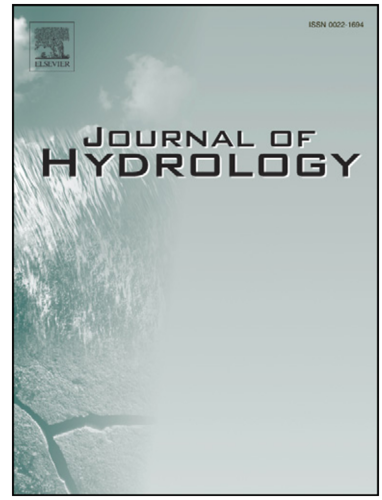
Accepted Manuscript



Estimation of annual baseflow at ungauged sites in Indiana USA

Laurent Ahiablame, Indrajeet Chaubey, Bernard Engel, Keith Cherkauer, Venkatesh Merwade

PII: S0022-1694(12)00876-1
DOI: http://dx.doi.org/10.1016/j.jhydrol.2012.10.002
Reference: HYDROL 18511

To appear in: *Journal of Hydrology*

Received Date: 6 June 2012
Revised Date: 24 September 2012
Accepted Date: 4 October 2012

Please cite this article as: Ahiablame, L., Chaubey, I., Engel, B., Cherkauer, K., Merwade, V., Estimation of annual baseflow at ungauged sites in Indiana USA, *Journal of Hydrology* (2012), doi: http://dx.doi.org/10.1016/j.jhydrol.2012.10.002

Estimation of annual baseflow at ungauged sites in Indiana USA

Laurent Ahiablame[1], Indrajeet Chaubey[1,2]

Bernard Engel[1], Keith Cherkauer[1], and Venkatesh Merwade[3]

[1]Department of Agricultural and Biological Engineering, [2]Department of Earth, Atmospheric, and Planetary Sciences, and Division of Environmental and Ecological Engineering, Purdue University, 225 South University Street, West Lafayette, IN 47907-2093, USA

[3]School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051

Corresponding author: ichaubey@purdue.edu

SUMMARY

This study developed regression models for predicting baseflow and baseflow index (BFI) (baseflow/total streamflow) at ungauged sites by using long-term baseflow data, land use, and watershed physiographic characteristics. Baseflow data were derived from daily streamflow records with a recursive digital filter method for baseflow separation for 22 watersheds in Indiana. Filtered average annual baseflow varies between 150 and 320 mm for the study watersheds. BFI varies between 0.40 to 0.88 with an average of 0.60, suggesting that 60% of long-term streamflow in Indiana is likely supported by ground water discharge and shallow subsurface flow. Three regression equations (one for baseflow and one for BFI) were developed and evaluated in the study watersheds. Results showed that the models reasonably estimate baseflow with relative errors (RE) ranging from 0 to 30%, except in one watershed where the RE was 50% during the validation period. These equations can be used to estimate baseflow and BFI at ungauged sites in Indiana. The present work would have implications for improving the capabilities of simple hydrologic/water quality models, and support planning and management of water resources. The methodology used in this study can be applied in other locations and regions.

1. Introduction

Records of streamflow consist of total flow which is the combination of direct runoff and baseflow. Baseflow or dry-weather flow refers to the portion of streamflow generated from groundwater and delayed shallow subsurface flow into the stream channel (Cherkauer and Ansari, 2005; Gebert et al., 2007; Santhi et al., 2008). Groundwater and surface water interact with each other such that changes in the amount or quality of one will generally affect the other

(Sophocleous, 2002). Effective management of water resources and water quality must consider interactions among surface and groundwater flows for determination of water availability and water use allocations, developing and improving management strategies for water supply systems and water quality, and understanding relationships between aquatic organisms and their environment (Reay et al., 1992; Stuckey, 2006; Santhi et al., 2008).

Hydrograph analysis studies to estimate baseflow from streamflow records have been conducted for many years (Nathan and McMahon, 1990; Arnold et al., 1995; Eckhardt, 2005; Lim et al., 2005). These studies have used graphical, analytical, and digital filtering techniques to partition baseflow from total streamflow (Lyne and Hollick, 1979; Nathan and McMahon, 1990; Szilagyi and Parlange, 1998; Arnold et al., 1995; Arnold and Allen, 1999). Although different in approach, all of these techniques share the same goal, which is to estimate baseflow under streamflow hydrograph, and provide methods for quantifying groundwater contribution to streamflow (e.g., Arnold et al., 1995; Eckhardt, 2005; Lim et al., 2005).

Estimation of baseflow using regression analysis is a process for transferring hydrologic information from gauged to ungauged watersheds (Gebert et al., 2007). Baseflow estimation at various spatial scales has previously attracted researchers in different parts of the world (Nathan and McMahon, 1990; Haberlandt et al., 2001; Mazvimavi et al., 2005; Longobardi and Villani, 2008; Santhi et al., 2008). The majority of these studies have made extensive use of statistical analysis to empirically relate baseflow to catchment characteristics (Table 1; Lacey and Grayson, 1998; Neff et al., 2005; Longobardi and Villani, 2008). The most common regression techniques reported in the literature include multiple linear regression and stepwise regression as shown in Table 1. For example, Santhi et al. (2008) used regression techniques to relate baseflow index to watershed relief and percentage of sand, and baseflow to watershed relief, precipitation, and potential evapotranspiration. Neff et al. (2005) observed spatial and geographic trends in baseflow within the Great Lakes region, suggesting that attention must be given to watershed specific properties when characterizing baseflow at different locations (Stuckey, 2006; Delin et al., 2007; Santhi et al., 2008). Baseflow is naturally influenced by a variety of watershed characteristics (Mazvimavi et al., 2005; Stuckey, 2006; Delin et al., 2007; Gebert et al., 2007; Santhi et al., 2008; Bloomfield et al., 2009; Zhu and Day, 2009). Examples of baseflow-related watershed characteristics, as well as methods used and study locations reported by previous studies are shown in Table 1.

In Indiana, Arihood and Glatfelter (1991) reported significant relationships between contributing drainage area, flow duration ratio and low flow in 82 watersheds using regression analysis. The authors showed that the equations predicted low flow with no bias in watersheds of various sizes, even in watersheds that were not used in model development. Tripathy (2007) reported an increasing trend in stream baseflow for a group of Indiana watersheds. Studies from other parts of the Midwestern United States also revealed increasing trends in baseflow over the last half century (Schilling and Libra, 2003; Zhang and Schilling, 2006). Kumar et al. (2009) recently assessed long-term flow trends in Indiana's streams and rivers. The authors highlighted the influence of tile drainage on increasing streamflow in the state, suggesting that attention must be given to local factors such as subsurface drainage in flow characteristic studies at watershed scales.

Most of Indiana is dominated by extensive networks of tile drainage systems, which are generally artificial channels more directly linking agricultural fields to naturally occurring streams or rivers than natural overland and subsurface flow pathways (Smith et al., 2010; Ahiablame et al., 2010; 2011). Construction of tile drainage systems in the region began in the second half of the 19th century (David et al. 1997; Richards et al. 2002). Many square kilometers of swamps and wetlands were artificially drained to provide highly fertile mollisols and alfisols (roughly 206, 390 km^2 of crop land) for agricultural production (Zucker and Brown, 1998; Richards et al. 2002; Schilling and Helmers, 2008). The estimated total land area artificially drained in the state was 44% in 1930, and 48% thirty years later (USDCBC, 1932-1961; Kumar et al., 2009). With more than 50% of subsurface tile drained agricultural lands (30,000 km^2), Indiana is portrayed as the second largest tile drained state in the Midwestern United States after Illinois (USDA-ERS, 1987). Up until recently, land drainage contributed to 87% of wetland loss (USDA-ERS, 1987).

In light of the above discussion, relationships between baseflow, local watershed characteristics, and landscape conditions including tile drainage merit to be investigated for Indiana. The goal of of this study was to relate baseflow to watershed characteristics. Specifically, this study developed and evaluated regression equations for baseflow prediction in ungauged watersheds. The methodology used in the present study can be applied in other locations, states, or regions to development baseflow equations for water resource planning and management.

2. Material and methods

2.1. Study area

The state of Indiana, located in the Midwestern U.S., covers a total area of 93,720 km^2 of which 57% is dedicated to agriculture (NALCC, 2002; Kumar et al., 2009). Indiana has a temperate and continental climate with warm summers and cold winters. Based on estimates from the National Climatic Data Center (http://www.ncdc.noaa.gov/oa/ncdc.html), daily air temperature varies from -10°C to -1°C, and -6°C to 4°C, respectively in the north and in the south for the coldest month of the year (typically, January). During the warmest month of the year (July), daily temperature ranges between 18°C and 29°C for the north, and between 21°C to 32°C in the south. The temperature in the far north is strongly influenced by water effects from Lake Michigan, giving a cool temperate climate in northern Indiana while a warmer temperate climate is noticeable in southern Indiana. Drastic variations in temperature are common and frequent throughout the year. Approximately 69% (710 mm) of average annual precipitation is returned to the atmosphere in the form of evapotranspiration (Fowler and Wilson, 1996).

Average annual precipitation varies from north to south between 890 and 1,100 mm (ISCO, 2011). Total annual snow depth also varies widely across the state with 2,000 mm in the north near Lake Michigan to 360 mm in the south (ISCO, 2011). The snow fall season varies from year to year beginning in late November and ending in early April (http://iclimate.org/narrative.asp). Some portions of the state are subject to flooding almost every year, especially between December and April (ISCO, 2011) due to increased runoff caused by frozen ground and low evapotranspiration (ET).

Prairies and wetlands emanating from the most recent ice age dominated the landscape of Indiana before settlement (Whitney, 1994; Kumar et al., 2009). There are three broad physiographic and geologic zones in Indiana: north, center, and south (Wayne, 1959; Schneider, 1966). The present landscape in the north is characterized by a pothole landscape and glaciated moraine, while most of the soils in the central portion of the state are compact soils of flat plains (Fowler and Wilson, 1996; Smith et al., 2010; ISCO, 2011). The topographic formation of the unglaciated south consists of hills, ridges, knolls, caves and waterfalls created through degradational processes such as weathering and stream erosion (Fowler and Wilson, 1996; ISCO, 2011). About 3/4 of the total land area of the state is used for agricultural activities, placing the state in the top 5 for corn and soybean production (NASDA, 2010). This rank would not be possible without exceptionally fertile farm fields brought into production through extensive drainage of wetlands throughout the state. Although Indiana's soils are highly productive, the majority of them is relatively flat (less than 2% slope) and poorly drained, resulting in frequent ponding (USDA, 2005).

Indiana's rivers are important parts of the natural features of the state. With a total drainage area of 85,000 km^2, the Wabash River (including the White River) is the state's longest river and drains the majority of Indiana's land surface (ISCO, 2011). Other river systems include the Maumee River in the far northeast, the St. Joseph (Lake Michigan) and Kankakee River (Illinois River) systems in the north central and northwest, and the Ohio River which drains only a small portion of the south and southeast.

2.2. Methodology of annual baseflow estimation and simulation

The methodology used to develop regression models for baseflow estimation in this study entails the following steps:

i. Development of a database to compile streamflow, climatic, and physical characteristics of watersheds: This step required various exploratory techniques to select gauging stations and potential explanatory variables to be used in the analysis.

ii. Baseflow separation and determination of baseflow index using a recursive digital filter technique for baseflow separation: This step involved partitioning of streamflow time series data into direct runoff and baseflow. The digital filter program for baseflow separation, herein referred to as BFLOW (Arnold and Allen, 1999), was used to estimate observed baseflow and baseflow index (BFI) for selected gauging stations. This program was selected for its widespread acceptance in the scientific community. A brief background description of the BFLOW program is provided in a section below. The baseflow computed with the BFLOW program is referred to as filtered baseflow in this study.

iii. Development of regression equations to estimate annual baseflow: This step used the multiple regression technique to generalize functional relationship between baseflow and watershed characteristics. The generalization of a mathematical relationship among variables is an advantage of multiple regression analysis compared to other statistical techniques such as ANOVA and ANCOVA which tend to be problem specific (McArthur and West, 1974). In addition, the multiple regression technique allows selection of a group of models for baseflow estimation rather than a single model which will be identified with stepwise regression. Previous studies show that, given a dataset, there is no single best model but a set of credible models with similar fit to the dataset (Box, 1979; Whittingham et al., 2006), leading to uncertainty analysis to measure the goodness of model results (Morgan and Henrion, 1990; Shirmohammadi et al., 2006). Whittingham et al. (2006) and Brodie and Dunn (2010) provided a detailed discussion of the limitations associated with stepwise regression analysis. Besides focusing on a single best model, other limitations of stepwise include bias in parameter estimation, inconsistencies among model selection algorithms, and intrinsic problems of multiple hypothesis testing (Whittingham et al., 2006).

iv. Validation of regression equations: Generally, hydrologic models are evaluated using split-sample methodology, investigation of errors, or by applying models at different locations (Koch and Smillie, 1986; McCuen, 2003; Gerbert et al., 2007; Zhu and Day, 2009). In this study, the reliability of the developed models was tested with a group of watersheds using different time periods and locations.

2.2.1. Database development

Daily streamflow data obtained from the USGS National Water Information System (NWIS) were used in this study. The USGS monitors a network of gauging stations and reports information on streamflow across the nation. The most important criteria used for selecting the gauging stations consisted of status of availability of long-term (more than 50 years) daily streamflow data availability, and regulatory status of streams draining into these stations. Many of Indiana's streams are subject to regulation or diversion that may induce biases in streamflow analysis. To avoid these biases, stations without regulation and diversion were selected using USGS annual water data reports (USGS, 2010). Thirty two gauging stations were identified as good candidates for the analysis. From this list, 10 stations were removed due to the fact that either all watershed characteristics were not available or they drain karst landscapes (common in southern Indiana). The karst landscapes were determined with spatial data available at IndianaMap.org (http://www.indianamap.org).

Fifty-seven years of streamflow data (1954-2010) for the remaining 22 gauging stations were used in the analysis (Table 2; Fig. 1). Thirty years of data were used for model development (as discussed in a section below) and the remaining 27 years were only included in model validation. Due to the limited number of watersheds meeting watershed selection criteria, the 30-year time frame was deemed reasonable to avoid short-term changes in precipitation, so that the remaining streamflow records can be used for validation while accounting for the long-term impacts of basin characteristics and climate on streamflow. A total of 18 watershed characteristics or metrics (Table 3) were evaluated as potential explanatory variables for use in developing the regression equations based on easy availability of the data for practical applications and a review of similar studies (e.g., Flynn and Tasker, 2004; Mazvimavi et al., 2005; Nejadhashemi et al., 2008; Price, 2011).

Precipitation data (downloaded on 01.22.2011) were extracted from digital gridded files of the PRISM (Parameter-elevation Regressions on Independent Slopes Model) climate mapping system (http://www.prism.oregonstate.edu) using the spatial extent of each delineated watershed. The PRISM datasets (Daly et al., 2000) are credited worldwide as high quality spatial climate data. The gridded datasets are produced and distributed by the PRISM Climate Group of Oregon State University at a spatial resolution of 4 km. The evapotranspiration (ET) data is also a digital gridded dataset generated as output from the Variable Infiltration Capacity (VIC) model at a spatial resolution of 1/8 degree as part of the Land Data Assimilation Project (LDAS) (Maurer et al., 2002).

Watershed characteristics pertaining to physical geography (Table 3) were derived from the Digital Elevation Model (Indiana Geological Survey, 2001, 30 meter DEM), the National Land Cover Data (NLCD, 2001), the digital soil of the Soil Survey Geographic database (SSURGO, 2001), and the National Hydrographic Dataset (High Resolution Flowlines, NHD, 2001). The DEM, NLCD and NHD were downloaded from the USGS database (http://seamless.usgs.gov; http://nhd.usgs.gov). The SSURGO data is produced and distributed by the USDA-Natural Resources Conservation Services (http://soils.usda.gov/survey/geography). The tile drained area for Indiana was estimated based on a modeling study conducted by Ale and Bowling (2010). ArcGIS software (version 9.3) and ArcHydro tools were used to create shape and raster files in order to extract the watershed characteristics (Table 3) needed for the analysis.

2.2.2. Baseflow separation

Baseflow was determined for the 22 gauging stations using the BFLOW program (Arnold et al., 1995). The BFLOW program has been widely used in baseflow separation studies (Lim et al., 2005; Eckhardt, 2008). The program is founded on the premise that streamflow has two components which are direct runoff and baseflow (Streamflow = Direct Runoff + Baseflow). Thus, streamflow time series data can be partitioned into these two components, similarly to the analysis of high and low frequency signals with a recursive filter technique initially proposed by Lyne and Hollick (1979). Baseflow represents, in this case, low frequency signals, whereas direct runoff can be considered as high frequency signals (Arnold et al., 1995; Eckhardt, 2008). The Lyne-Hollick equation described by Nathan and McMahon (1990), and Arnold et al. (1995) can be expressed in terms of baseflow filtering as (Eckhardt, 2008):

$$b_t = ab_{t-1} + \frac{1-a}{2}\left(Q_t + Q_{t-1}\right) \qquad (1)$$

where, b is the baseflow; a is the recession constant; Q is the total streamflow; and t is the time step number. This equation is restricted with the condition that $b_t \leq Q_t$ (Eckhardt, 2008).

The BFLOW program computes baseflow by passing the filter over streamflow data three times consisting of 1-Pass, 2-Pass, and 3-Pass (i.e., forward, backward, and forward). Each pass leads to a reduction in baseflow as a percent of streamflow. Arnold et al. (1995) provided a detailed description of the percent reduction for each pass. The authors reported that annual filtered baseflow with 1-Pass is consistent with baseflow estimated with manual and graphical techniques within ±11%. Although the BFLOW program combines two techniques, one to separate baseflow based on the Lyne-Hollick recursive filter technique, and the other one to estimate streamflow recession constant (RC) using a matching strip approach (Arnold et al., 1995), only BFI values were reported in this study because the focus of this study was to develop equations for baseflow and BFI estimation. It should be noted that the digital filter approach for baseflow separation, as implemented in the BFLOW program, uses a single value of 0.925 for the RC (Arnold and Allen, 1999; Eckhardt, 2008). Following Eckhardt (2008), the baseflow and BFI values generated with 1-Pass were used for the regression analysis described in the next section. Annual baseflow was calculated in cubic meters per second and normalized by unit area per year to allow comparison across watersheds.

2.2.3. Regression analysis

Multiple regression analysis was utilized to develop equations for annual baseflow and BFI prediction at ungauged sites using physical and climatic characteristics of the study watersheds. It should be recalled that generalizing relations between baseflow and watershed characteristics with multiple regression may lead to a group of models with similar fit to the dataset, not a single best model as would be the case with the use of stepwise regression. Baseflow calculated for the period of 1974-2003 in 18 watersheds was used for model development (Fig. 1; Table 2). The remaining four watersheds were not used for model development due to the relative inconsistency of the delineated area for these watersheds compared to the areas published by the USGS (2-19% error) (see Table 2 for these watersheds). These remaining four watersheds were included in the validation process to evaluate the performance of the developed models for different spatial conditions over the period of 1974-2003. Additionally, the models were validated in all 22 watersheds for two different periods: 1954-1973 and 2004-2010. This allowed assessment of the accuracy of the models outside the flow period used for model development.

Regression equations relate watershed characteristics to baseflow and BFI. To select watershed characteristics that are statistically significant, the Spearman's rank correlation test was used. The p-values generated from this test helped identify the watershed characteristics that have the potential to explain variability in baseflow and BFI (Table 4). The correlation test also provided a mechanism for whether or not these watershed characteristics were statistically independent from each other (Table 4). After this initial screening, three additional steps were utilized to further refine the selection of explanatory variables. First, judgment, drawing from the literature, was used to select supplementary variables that would normally influence baseflow but were not significant with the test (e.g., precipitation, percent tile drained area). Second, the best 20 fitted models were selected with an option of BEST = 20 in SAS "proc reg" procedure (SAS Institute Inc., 2010). The BEST = 20 requests SAS to output the best 20 models with the highest R^2 and adjusted R^2 values based on different sets of combinations of independent variables. An inspection of adjusted R^2 values of these best 20 models was then used to identify models with simple independent variables that have similar high level of good fit (i.e. $R^2 > 0.70$). Third, individual p-values of independent variables in the remaining models were inspected for significance. The significant independent variables were used in a heuristic process (i.e. trial and error) to assess their importance in the models, translated by variations in adjusted R^2 values, when present or not present in the models. The final independent variables for the equations that would best predict baseflow and BFI in Indiana were selected at the end of this third step. All analyses were conducted with the Statistical Analysis System, version 9.2 (SAS Institute Inc., 2010), at the 5% significance level. The regression equations for baseflow are expressed as:

$$Q_b = b_0 X_1^{b_1} X_2^{b_2} X_3^{b_3} \ldots X_n^{b_n} \qquad (2)$$

where, Q_b is the annual baseflow (m^3 per year); b_0 is the regression constant, b_1, …, b_n are regression coefficients; and X_1, X_2, …, X_n are watershed characteristics. The regression equations were developed by applying a logarithmic transformation to all variables to meet normality requirements, and the data were analyzed as ordinary linear regression in the form of:

$$\log(Q_b) = \log(b_0) + b_1 \log(X_1) + b_2 \log(X_2) + b_3 \log(X_3)... + b_n (\log X_n) \quad (3)$$

The final regression equations were reported in two forms using equations 2 and 3. Equation 2 is a retransformation of equation 3 with the inverse of the logarithmic. Similarly, an equation to predict BFI was developed as function of watershed characteristics using the format of equations 2 and 3.

The predictive capacity of the models was assessed with relative error (RE) and Nash-Sutcliffe efficiency (E_{NS}; Nash and Sutcliffe, 1970), respectively, calculated as:

$$RE = \frac{Q_{b(predicted)} - Q_{b(filtered)}}{Q_{b(filtered)}} \times 100 \quad (4)$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{n} \left(Q_{b_i(filtered)} - Q_{b_i(predicted)} \right)^2}{\sum_{i=1}^{n} \left(Q_{b_i(filtered)} - \overline{Q}_{b_i(filtered)} \right)^2} \right] \quad (5)$$

where is $Q_{b_i\,(filtered)}$ the filtered (observed) annual baseflow; $Q_{b_i\,(predicted)}$ is the simulated annual baseflow; $\overline{Q}_{b_i\,(filtered)}$ is the average annual filtered baseflow during the period of interest; and n is the total number of years. Guided by easy accessibility and availability of independent variables, two equations for annual baseflow and one equation for BFI estimation were retained to be evaluated. The equation retained for BFI prediction was used to estimate BFI values for the two baseflow prediction models.

3. Results and discussion

3.1. Filtered (observed) baseflow

Filtered average annual baseflow of the 1974-2003 period varied between 150 mm per year for the Mississinewa River near Ridgeville and 320 mm per year for the Kankakee River at Porter (Fig. 1; Tables 2 and 5), with an average of 220 mm per year for the 22 watersheds. The average annual total streamflow for the 22 watersheds during the same period was 380 mm per year. These baseflows are comparable to published studies in the Midwest which reported a range of 80 to 200 mm per year in Iowa streams (schilling and Helmers, 2008). The filtered baseflow tends to be high in larger watersheds (Table 5). This could be due to the fact that the BFLOW algorithm estimates baseflow with considerable smoothing (Eckhardt, 2008). Without explicitly taking drainage area into account, BFLOW is more likely to overestimate baseflow in large watersheds. For the 30 years of records, average baseflow appears to be slightly larger in the northern part of the state (Table 3; Fig. 1). Although the difference is not significant, Fowler and Wilson (1996) argued that stream channels in the glaciated north are highly influenced by ground water discharge than those in the south, allowing more sustained flows. The presence of glacial features such as moraines and morainal lakes, kames, eskers, melt-water channels, and ice-block depression lakes in northern Indiana play an important role in ground water discharge to streamflow (Arihood and Glatfelter, 1991). The central Indiana landscape consists of nearly uniform flat plains with lower sustained flows (Arihood and Glatfelter, 1991). Although precipitation and geology have been recognized to influence streamflow in Indiana, the south which receives more precipitation with higher temperatures and resultant ET, tends to have lower sustained flows than the north (Fowler and Wilson, 1996). This suggests that spatial variation of baseflow, when moving from the north to the south, could be the result of other factors beyond landscape, geology and precipitation patterns. In a study conducted over 1970-2000, Tripathy (2007) demonstrated that the increasing baseflow trend observed in Indiana is strongly influenced by tile drainage in rural watersheds, a combination of leakage from sewers, water supply lines, stormwater management facilities, groundwater withdrawal, and water release to surface water after human uses. Recently, Kumar at al. (2009) also reported that increased streamflow trends in Indiana were related to factors such as subsurface tile drainage. The analysis conducted in the present study concurred with these previous findings. Although geomorphological differences exist between the regions of the state (Wilkerson and Merwade, 2010), annual baseflow across the state lie within the range of 150 to 320 mm per year in the 22 watersheds without any particular trend respective to the regions (Fig. 2), suggesting that these differences were not noticeably reflected in baseflow.

BFI values for the study watersheds range from 0.40 to 0.88 with an average of 0.60 (Table 5), suggesting that on average 60% of long-term streamflow in Indiana is assumed to be coming from groundwater discharge and shallow subsurface flow. The logic of this assumption resides in the fact that baseflow is generally linked to discharge from groundwater storage (Eckhardt, 2008) and soil permeability (Santhi et al., 2008). As mentioned earlier in “Baseflow separation”, analysis of streamflow time series data can be compared, by analogy, to signal analysis with low frequency variability associated with baseflow, and high frequency variability, being the effect of direct runoff (Arnold et al., 1995; Eckhardt, 2008). Therefore, baseflow can be isolated by low-pass filtering streamflow hydrograph (Eckhardt, 2008).

Gebert et al. (2007) reported similar baseflow proportion for Wisconsin, where the authors showed that baseflow was 62% of annual total at gauging stations. BFI values for Indiana are also consistent with the range of 60-80% of streamflow as baseflow (BFI) reported by Schilling and Helmers (2008) for Iowa streams. A close look at BFI values in the northern, central, and southern regions of Indiana, revealed that the northern heavily drained region (Kumar et al., 2009) tends to have slightly higher BFI values compared to the remainder of Indiana, indicating that increasing baseflow may be affected by water storage (created by tile drainage or natural or the combination of both). This also weakens the theory of direct dependence of baseflow on geology and precipitation. While the quick removal of water from the fields can provide temporary storage in the soil profile, the water added in the ditches can result in increased streamflow as discussed earlier. In agricultural settings, tile drainage networks act as flow conduit systems that may lead to high BFI values due to increase in baseflow. However, further research is needed to investigate this hypothesis, which is beyond the scope of this study.

3.2. Regression equations

3.2.1. Model application: 1974-2003

The regression analysis resulted in the development two regression equations for baseflow (Models A and B), and one equation for BFI estimation with watershed characteristics (Table 6). It should be noted that watershed characteristics that vary substantially on annual basis include precipitation and ET (Table 3). The watershed characteristics that influence variation in baseflow in Indiana include drainage area, percent of tile drained area, precipitation, and the long-term ratio of baseflow to total flow (i.e. BFI) (Table 6; $p < 0.0001$). Except watershed drainage area, these variables were not statistically significant based on the initial screening of the independent variables with correlation test (Table 4). The non-significant baseflow predictive capability of these variables could simply be the fact that their significance was inhibited by the presence of the other predictors. Therefore, additional steps were completed as described in the Regression analysis section above to identify several models with the "BEST = 20" procedure. This procedure examines several possible combinations of independent variables that provide models of best fit to the data. For example, precipitation was not highly correlated with baseflow with the initial correlation test (Table 4). However, the continuation of the regression analysis showed that variation in baseflow was explained by variation in precipitation ($p < 0.0006$). The non-significance correlation between precipitation and baseflow could also probably be a delayed response to precipitation, especially at the watershed scale. In addition, these variables were previously used to predict baseflow with high accuracy in other studies (e.g., Gerbert et al., 2007; Zhu and Day, 2009). Gerbert et al. (2007) developed baseflow models, in which drainage area, soil infiltration, and baseflow were significant factors (referred to as BFI in this study). For Zhu and Day (2009), the statistically significant independent variables for their baseflow equation include drainage area, precipitation, evapotranspiration, and elevation.

The variables that best predict BFI in Indiana include percent surface water in the form of land use, and percent hydrologic soils B and C in the study watersheds, ($p < 0.0001$). Surface water is generally stored in low relief areas within a watershed, and permeability is controlled by soil texture (fine to coarse). In Indiana, surface water storage may be linked to tile drainage, which contributes to baseflow. Beside the apparent influence of climate on baseflow, watershed geological characteristics may also influence baseflow. The influence of hydrologic soils on BFI has a geological meaning because the formation of different types of rocks results in the formation of different types of soils (Bloomfield et al., 2009). Soils with different infiltration capacities will influence baseflow differently (Lacey and Grayson, 1998). Santhi et al. (2008) found a high correlation between BFI and watershed relief and percentage of sand for the conterminous United States using regression analysis. Longobardi and Villani (2008) also reported relationships between BFI, elevation and permeability index, which was calculated as the ratio of permeable area to the total area of the watershed (calcareous and dolomitic complex areas).

The two baseflow models were applied to the 22 watersheds, of which 18 were used in model development for the study period of 1974 to 2003 (Fig. 1 and Table 2). The performance of the regression equation developed to predict BFI values in ungauged sites in Indiana is shown in Fig. 3. The data points are evenly scattered around the 1:1 line, and the filtered and predicted BFI values are found to be significantly related to each other ($p < 0.0001$). To screen the accuracy of the models before using them with predicted BFI values, they were first applied to the study watersheds with calculated BFI values (BFLOW BFI values). The logic of using calculated BFI values was to demonstrate that the models can be used to predict baseflow at ungauged sites with calculated BFI values at gauged sites which have similar conditions as the ungauged sites. Results hold promise as shown in Fig. 4 and Table 7 for simulations with calculated BFI values and simulations with predicted BFI values. The developed models predicted baseflow with minimal relative error (Table 7; Fig. 4), except in 03342500 (Busseron Creek near Carlisle) and 04099510 (Pigeon Creek near Angola) that have relatively higher relative error compared to the others (Table 7). While there is no particular pattern in the degree of accuracy of the models in terms of watershed characteristics, model B performed better than model A, indicating that tile drainage is an important factor that should be taken into consideration for streamflow analysis in Indiana (Table 7).

3.2.2. Model validation: 1954-1973 and 2004-2010

The regression models were validated for two periods: 1954 to 1973 and 2004 to 2010. The models were validated with simulated BFI values during the 1954-1973 period in 22 watersheds consisting of the 19 watersheds used for model development and the remaining three of the 22 watersheds (Fig. 1; Table 2). Variations in predicted baseflow could be the result of changes in precipitation pattern reported for the Midwestern region over the past few decades (Mishra et al., 2010). A two sample t-test to compare precipitation during the study period and the validation period revealed that average annual precipitation significantly increased in Indiana during the study period ($p = 0.0101$). Kumar et al. (2009) also observed an increasing trend in precipitation at many locations in Indiana, due especially to increased summer rainfall. In addition, Indiana has experienced increased land transformation with development of more agricultural and urban areas. Palmer and Ottensmann (2003), for example, reported an increasing trend in urban growth in the central part of the state, which is expected to double in urban areas by 2040. Agricultural and urban expansion resulted in losses of hardwood forests, placing Indiana 48th out of the 50 states for the amount of remaining natural vegetation (Santelmann et al., 2004; Rizkalla and Swihart, 2009). Land use change affects ecohydrological responses (Rizkalla and Swihart, 2009), including precipitation and streamflow. The combination of these effects could limit the ability of the two models to predict baseflow for the past (e.g., 1954-1973). The use of land cover information of different time frame in future research may provide more understanding of this behavior.

A close investigation of different sets of BFI values (filtered versus predicted) revealed that the lower the BFI values, the better the performance of the models. It should be recalled that BFI is calculated as the long-term ratio of baseflow to total flow. The accuracy of the baseflow models rely on the accuracy of BFI values. While BFI cannot be measured directly in the field, it is a sensitive parameter which exerts a strong influence on baseflow (Eckhardt, 2005). Previous studies showed that BFI is related to climate, topography, vegetation, soil types, and geology within a watershed, and constitutes the most dominant low flow indicator (Vogel and Kroll, 1992; Lacey and Grayson, 1998; Haberlandt et al., 2001; Longobardi and Villani, 2008).

Although there is a slight prediction difference between the two annual models (models A and B), they reasonably predicted baseflow during the validation period (Figs. 5 and 6; Table 8), except in the Busseron Creek near Carlisle and the Pigeon Creek near Angola) that show large discrepancies (Figs. 5 and 6; Table 8). The presence of floodwater retarding structures and the practice of surface mining in Busseron Creek have been reported to influence natural water flow and could potentially affect baseflow (USGS, 2010). Busseron Creek is sometimes affected by backwater from the Wabash River (USGS, 2010). Pigeon Creek also receives discharge from four main outlets, which include Jackson Ditch, John Leach Drain, a wastewater treatment plant outfall, and a storm sewer outfall from the city of Angola (V3 Companies and SCSWCD, 2006). The discharge from these outlets, especially the wastewater treatment plant and the storm sewer system, would influence streamflow, driving the differences between simulated and filtered baseflows. Although natural flow in these two watersheds is impacted by anthropogenic activities, they were intentionally included in the spatial validation process to assess the performance of the models in watersheds with similar problems.

Table 8 shows the performance of the models with respect to R^2 and E_{NS} which range from 0.04 to 0.87, and -8.47 to 0.71, respectively, for model A and model B. While these two statistics have gained widespread use for model evaluation, the use of only one of them to evaluate the accuracy of a model is not recommended (Jain and Sudheer, 2008). Values greater than 0.5 for the R^2, and values between 0 to 1 for the E_{NS} are usually considered adequate in terms of model performance (Santhi et al., 2001; Moriasi et al., 2007). The RE varies between 0 and 53% for the two models. Based on these statistics along with the RE, it appears that the models show satisfactory predictions of baseflow in most of the 22 watersheds. This performance can be considered good, given the study assumptions (see limitations of models section below), the increasing statewide anthropogenic activities and their effects on natural water systems (Palmer and Ottensmann 2003; Yang et al., 2011), and the changing climatic characteristics reported for Indiana (Tripathy, 2007; Kumar et al., 2009). There is also a consistent pattern across the three statistics (R^2, E_{NS}, RE; Table 8), suggesting that discrepancies between predicted and filtered baseflow are not related to the models suitability but may be attributed to other factors as discussed above for the Busseron Creek near Carlisle and Pigeon Creek near Angola for example.

Results for model evaluation during the 2004-2010 validation period are shown in Figure 7. For the two models, the data points are spread along the 1:1 line, indicating that the regression equations yield satisfactory results in predicting baseflow (Fig. 7). The data points representing the Busseron Creek near Carlisle and Pigeon Creek near Angola are clearly outliers (Fig. 7). Streamflow in these 2 watersheds is impacted by anthropogenic activities as discussed for the first validation period (1954-1973). Although the models performed well, the majority of the data points fell below the 1:1 line (Fig. 7), suggesting that the models display a slight negative bias. The models over predict baseflow before the 1973-2003 period used for model development, while the evaluation of the models after the 1973-2003 period reveals a slight tendency to under predict baseflow. This could be the result of the changes in land use/land cover, streamflow pattern, and climate in this period with respect to the period for which the equations were developed. With BFI values that correctly represent hydrological and/or hydrogeological characteristics of the watersheds, these regression models are viable options for baseflow prediction in Indiana and other locations with similar landscape and climate characteristics.

3.2.3. Model limitations

The use of the regression equations developed in this study is limited by a range of conditions over which input data were collected. During the analysis, watershed characteristics, including land uses were assumed constant over the study period (1974-2003), while changes in land use/land cover were reported for Indiana over the past several decades (Palmer and Ottensmann 2003; Rizkalla and Swihart, 2009). Indiana has three physiographic regions with varying hydrologic responses (Fowler and Wilson, 1996). Developing regression equations for each of these regions could improve accuracy for baseflow estimation within these physiographic regions in accordance to Smakhtin (2001), who reported some advantages of developing distinct regression equations for separate regions within study areas. However, developing several separate equations would likely result in computational and resource-intensive tasks, especially, for estimation of baseflow over large areas including two or more physiographic regions. The focus of this study was to provide a general representation of baseflow with watershed characteristics across landscape variation at the local level. The statewide model has the advantage to generalize baseflow and BFI characteristics over the region.

Another limitation is the uncertainty in the choice of the baseflow separation technique. There is no standard in choosing a particular hydrograph separation technique for developing regression equations and/or for comparing simulated to filtered baseflows over another technique. There is no roadmap to follow when selecting the best baseflow separation technique in a specific region since there is no baseflow observational data supporting any choice. The BFLOW program was used in this study because results from the BFLOW program were used to validate other baseflow separation techniques for Indiana (Lim et al., 2005). The outcomes from the present study may somewhat be different outcomes if other baseflow separation techniques were used.

Although the majority of the independent variables in the developed models are easily accessible, available, and applicable to other locations, caution should be observed in the use of these equations in entirely different landscape characteristic regions. These models were developed with specific topographic, geologic, soil, vegetation and climatic attributes.

4. Summary and Conclusions

This study developed equations for annual baseflow estimation for Indiana using regression analysis. Twenty-two gauged watersheds were delineated and their physical and climatic characteristics compiled. Thirty years of streamflow records were used to compute baseflow in these watersheds. Watershed characteristics and baseflow records were then utilized to develop equations for baseflow and baseflow index estimation in Indiana ungauged watersheds. The methodology used for model development and evaluation in the present study can be used in other states and regions for baseflow estimation in ungauged sites. Important conclusions from this study include:

- Annual baseflow ranges from 150 to 320 mm per year in all watersheds. On average 60% of long-term streamflow in Indiana is likely supported by ground water discharge and shallow subsurface flow, and the remaining 40% would be the contribution of direct runoff. If 60% of streamflow in Indiana is sustained by groundwater discharge, more attention should be given to understanding pollutant transport in baseflow for improving management of water resources and water quality.

- The independent variables influential in baseflow prediction consist of watershed drainage area, precipitation, percent tile-drained area in the watershed, and baseflow index. Guided by the simplicity, easy accessibility, and availability of the independent variables, the model with drainage area, precipitation, and baseflow index (model A) is recommended. The independent variables in the baseflow index model include percent surface water, and watershed percent hydrologic soils B and C.
- Validation of the baseflow equations during two time periods indicated that the models can be used to accurately estimate baseflow at ungaged sites. The accuracy of the predictive capacity of the models will likely depend on the annual precipitation and the accuracy of baseflow index in the study area.
- Limitations should be considered when using these models. This study used streamflow data from Indiana to develop a methodology for baseflow estimation at ungauged sites. Although the independent variables that explain variability in baseflow are easily accessible for almost any location, equations in geographic regions different from Indiana conditions should be validated and modified, if needed, before making baseflow predictions.
- The regression models developed in this study can be implemented in simple hydrologic/water quality models to expand the capability of these models for assessing the benefits of best management practices on runoff, baseflow, and streamflow at various scales in the context of water resources planning and management.

Acknowledgements

This study was partially funded by the Illinois-Indiana Sea Grant, College Program Research and Outreach Development and Capacity Building Projects, the U.S. Environmental Protection Agency (USEPA)-Great Lakes Restoration Initiative program, and the USEPA-Region 5 Water Division. The authors would like to thank Dr. Kossi Nouwakpo, Larry Theller, Dr. Sanjiv Kumar, Dr. Laura Bowling, Cibin Raj, and Lili Wang with their help in data compilation and processing during this study. Detailed comments provided by two reviewers greatly improved an earlier version of this manuscript.

References

Ahiablame, L., Chaubey, I., Smith, D.R., 2010. Nutrient content at the sediment-water interface of tile-fed agricultural drainage ditches. Water 2, 411-428.

Ahiablame, L., Chaubey, I., Smith, D.R., Engel, B.A., 2011. Effect of tile effluent on nutrient concentration and retention efficiency in agricultural drainage ditches. Agric. Water Manage. 98, 1271-1279.

Arihood, L.D., Glatfelter, D.R., 1991. Method for estimating low-flow characteristics of ungaged streams in Indiana. United States Geological Survey Water-Supply Paper 2372. Prepared in cooperation with the Indiana State Board of Health.

Ale, S., Bowling, L.C., 2010. Estimating potentially subsurface drained areas in Indiana and their influence on streamflow pattern. National Water Conference. Hilton Head, SC, February 21-25.

Arnold, J.G., Allen, P.M., Muttiah, R., Bernhardt, G., 1995. Automated baseflow separation and recession analysis techniques. Ground Water 33, 1010-1018.

Arnold, J.G., Allen, P.M., 1999. Automated methods for estimating baseflow and ground water recharge from streamflow records. J. Am. Water Resour. Assoc. 35 (2), 411-424.

Bloomfield, J.P., Allen, D.J., Griffiths, K.J., 2009. Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. J. Hydrol. 373, 164-176.

Box, G.E.P., 1979. Robustness in the strategy of scientific model building, in: Launer, R.L., Wilkinson, G.N. (Eds), Robustness in statistics. Academic Press, San Diego, CA, p. 202.

Brodie, I.M., Dunn, P.K., 2010. Commonality of rainfall variables influencing suspended solids concentrations in storm runoff from three different urban impervious surfaces. J. Hydrol. 387, 202-211.

Cherkauer, D.S., Ansari, S.A., 2005. Estimating ground water recharge from topography, hydrogeology, and land cover. Ground Water 43 (1), 102-112.

Daly, C., Taylor, G.H., Gibson, W.P., Parzybok, T.W., Johnson, G.L., Pasteris, P.A., 2000. High-quality spatial climate data sets for the United States and beyond. Trans. ASAE 43(6), 1957-1962.

David, M. B., Gentry, L.E., Kovacic, D.A., 1997. Nitrogen balance in and export from an agricultural watershed. J. Environ. Qual. 26, 1038-1048.

Delin,G.N., Healy, R.W., Lorenz, D.L., Nimmo, J.R., 2007. Comparison of local to regional scale estimates of groundwater recharge in Minnesota, USA. J. Hydrol. 334, 231-249.

Eckhardt, K., 2005. How to construct recursive digital filters for baseflow separation. Hydrol. Proc. 19 (2), 507-515.

Fowler, K.K., Wilson, J.T., 1996. Low flow characteristics of Indiana streams. USGS Water Resources Investigations Report 96-4128.

Fynn, R.H., Tasker, G.D., 2004. Generalized estimates from streamflow data of annual and seasonal groundwater-recharge rate from drainage basins in New Hampshire. USGS Scientific Investigations Report 5019-5090.

Gebert, W.A., Radloff, M.J., Considine, E.J., Kennedy, J.L., 2007. Use of streamflow data to estimate baseflow/ground-water recharge for Wisconsin. J. Am. Water Resour. Assoc. 43(1), 220-236.

Haberlandt, U, Klöcking, B., Krysanova, V., Becker, A., 2001. Regionalization of the baseflow index from dynamically simulated flow components: a case study in the Elbe River Basin. J. Hydrol. 248, 35-53.

ISCO (Indiana State Climate Office), 2011. Indiana Climate Atlas. <http://www.iclimate.org/facts.asp> (04.25.11).

Jain, S.K., Sudheer, K.P., 2008. Fitting of hydrologic models: a close look at the Nash-Sutcliffe Index. J. Hydrol. Eng. 13(10), 981-986.

Koch, R.W., Smillie, G.M., 1986. Bias in hydrologic prediction using log-transformed regression models. J. Am. Water Resour. Assoc. 22 (5), 717-723.

Kumar, S., Merwade, V., Kam, J. Thurner, K., 2009. Streamflow trends in Indiana: Effects of long term persistence, precipitation and subsurface drains. J. Hydrol. 374, 171-183.

Lacey, G.C., Grayson, R.B., 1998. Relating base flow to catchment properties in Southeastern Australia. J. Hydrol. 204, 231-250.

Lim, K.J., Engel, B.A., Tang, Z., Choi, J., Kim, K., Muthukrishnan, S., Tripathy, D., 2005. Automated web GIS-based hydrograph analysis tool, WHAT. J. Am. Water Resour. Assoc. 41 (6), 1407-1416.

Lyne, V.D., Hollick, M., 1979. Stochastic time-variable rainfall runoff modelling. Hydrology and Water Resources Symposium. Institution of Engineers, Australia, Perth, pp. 89-92.

Longobardi, A., Villani., P., 2008. Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: role of the catchment permeability index. J. Hydrol. 355, 63-75.

Mazvimavi, D., Meijerink, A.M.J., Savenije, H.H. G., Stein, A., 2005. Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe. Phys. Chem. Earth 30, 639-647.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., Nijssen, B., 2002. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. J. Climate 15, 3237-3251.

McArthur, J.T., West, L.H.T., 1974. Multiple regression as a general data analysis technique. Res. Sci. Educ. 4,185-192.

McCuen, R.H., 2003. Modeling hydrologic change: statistical methods. Lewis Publishers, Boca Raton, FL 433, pp.

Meyer, S.C., 2005. Analysis of base flow trends in urban streams, northeastern Illinois, USA. Hydrogeology J.13,871-885.

Mishra, V., Cherkauer, K.A., Niyogi, D., Lei, M., Pijanowski, B.C., Ray, D.K., Bowling, L.C., Yang, G., 2010. A regional scale assessment of land use/land cover and climatic changes on water and energy cycle in the upper Midwest United States. Int. J. Climate 30, 2025-2044.

Morgan, M., G., Henrion, M., 1990. Uncertainty : A Guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, New York.

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. Trans. ASABE 50 (3), 885-900.

NALCC (North American Land Cover Characteristics), 2002. 1-Kilometer Resolution. National Atlas of the United States. <http://seamless.usgs.gov> (04.25.11).

Nathan, R. J., McMahon, T.A., 1990. Evaluation of automated techniques for baseflow and recession analysis. Water Resour. Res. 26, 1465-1473.

Nash, J. E., Sutcliffe, and J. V., 1970. River flow forecasting through conceptual models: part 1. A discussion of principles. J. Hydrol. 10 (3), 282-290.

NASDA (National Association of State Departments of Agriculture), 2010. Indiana State Department of Agriculture. < www.nasda.org> (04.25.11).

Neff, B.P., Day, S.M., Piggott, A.R., Fuller, L.M., 2005. Baseflow in the Great Lakes Basin. USGS Scientific Investigation Report 2005-5217.

Nejadhashemi, A. P., Shirmohammadi, A., Montas, H.J., Sheridan, J.M., Bosch, D.D., 2008. Watershed physical and hydrological effects on baseflow separation. J. Hydrol. Eng. 13(10), 971-980.

Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. Progr. Phys. Geogr. 35(4), 465-492.

Palmer J., Ottensmann J., 2003. New model predicts growth patterns in central Indiana. Center for Urban Policy and the Environment. Indiana University-Purdue University Indianapolis. <http://www.urbancenter.iupui.edu/PubResources/pdf/37_LUCI_rev.pdf>. (11.02.10).

Reay, W.G., Gallagher, D.L., Simmons Jr, G.M., 1992. Groundwater discharge and its impact on surface water quality in a Chesapeake Bay inlet. Ground Water 28(6),1121-1133.

Richards, R.P., Calhoun, F.G., Matisoff, G., 2002. The Lake Erie agricultural systems for environmental quality project: an introduction. J. Environ. Qual. 31(1), 6-16.

Rizkalla, C.E., Swihart, R.K., 2009. Forecasting the effects of land-use change on forest rodents in Indiana. Environ. Manage. 44, 899-908.

Santelmann, M.V., White, D., Freemark, K., Nassauer, J.I., Eilers, J.M., Vache, K.B., Danielson, B.J., Corry, R.C., Clark, M.E., Polasky, S., Cruse, R.M., Sifneos, J., Rustigian, H., Coiner, C., Wu, J., Debinski, D., 2004. Assessing alternative futures for agriculture in Iowa, USA. Landscape Ecol. 19, 357-374.

Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Srinivasan, R., Hauck, L.M., 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. J. Am. Water Resour. Assoc. 37(5), 1169-1188.

Santhi, C., Allen, P.M., Muttiah, R.S., Arnold, J.G., Tuppad, P., 2008. Regional estimation of baseflow for the conterminous United States by hydrologic landscape regions. J. Hydrol. 351,139-153.

SAS Institute, Inc., 2010. Version 9.2 for Windows. SAS, Cary, NC.

Schilling, K.E., Libra, R.D., 2003. Increased baseflow in Iowa over the second half of the 20th century. J. Am. Water Resour. Assoc. 39, 851-860.

Schilling, K.E., Helmers, M., 2008. Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: Exploratory hydrograph analysis. J. Soil Water Cons. 62(4), 252- 261.

Schneider, A.F., 1966. Physiography, in: Lindsey, A.A. (Eds), Natural features of Indiana. Indiana Academy of Science, Indianapolis, pp. 40-56.

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

Smakhtin, V.Y., 2001. Low flow hydrology: a review: J. Hydrol. 240, 147-186.

Smith, D.R., Huang, C., 2010. Assessing nutrient transport following dredging of agricultural drainage ditches. Trans. ASABE 52 (3), 429-436.

Sophocleous, M., 2002. Interactions between groundwater and surface water: the state of the science. Hydrogeology J. 10, 52-67.

Stuckey, M.H., 2006. Lowflow, baseflow, and mean flow regression equations for Pennsylvania streams. USGS Scientific Investigations Report 2006-5130.

Szilagyi, J., Parlange, M.B., 1998. Baseflow separation based on analytical solutions of the Boussinesq equation. J. Hydrol. 204, 251-260.

Tripathy, D., 2007. Development of a decision support tool for assessing impacts of land-use change on groundwater quantity. Ph.D. Thesis, Purdue University, West Lafayette, Indiana, USA.

USDA (United States Department of Agriculture), 2005. Underground drainage-a secret of Americas bounty. Agr. Res. Magazine 53 (9), 4-6.

USDA-ERS (United States Department of Agriculture - Economic Research Service), 1987. Farm Drainage in the United States: History, Status, and Prospects. Misc. Publication No. 1455. USDA-ERS.

USDCBC (United States Department of Commerce, Bureau of the Census). 1932. Fifteenth Census of the United States: 1930. Drainage of Agricultural Lands.

USDCBC (United States Department of Commerce, Bureau of the Census). 1952. US Census of Agriculture: 1950. vol. IV. Drainage of agricultural lands.

USDCBC (United States Department of Commerce, Bureau of the Census). 1961. US Census of Agriculture: 1959. Final Report, vol. IV. Drainage of agricultural lands.

USGS (United States Geological Survey), 2010. Water resources data for the United States water year 2007. <http://wdr.water.usgs.gov/wy2007/search.jsp> (05.26.11).

Vivoni, E.R., Entekhabi, D., Bras, R.L., Ivanov, V.Y., 2007. Controls on runoff generation and scale-dependence in a distributed hydrologic model. Hydrol. Earth Syst. Sci. 11 (5), 1683-1701.

Vogel, R.M., Kroll, C.N., 1992. Regional geohydrologic-geomorphic relationships for the estimation of low flow statistics. Water Resour. Res. 28 (9), 2451-2458.

Vogel, R.M., Kroll, C.N., 1996. Estimation of baseflow recession constants. Water Resour. Manage. 10, 303-320.

V3 Companies, Ltd and SCSWCD (Steuben County Soil and Water Conservation District), 2006. Pigeon Creek watershed management plan. Steuben County Soil and Water Conservation District. Angola, IN.

Ward, R.C., Robinson, M., 1990. Principles of hydrology. Maidenhead. McGraw-Hill, 365 pp.

Wayne, W.J., 1956. Thickness of drift and bedrock physiography of Indiana north of the Wisconsin glacial boundary. Indiana Geological Survey, Report of Progress No. 7, 70 p.

Whitney, G.G., 1994. From coastal wilderness to fruited plain: a history of environmental change in temperate North America from 1500 to the present. Cambridge University Press, New York.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behaviour? J. Animal Ecol. 75, 1182-1189.

Wilkerson, J., Merwade, V., 2010. Incorporating surface storage and slope to estimate Clark unit Hydrographs for ungauged Indiana watersheds. J. Hydrol. Eng. 15 (11), 918-930.

Wright, C.E., 1970. Catchment characteristics influencing low flows. Water Water Eng.74, 468-471.

Yang, G., Bowling, L. C., Cherkauer, K. A., Pijanowski, B. C., 2011. The impact of urban development on hydrologic regime from catchment to basin scales. Landscape Urban Plann. 103(2), 237-247.

Zhang, Y.K., Schilling, K.E., 2006. Increasing streamflow and baseflow in the Mississippi River since 1940: Effect of land use change. J. Hydrol. 324, 412-422.

Zhu, Y., Day, R.L., 2009. Regression modeling of streamflow, baseflow, and runoff using geographic information systems. J. Environ. Manage. 90, 946-953.

Zucker, L.A., Brown, L.C., 1998. Agricultural drainage: Water quality impacts and subsurface drainage studies in the Midwest. Ohio State University Extension Bulletin 871: The Ohio State University.

Table 1. Review of regression models describing relationships between baseflow and watershed characteristics.

Author	Method	Location	Watershed characteristics	Related to
Lacey and Grayson (1998)	Multiple linear regression	Australia	Basin drainage area, elevation, potential evapotranspiration, forest land cover, rainfall, stream length.	BFI
Haberlandt et al. (2001)	Stepwise multiple regression	Germany	Slope, topographic index, saturated hydraulic conductivity, mean annual precipitation.	BFI
Mazvimavi et al. (2005)	Multiple linear regression	Zimbabwe	Mean annual precipitation, slope, and wooded grassland and grassland cover.	BFI
Neff et al. (2005)	Multiple regression	Great Lakes, Canada and United States	Bedrock, coarse, fine, organic, till, proportion of surface water.	BFI and Baseflow
Gebert et al. (2007)	Multiple linear regression	Wisconsin, United States	Basin drainage area, soil infiltration rate, basin storage, baseflow factor.	Baseflow
Longobardi and Villani (2008)	Simple and multiple linear regression	Italy	Permeability index	BFI
Santhi et al. (2008)	Stepwise multiple regression	Conterminous, United States	Relief, percent sand, precipitation, potential evapotranspiration.	BFI and Baseflow
Bloomfield et al. (2009)	Stepwise multiple regression	United Kingdom	Urban area Lithology	BFI and Baseflow
Zhu and Day (2009)	Multiple linear regression	Pennsylvania, United States	Basin drainage area, annual precipitation minus evapotranspiration, elevation, annual precipitation.	Baseflow

Table 2. Streamflow gauging stations with corresponding coordinates, period of record, and delineated drainage area of watersheds used for the regression analysis.

USGS Site ID	Station name and location	Latitude	Longitude	Period of record	Drainage area (km^2)
Watersheds used for model development					
03275000	Whitewater River near Alpine, IN	39.579	-85.158	1928-present	1272.7
03324000	Little River near Huntington, IN	40.904	-85.406	1944-present	657.3
03325500	Mississinewa River near Ridgeville, IN	40.280	-84.992	1946-present	342.7
03328000	Eel River at North Manchester, IN	40.994	-85.781	1930-present	1030.4
03329700	Deer Creek near Delphi, IN	40.590	-86.621	1944-present	707.7
03331500	Tippecanoe River near Ora, IN	41.157	-86.564	1943-present	2227.0
03339500	Sugar Creek at Crawfordsville, IN	40.049	-86.899	1938-present	1299.1
03351500	Fall Creek near Fortville, IN	39.955	-85.867	1941-present	447.8
03361500	Big Blue River at Shelbyville, IN	39.529	-85.782	1943-present	1087.0
03362000	Youngs Creek near Edinburgh, IN	39.419	-86.005	1942-present	280.0
03363500	Flatrock River at St. Paul, IN	39.418	-85.634	1930-present	772.0
03364500	Clifty Creek at Hartsville, IN	39.275	-85.702	1948-present	235.1
03366500	Muscatatuck River near Deputy, IN	38.804	-85.674	1948-present	745.5
04094000	Little Calumet River at Porter, IN	41.622	-87.087	1945-present	171.4
04180000	Cedar Creek near Cedarville, IN	41.219	-85.076	1946-present	671.6
05515500	Kankakee River at Davis, IN	41.400	-86.701	1925-present	1416.5
05516500	Yellow river at Plymouth, IN	41.340	-86.304	1948-present	740.6
05524500	Iroquois River near Foresman, IN	40.871	-87.307	1949-present	1168.4
Watersheds used for model validation only					
03342500	Busseron Creek near Carlisle, IN	38.974	-87.426	1943-present	579.8
03334500	South Fork Wildcat Creek near Lafayette, IN	40.418	-86.768	1943-present	642.0
04099510	Pigeon Creek near Angola, IN	41.634	-85.110	1945-present	263.8
05536190	Hart Ditch at Munster, IN	41.561	-87.481	1942-present	218.3

Table 3. Watershed characteristics used as independent variables for regression analysis

Variable	Unit	Notation
Basin Drainage Area	km^2	BDA
Tile Drained Area	%	TDA
Basin Relief	m	BH
Average Basin Slope	%	ABS
Total Channel Length	km	TCL
Average Bedrock Depth	m	ABD
Annual Precipitation	mm	APCP
Annual Evapotranspiration	mm	AET
Forest Land Cover	%	FLC
Urban Land Cover	%	ULC
Grass Land Cover	%	GLC
Agricultural Land Cover	%	ALC
Water Land Cover	%	WLC[a]
Hydrologic Soil Group (A-D)	%	HSG
Baseflow Index	No unit	BFI

[a]WLC represents proportion of surface water in the watershed.

Table 4. Correlation matrix for variables used in baseflow regression analysis.

	Qb	BDA	TDA	BH	ABS	TCL	ABD	APCP	AET	FLC	ULC	GLC	ALC	WLC	HSGA	HSGB	HSGC	HSGD	BFI
Qb	1.000																		
BDA	0.906	1.000																	
TDA	0.029	0.086	1.000																
BH	0.277	0.365	-0.282	1.000															
ABS	0.005	-0.037	-0.826	0.512	1.000														
TCL	0.861	0.946	0.059	0.212	-0.045	1.000													
ABD	0.224	0.124	-0.293	-0.325	0.161	0.227	1.000												
APCP	0.183	0.001	-0.052	0.234	0.131	-0.065	-0.263	1.000											
AET	0.206	0.120	-0.161	0.179	0.147	0.121	0.009	0.012	1.000										
FLC	0.256	0.210	-0.768	0.279	0.785	0.280	0.214	0.040	0.338	1.000									
ULC	-0.043	-0.185	-0.176	-0.338	0.046	-0.217	0.550	-0.126	0.085	0.035	1.000								
GLC	0.049	-0.046	-0.697	0.237	0.655	-0.072	0.149	0.102	0.291	0.727	0.422	1.000							
ALC	-0.095	-0.012	0.764	-0.095	-0.657	-0.034	-0.320	-0.027	-0.282	-0.796	-0.518	-0.911	1.000						
WLC	0.423	0.291	-0.282	-0.299	0.188	0.423	0.829	-0.248	0.277	0.466	0.569	0.368	-0.528	1.000					
HSGA	0.248	0.179	-0.075	-0.658	-0.313	0.370	0.329	-0.187	0.163	0.156	0.327	0.095	-0.347	0.543	1.000				
HSGB	0.463	0.425	-0.002	0.052	-0.098	0.309	-0.040	0.040	0.260	0.159	0.152	0.188	-0.159	0.195	0.360	1.000			
HSGC	-0.309	-0.260	-0.021	0.152	0.255	-0.201	0.032	0.001	-0.207	-0.057	-0.199	-0.132	0.103	-0.211	-0.604	-0.903	1.000		
HSGD	0.379	0.303	-0.289	-0.276	0.262	0.486	0.615	-0.161	0.240	0.604	0.386	0.354	-0.527	0.838	0.749	0.277	-0.405	1.000	
BFI	0.542	0.417	-0.177	-0.095	0.000	0.426	0.399	-0.119	0.374	0.345	0.419	0.323	-0.351	0.693	0.649	0.657	-0.733	0.674	1.000

Table 5. Calculated average annual baseflow and baseflow index (BFI) from 1973-2003 in 22 Indiana watersheds. Highlighted watersheds were used in model validation only.

USGS Site ID	Baseflow (mm/yr)	Total flow (mm/yr)	BFI (BFLOW)	Location
03275000	259	403	0.64	center
03324000	160	334	0.48	north
03325500	146	357	0.41	center
03328000	221	357	0.62	north
03329700	193	323	0.60	north
03331500	306	385	0.79	north
03339500	187	339	0.55	center
03351500	253	385	0.66	center
03361500	258	397	0.65	center
03362000	199	381	0.52	center
03363500	229	395	0.58	center
03364500	193	394	0.49	south
03366500	173	427	0.40	south
04094000	297	424	0.70	north
04180000	213	353	0.60	north
05515500	323	365	0.88	north
05516500	234	365	0.64	north
05524500	228	344	0.66	north
03334500	216	351	0.61	center
03342500	196	386	0.51	south
04099510	310	394	0.79	north
05536190	212	415	0.51	north

Table 6. Regression equations for estimating annual baseflow and baseflow index in Indiana.

Model Description	Equation	R^2
Model A	$\log(Q_b) = 1.476 + 0.953\log(BDA) + 1.424\log(APCP) + 1.260\log(BFI)$ $Q_b = 29.896 BDA^{0.953} APCP^{1.424} BFI^{1.260}$	0.94
Model B	$\log(Q_b) = 1.626 + 0.963\log(BDA) - 0.077(TDA) + 1.400\log(APCP) + 1.224\log(BFI)$ $Q_b = 42.253 BDA^{0.963} TDA^{-0.077} APCP^{1.400} BFI^{1.224}$	0.94
BFI	$\log(BFI) = -0.397 + 0.105\log(WLC) + 0.152\log(HSGB) - 0.045\log(HSGC)$ $BFI = 0.401 WLC^{0.105} HSGB^{0.152} HSGC^{-0.045}$	0.91

Qb = annual baseflow (m^3); BDA = basin drainage area (km^2); BFI = baseflow index; APCP = annual precipitation (mm); TDA = tile drained area (%);WLC = percent of open water bodies in the watershed; HSG B = hydrologic soil group B; and HSG C = hydrologic soil group C.

Table 7. Relative error (%) between predicted baseflow and filtered baseflow using models A and B during 1974-2003 period in 22 Indiana watersheds. Highlighted watersheds were used in model validation only.

USGS Site ID	Model A	Model B	Model A	Model B
	Calculated BFI[a]		Calculated BFI[b]	
03275000	5	2	12	5
03324000	5	6	6	8
03325500	10	10	14	15
03328000	2	1	10	7
03329700	5	1	4	9
03331500	10	7	8	5
03339500	3	2	8	4
03351500	4	3	9	7
03361500	0	2	10	12
03362000	2	0	10	8
03363500	0	1	8	10
03364500	5	3	7	5
03366500	6	2	5	12
04094000	9	3	6	0
04180000	11	7	6	3
05515500	5	5	6	5
05516500	2	5	12	9
05524500	1	3	2	3
03342500	3	7	29	31
03334500	5	2	5	2
04099510	5	16	5	16
05536190	0	0	0	0

[a] BFI calculated with the BFLOW program; [b] BFI calculated with BFI regression equation.

Table 8. Performance of models A and B during 1954-1973 validation period in 22 Indiana watersheds. Highlighted watersheds were used in model validation only.

	Model A			Model B		
Watershed	R^2	E_{NS}	RE	R^2	E_{NS}	RE
03275000	0.42	0.40	4	0.43	0.25	11
03324000	0.64	0.52	9	0.63	0.53	8
03325500	0.57	0.57	1	0.57	0.55	1
03328000	0.49	0.42	2	0.49	0.37	5
03329700	0.39	0.34	7	0.38	0.37	2
03331500	0.35	0.20	8	0.34	0.11	10
03339500	0.87	0.53	14	0.87	0.63	10
03351500	0.73	-0.16	20	0.73	0.01	18
03361500	0.54	0.52	3	0.54	0.50	4
03362000	0.78	0.06	21	0.77	0.20	19
03363500	0.58	0.53	2	0.57	0.52	0
03364500	0.59	0.10	17	0.59	0.21	16
03366500	0.53	-0.06	20	0.53	-0.56	25
04094000	0.51	-0.24	6	0.51	-0.74	11
04180000	0.47	0.38	6	0.48	0.28	9
05515500	0.36	-0.01	1	0.37	0.01	0
05516500	0.66	-0.10	20	0.66	0.16	17
05524500	0.51	-0.08	22	0.51	0.11	18
03342500	0.04	-7.38	54	0.04	-8.47	56
03334500	0.81	0.67	8	0.83	0.71	5
04099510	0.26	-1.35	26	0.24	-3.13	34
05536190	0.50	-1.28	30	0.50	-1.29	30

List and description of figures

Fig. 1. Indiana map with delineated watersheds, USGS stream gauging stations, and karst areas.

Fig. 2. Annual baseflow in 22 Indiana watersheds during 1974-2003.

Fig. 3. Predicted BFI versus BFLOW BFI in 18 Indiana watersheds during 1974-2003.

Fig. 4. Predicted versus filtered average annual baseflow for models A and B in 22 Indiana watersheds during 1974-2003. Calculated BFI with the BFLOW program and regression equation was used for baseflow predictions.

Fig. 5. Comparison of simulated to filtered annual baseflow in 22 Indiana watersheds during the 1954-1973 validation period: Model A.

Fig. 6. Comparison of simulated to filtered annual baseflow in 22 Indiana watersheds during the 1954-1973 validation period: Model B.

Fig. 7. Comparison of simulated to filtered average annual baseflow in 22 Indiana watersheds during the 2004-2010 validation period for Models A and B.

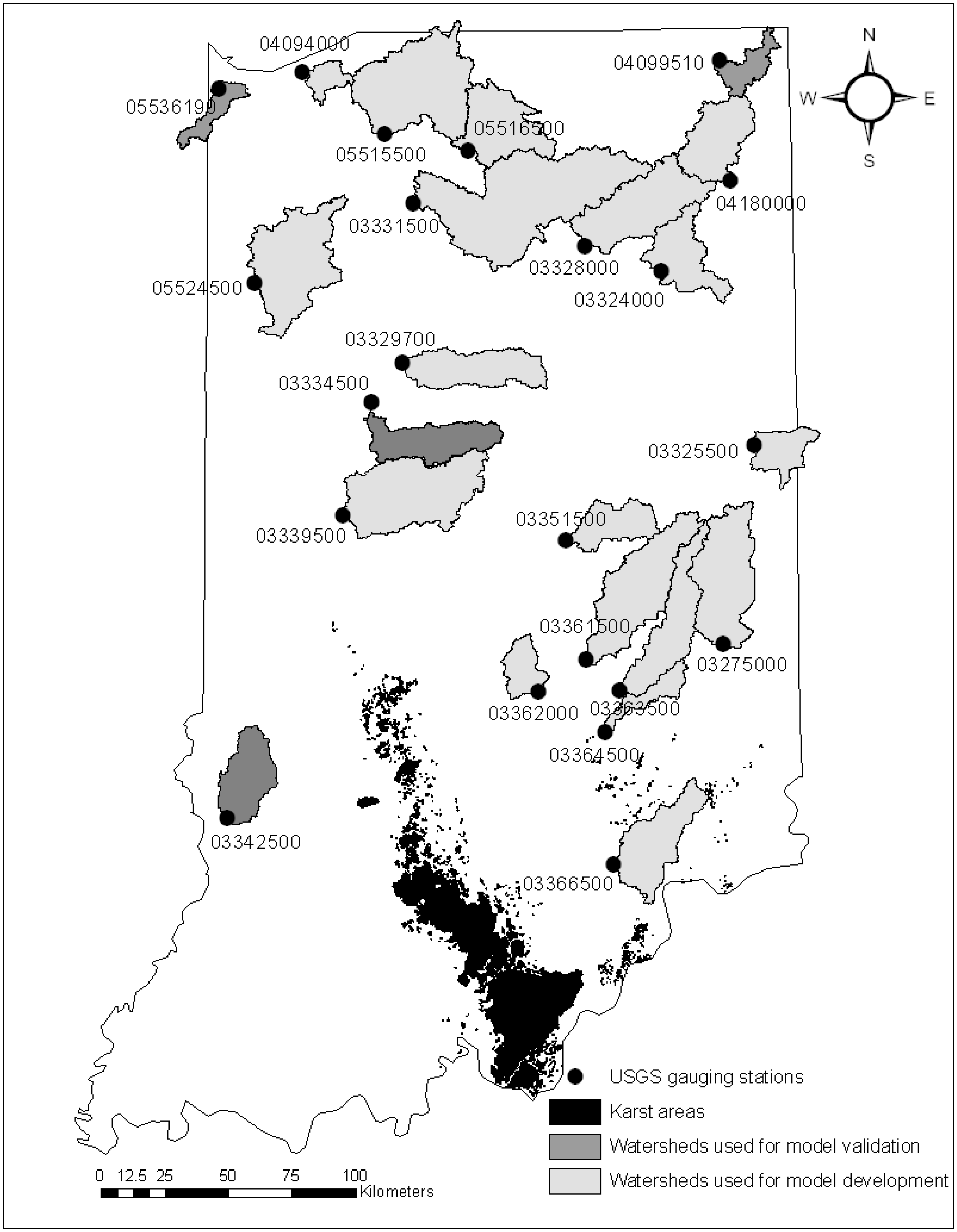
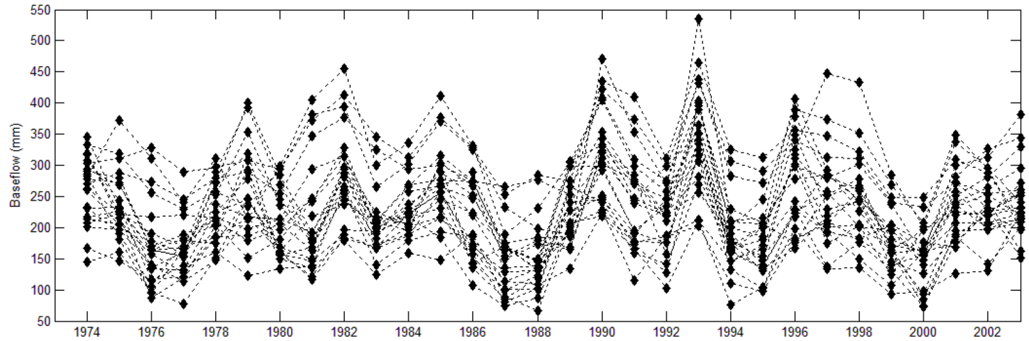

N
W
E
S
04094000
04099510
05536190
05516500
05515500
04180000
03331500
03328000
03324000
05524500
03329700
03334500
03325500
03351500
03339500
03361500
03275000
03362000
03363500
03364500
03342500
03366500
USGS gauging stations
Karst areas
Watersheds used for model validation
Watersheds used for model development
0 12.5 25 50 75 100
Kilometers

Fig. 1.



Baseflow (mm)
550
500
450
400
350
300
250
200
150
100
50
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002

Fig. 2.

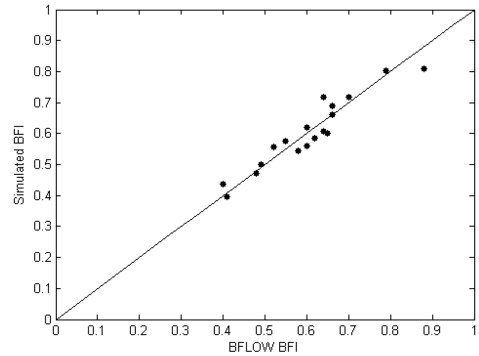


Fig. 3.

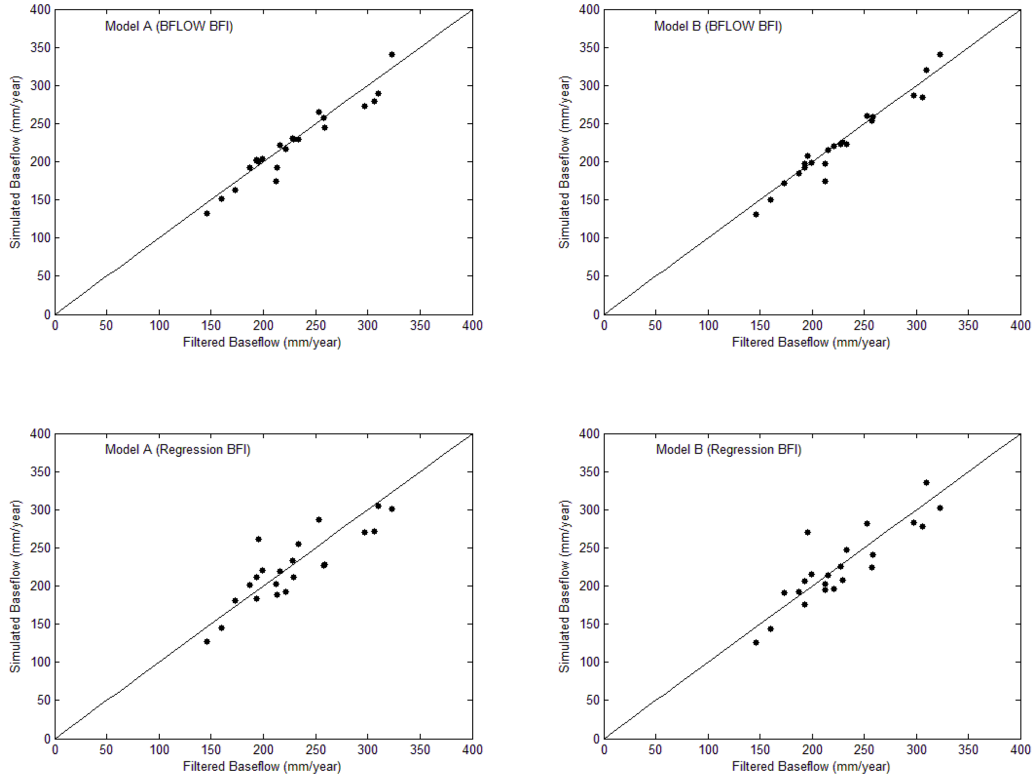
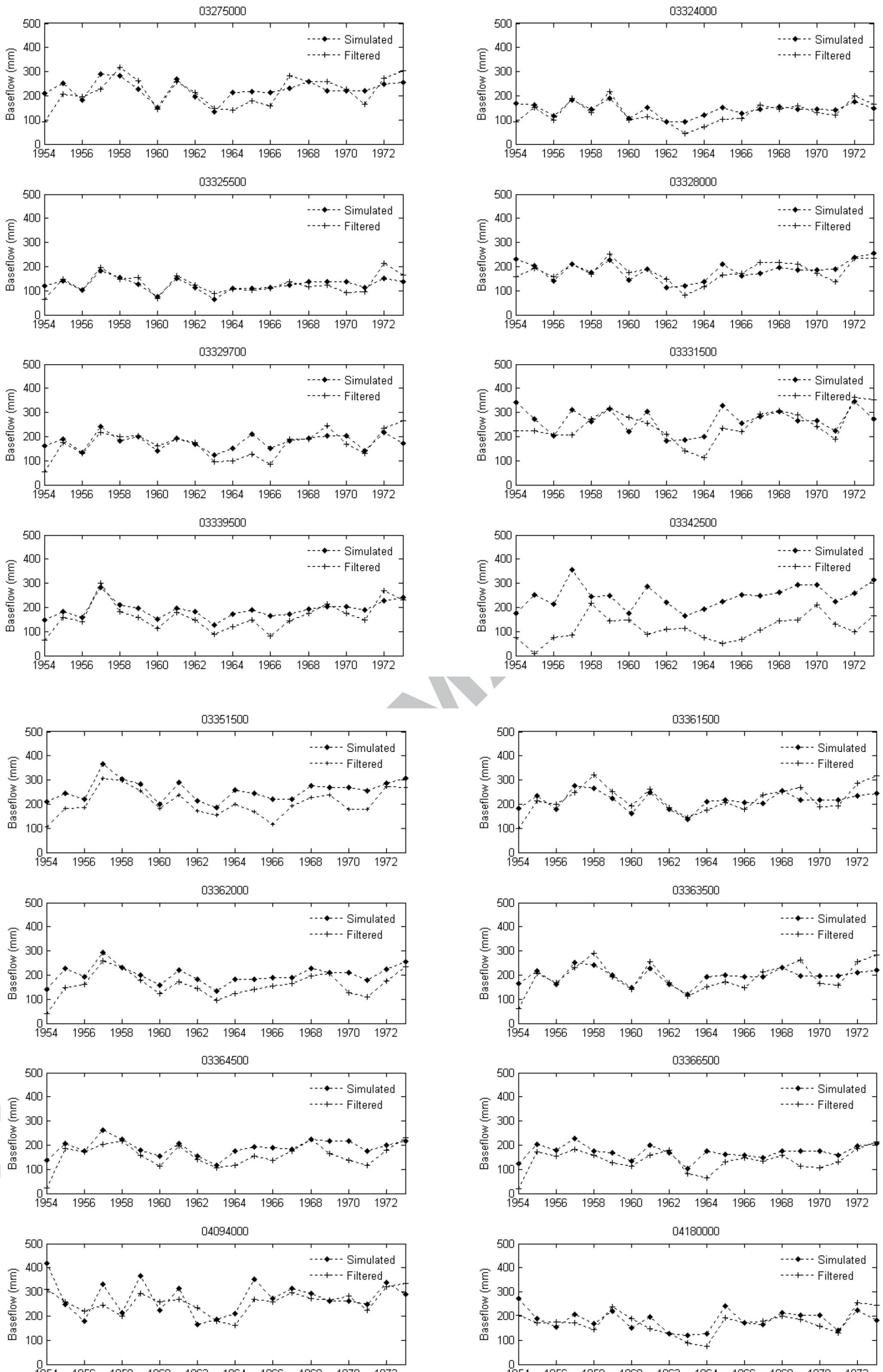

Model A (BFLOW BFI)
Model B (BFLOW BFI)
Model A (Regression BFI)
Model B (Regression BFI)
Simulated Baseflow (mm/year)
Filtered Baseflow (mm/year)
0
50
100
150
200
250
300
350
400

Fig. 4.


03275000
03324000
03325500
03328000
03329700
03331500
03339500
03342500
03351500
03361500
03362000
03363500
03364500
03366500
04094000
04180000
Baseflow (mm)
Simulated
Filtered
1954 1956 1958 1960 1962 1964 1966 1968 1970 1972
0 100 200 300 400 500

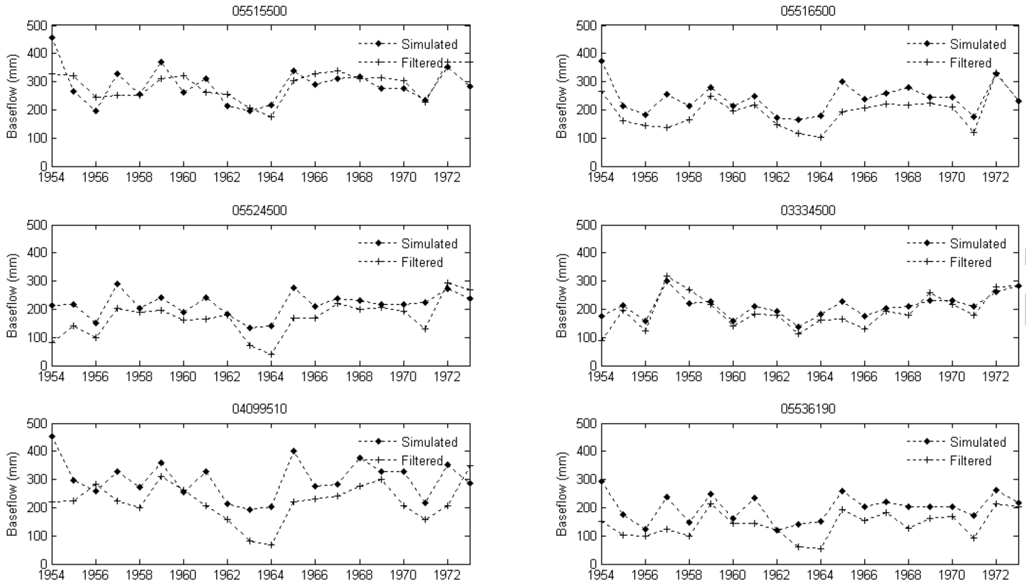
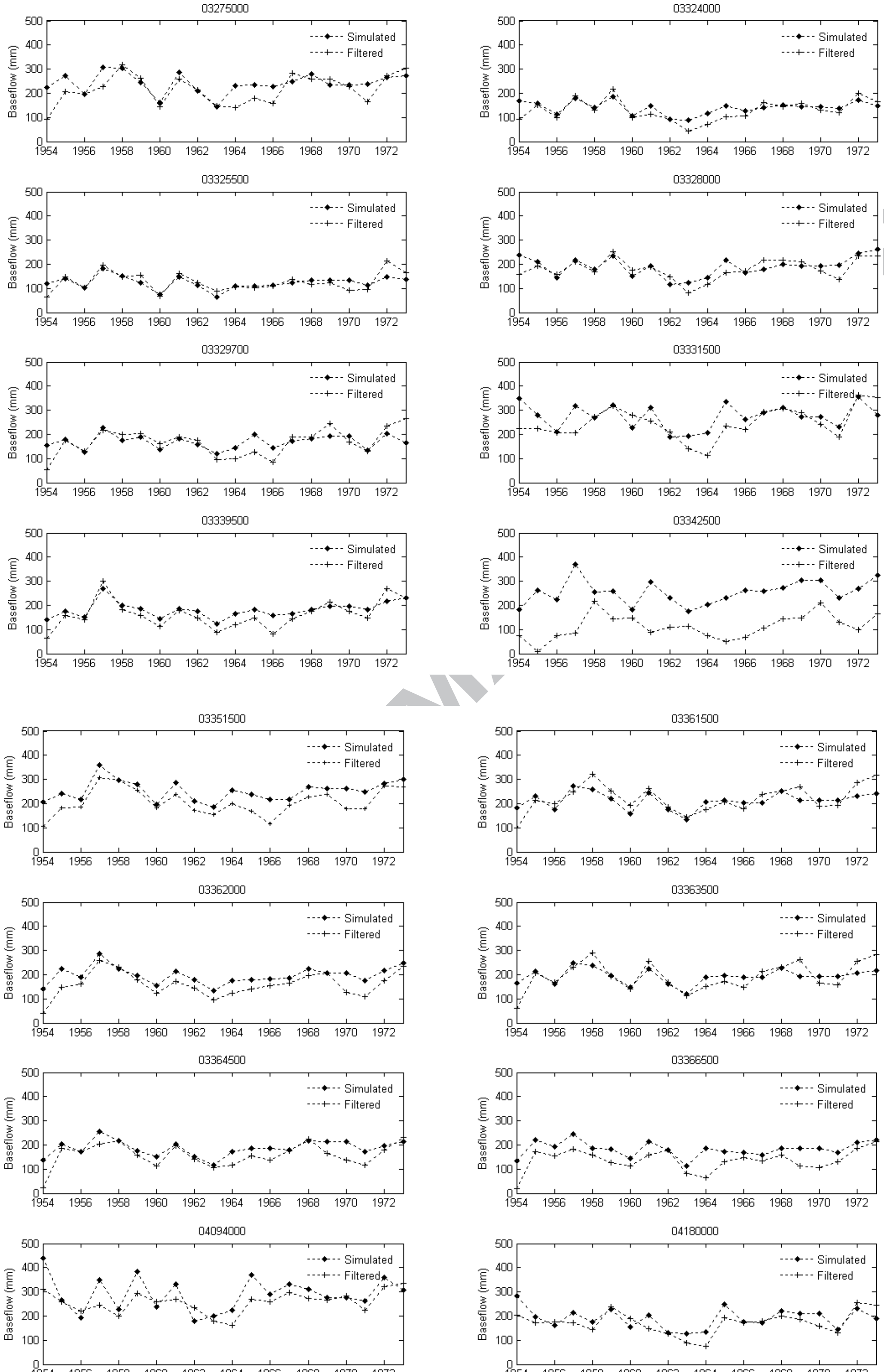

05515500
05516500
05524500
03334500
04099510
05536190
Simulated
Filtered
Baseflow (mm)
1954 1956 1958 1960 1962 1964 1966 1968 1970 1972
0 100 200 300 400 500

Fig. 5.


03275000
03324000
03325500
03328000
03329700
03331500
03339500
03342500
03351500
03361500
03362000
03363500
03364500
03366500
04094000
04180000
Baseflow (mm)
Simulated
Filtered
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
0
100
200
300
400
500

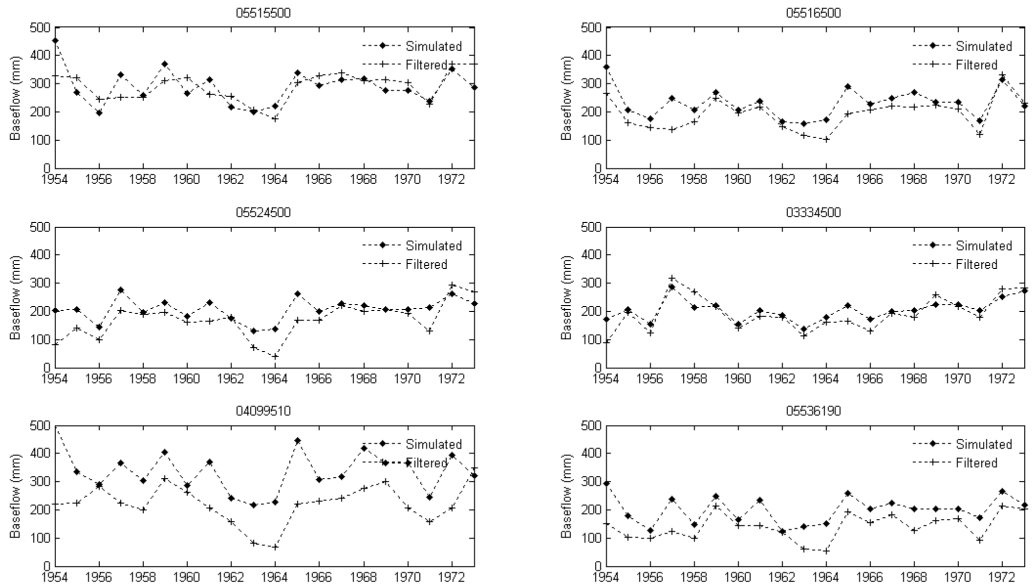

05515500
05516500
05524500
03334500
04099510
05536190
Baseflow (mm)
Simulated
Filtered
0
100
200
300
400
500
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972

Fig. 6.

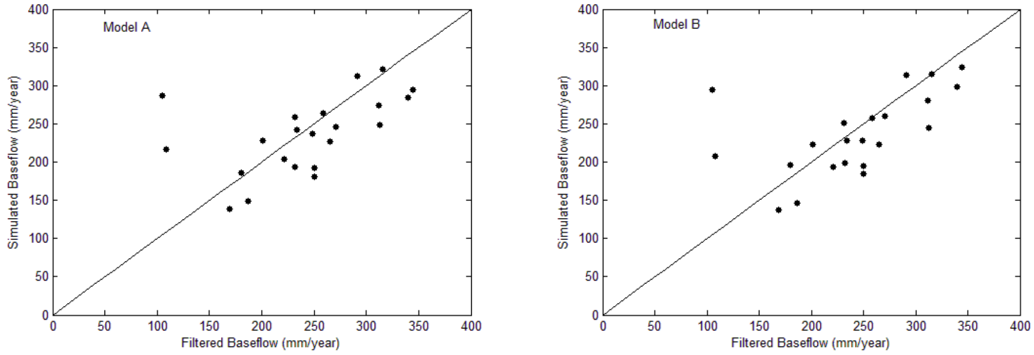

Model A
Simulated Baseflow (mm/year)
Filtered Baseflow (mm/year)
Model B
Simulated Baseflow (mm/year)
Filtered Baseflow (mm/year)

Fig. 7.

Research Highlights

1. Methods for for predicting baseflow and baseflow index at ungauged sites are developed.
2. Baseflow data from 22 watersheds in Indiana are used in this study.
3. Four equations for baseflow and one equation for BFI were developed and evaluated.
4. The methods reasonably estimate baseflow.
5. These equations can be used to estimate baseflow and BFI at ungauged sites in Indiana.