Ecohydrological model parameter selection for stream health evaluation

Sean A. Woznicki [a], A. Pouyan Nejadhashemi [a,*], Dennis M. Ross [b], Zhen Zhang [c], Lizhu Wang [d], Abdol-Hossein Esfahanian [b]

[a] Department of Biosystems and Agricultural Engineering, 524 S. Shaw Lane, Room 216, Michigan State University, East Lansing, MI 48824, USA
[b] Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
[c] Physical Sciences Division, Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[d] International Joint Commission, Great Lakes Office, Windsor, ON N9A 6T3, Canada

HIGHLIGHTS

- A watershed model generated over 200 flow regime and water quality variables.
- We defined appropriate scale for stream health models.
- Three techniques were used to identify variables for ANFIS stream health models.
- Clustered models with Bayesian variable selection generally performed the best.
- This study guides selection of influential variables for stream health models.

ARTICLE INFO

ABSTRACT

Variable selection is a critical step in development of empirical stream health prediction models. This study develops a framework for selecting important in-stream variables to predict four measures of biological integrity: total number of Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa, family index of biotic integrity (FIBI), Hilsenhoff biotic integrity (HBI), and fish index of biotic integrity (IBI). Over 200 flow regime and water quality

Abbreviations: ANFIS, artificial neuro-fuzzy inference system; DH11, annual maximum of 1-day moving average flows divided by the median for the entire record (cms); DH14, flood duration, the 95th percentile of mean monthly flows divided by the mean of the monthly means; DH19, high flow duration above seven times the median flow for the entire flow record; DH22, flood interval for a 1.67 year recurrence interval; DH24, flood free days, maximum number of days flow is below a flood recurrence of 1.67 years; DL2, annual minimum of 3-day moving average flow (cms); DL4, annual minimum of 30-day moving average flow (cms); DL11, annual minimum flow divided by the median flow for the entire record (cms); DL12, annual minimum of 7-day moving average flow divided by the median for the entire record (cms); DL15, low exceedance flows, the 90% exceedance value divided by the median for the entire record; DL16, low flow pulse duration, the median of the average annual pulse durations for flow events below the 25th percentile value for the entire flow record; DL18, number of zero-flow days for the entire flow record; EPT taxa, number of Ephemeroptera, Plecoptera, and Trichoptera taxa; FH2, variability in high pulse count; FH6, flood frequency (average number of events) above three times the median flow for the entire record; FH7, flood frequency (average number of events) above seven times the median flow for the entire record; FH10, flood frequency (average number of events) above median of the annual flow minima; FIBI, Family Index of Biotic Integrity; FL1, low flood pulse count, number of flow events with flows below the 25th percentile of the entire flow record; FL2, variability in low pulse count; FL3, frequency of low pulse spells; HBI, Hilsenhoff Biotic Index; IBI, fish Index of Biotic Integrity; MA1, mean of the daily flow values for the entire flow record (cms); MA3, mean of the coefficients of variation for each year (cms); MA4, standard deviation of the percentiles of the logs of the entire flow record divided by the mean of percentiles of the logs; MA12, mean of January flows (cms); MA16, mean of May flows (cms); MA24, variability of January flows; MA26, variability of March flows; MA27, variability of April flows; MA28, variability of May flows; MA32, variability of September flows; MA33, variability of October flows; MA34, variability of November flows; MA35, variability of December flows; MA38, variability across monthly flows; MA39, variability across monthly flows; MA43, variability across annual flows; MA45, skewness in annual flows; MH10, mean of maximum October flows (cms); MH12, mean of maximum December flows (cms); MH13, variability across maximum monthly flow values; MH15, high flow discharge index (the 1% exceedance value divided by the median flow for the entire record); MH18, variability across annual maximum flows; MH22, high flow volume, the average flow volume for flow events above three times the median flow for the entire record divided by the median flow for the entire record (m^3); MH24, high peak flow for flow events above the median flow for the entire record (cms); MH25, high peak flow for flow events above three times the median flow for the entire record (cms); MH27, high peak flow for flow events above the 75th percentile for the entire record (cms); MinPann, annual mineral phosphorus concentration; ML1, mean minimum flows for January (cms); ML2, mean minimum flows for February (cms); ML4, mean minimum April flows (cms); ML9, mean of minimum September flows (cms); ML11, mean minimum flows (cms); ML12, mean minimum December flows (cms); ML13, variability across minimum monthly flows; ML14, mean of the ratios of minimum annual flows to the median flow for each year; NH4ann, annual average ammonium concentration (mg/L); NH4djf, average daily ammonium concentration of December–January–February (mg/L); NH4mam, average daily ammonium concentration of March–April–May (mg/L); NH4son, average daily ammonium concentration of September–October–November (mg/L); NO2mam, average daily nitrite concentration of March–April–May (mg/L); NO3djf, average daily nitrate concentration of December–January–February (mg/L); NO3jja, average daily nitrate concentration of June–July–August (mg/L); NO3mam, average daily nitrate concentration of March–April–May (mg/L); NO3son, average daily nitrate concentration of September–October–November (mg/L); OrgNjja, average daily organic nitrogen concentration of June–July–August (mg/L); OrgPson, average daily organic phosphorus concentration of September–October–November (mg/L); RA2, variability in rise rate; RA4, variability in fall rate; RA5, number of day rises, in which the flow is greater than the previous day; RA6, change of flow; RA8, number of flow reversals; RA9, variability in flow reversals; SEDjja, average daily sediment concentration of June–July–August (mg/L); SWAT, Soil and Water Assessment Tool; TA1, constancy; TA2, predictability; TH2, variability in Julian date of annual maxima; TL3, seasonal predictability of low flow; TL4, seasonal predictability of non-low flow.

* Corresponding author.
E-mail address: pouyan@msu.edu (A.P. Nejadhashemi).

variables were calculated using the Hydrologic Index Tool (HIT) and Soil and Water Assessment Tool (SWAT). Streams of the River Raisin watershed in Michigan were grouped using the Strahler stream classification system (orders 1–3 and orders 4–6), *k*-means clustering technique (two clusters: C1 and C2), and all streams (one grouping). For each grouping, variable selection was performed using Bayesian variable selection, principal component analysis, and Spearman's rank correlation. Following selection of best variable sets, models were developed to predict the measures of biological integrity using adaptive-neuro fuzzy inference systems (ANFIS), a technique well-suited to complex, nonlinear ecological problems. Multiple unique variable sets were identified, all which differed by selection method and stream grouping. Final best models were mostly built using the Bayesian variable selection method. The most effective stream grouping method varied by health measure, although *k*-means clustering and grouping by stream order were always superior to models built without grouping. Commonly selected variables were related to streamflow magnitude, rate of change, and seasonal nitrate concentration. Each best model was effective in simulating stream health observations, with EPT taxa validation R^2 ranging from 0.67 to 0.92, FIBI ranging from 0.49 to 0.85, HBI from 0.56 to 0.75, and fish IBI at 0.99 for all best models. The comprehensive variable selection and modeling process proposed here is a robust method that extends our understanding of watershed scale stream health beyond sparse monitoring points.

1. Introduction

Rivers are an important resource for humans and natural systems alike. To humans, rivers provide water for consumption, economic and recreational opportunities, and innumerable ecosystem services. As a natural system, rivers are complex webs that support an array of flora and fauna. Although rivers are critical to humans and nature, they have degraded at an alarming rate. For example, the Environmental Protection Agency (EPA) has determined that 44% of surveyed streams and rivers in the United States are impaired for at least one designated use (USEPA, 2009). The primary causes of impairment are diverse, including pathogens, oxygen depletion, habitat alteration, nutrients, and sediment. Sources of impairment are mostly anthropogenic, including agriculture, hydromodification, habitat alteration, and municipal discharge. It is increasingly recognized that these activities are a primary threat to ecological integrity of freshwater systems because they affect water quality, biota, and habitat in many ways that are often complex (Allan, 2004).

By the 1970s, a majority of lakes, rivers, and coastal waters were deemed unsafe for fishing and swimming. Upon realizing the level of freshwater impairment due to unchecked pollution dumping in the United States, the Clean Water Act was enacted in 1972 with the goal "to restore and maintain the chemical, physical, and biological integrity of the Nation's waters". Since the act's passage, major water quality improvements have been accomplished; however, the degradation of native aquatic communities continues (Bryce et al., 2008). As a result, there has been a push to consider biological criteria as in addition to chemical or physical criteria to achieve water quality goals, because biological integrity is the ultimate endpoint of concern (Karr and Yoder, 2004). Biological integrity describes an ecosystem comparable to a natural habitat in terms of species composition, diversity, and functional organization, while supporting a balanced, integrated, and adaptive community of organisms (Karr and Dudley, 1981). The strength in using biological assessment to measure biological integrity is that biota residing in a stream are continual monitors of environmental quality and stream conditions (USEPA, 2011).

Multimetric biological indicators (e.g., indices of biotic integrity for macroinvertebrates or fish) are commonly used to measure stream health. These indicators represent large array of aquatic species that are sensitive to a variety of stressors (Wang et al., 2008; USEPA, 2011). Measuring multiple indicators, rather than a single endpoint, provides a holistic view of ecological integrity (Clapcott et al., 2012). However, assessing biological integrity is usually limited to a small portion of a watershed where data about aquatic fauna communities is collected (Wang et al., 2008). Meanwhile, biota often respond to landscape factors and stressors in a complex and nonlinear manner (Wang et al., 2008; Johnson and Host, 2010; Waite et al., 2010; Einheuser et al., 2012). In light of these limitations, modeling can play a major role in characterizing biological integrity in large study areas such as watersheds. However, modeling approaches are not straightforward because of the challenges in connecting specific disturbances or variables with multimetric measures of integrity (Niemi and McDonald, 2004; Wang et al., 2008). In reality, these indicators represent the lumped effects of different stressors (USEPA, 2011). Given these challenges, guidance is needed when developing and parameterizing stream health models.

Parameterization in ecological modeling is a common problem due to the complex, nonlinear, and uncertain relationships in these systems. Among hundreds of variables that influence stream health (e.g., landscape attributes, hydrologic indices, and water quality parameters), a limited number should be selected that are relevant to the problem and can be reliably obtained (Maier and Dandy, 2000). Selection of variables for assessment is subjective and it is impractical to monitor a large suite of variables (Pinto and Maheshwari, 2011). In addition, increased parameterization often leads to model over-fitting (Whittaker et al., 2010). Variable selection can be performed using expert knowledge or data-driven techniques (Adriaenssens et al., 2004). Using expert knowledge for selection of large number of variables is difficult, since the interactions between variables are often composite and confounding. Conversely, data-driven techniques for input variable selection such as artificial neural networks (ANNs) and Principal Component Analysis (PCA) do not require prior understanding of relationships between variables (Adriaenssens et al., 2004) and have been used extensively in characterizing stream health (Olden and Poff, 2003; Kennard et al., 2005; Fellows et al., 2006; Lencioni et al., 2007; Lücke and Johnson, 2009; Mondy and Usseglio-Polatera, 2014).

In ecological and environmental applications, model selection is especially important (Lek and Guégan, 1999; Metternicht, 2001; Chen and Mynett, 2003; Adriaenssens et al., 2004; Mathon et al., 2013). Multiple approaches have been used to develop predictive relationships between landscape factors or human disturbance stressors to stream health. Linear regression is widely used to analyze ecological data for prediction and explanation (Hawkins et al., 2000; Frimpong et al., 2005; Kennard et al., 2006; Pont et al., 2009; Maret et al., 2010; Moya et al., 2011). This approach implies linearity among variables, which is rare in ecology (Lek et al., 1996). The major drawback of multiple linear regression is its inability to account for nonlinearity between independent and dependent variables (Gevrey et al., 2003). Conversely, soft computing methods such as ANNs and fuzzy logic are well-suited to ecological and environmental applications because of the inherent uncertainty, complexity, ambiguity, and non-linearity of these systems and their data (Metternicht, 2001; Chen and Mynett, 2003; Adriaenssens et al., 2004). Developed by Zadeh (1965), fuzzy logic requires linguistic variables and rules rather than numerical values and Boolean logic (Marchini, 2011). Fuzzy logic is useful for prediction of multimetric indices because of its linguistic nature. For example,
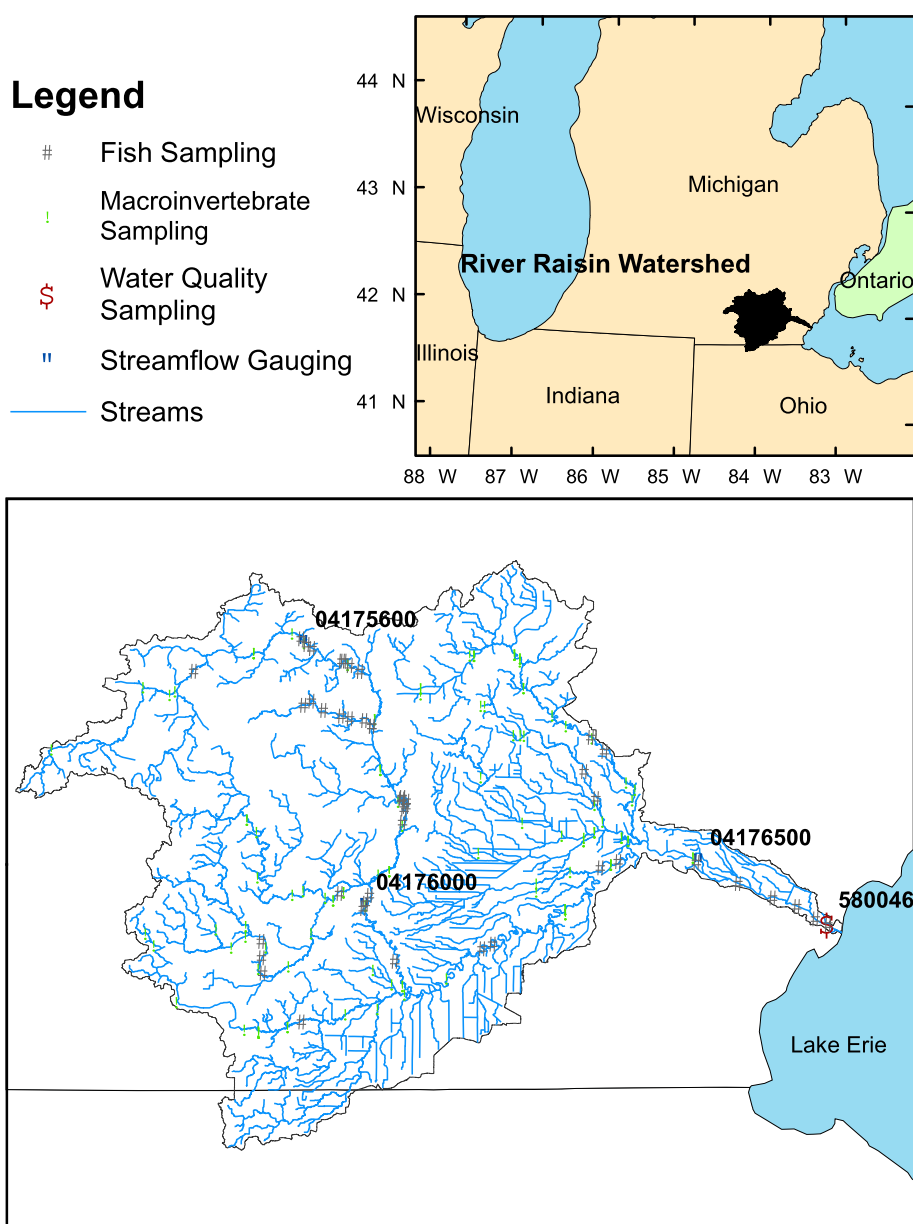
Fig. 1. River Raisin watershed.

the fish index of biotic integrity (Karr, 1981), uses linguistic terms such as 'poor', 'good', and 'excellent'. Several studies have demonstrated the utility of ANNs and fuzzy logic over linear regression methods in ecological settings (Lek et al., 1996; Einheuser et al., 2012, 2013a; Hamaamin et al., 2013). Fuzzy models have been used in development of environmental condition indices (Lermontov et al., 2009; Marchini et al., 2009), modeling wetland conditions (Mah and Bustami, 2012), and predicting habitat suitability (Mouton et al., 2009), algal biomass (Chen and Mynett, 2003), macroinvertebrate taxa (Adriaenssens et al., 2006; Einheuser et al., 2012), biotic integrity of fish (Einheuser et al., 2013a), baseflow (Hamaamin et al., 2013), and water quality (Ocampo-Duque et al., 2006). Similarly, ANNs have been used to link biological integrity to stream habitat and geomorphic conditions (Mathon et al., 2013), and model macroinvertebrate assemblages (Park et al., 2003; Compin and Céréghino, 2007; Lencioni et al., 2007; Mouton et al., 2010) and brown trout density (Lek et al., 1996). Studies have been

particularly successful in combining ANNs and fuzzy logic using adaptive-neuro fuzzy inference systems (ANFIS) to model biotic integrity of macroinvertebrates (Einheuser et al., 2012, 2013b) and fish (Einheuser et al., 2013a). Due to these recent successes, ANFIS was used to model stream health in this study. However, variable selection in ANFIS has no commonly agreed upon methods (Einheuser et al., 2012) and the number of variables needs to be controlled to limit noise.

Given the large array of variables and their numerous selection techniques available, there is a lack of guidance on how to proceed when developing stream health prediction models. Meanwhile, limiting the number of predictor variables used in ANFIS is critical to reduce the fuzzy logic rule set size and improve efficiency (Chen and Mynett, 2003), because the number of if-then rules in fuzzy logic is a function of the number of input variables and membership functions (Mahabir et al., 2006). This study explores multiple model parameterization methods ranging from simple (Spearman's Rank Correlation) to

complex (Bayesian variable selection), and then use those variables in model development for prediction of four stream health measures. The specific objectives of this study were to: (1) identify appropriate scales for stream health modeling, (2) select variable sets for further exploration in four stream health models, and (3) develop predictive models for stream health measures.

2. Materials and methods

2.1. Study area

The River Raisin watershed, hydrologic unit code 04100002, is located in south-eastern Michigan and a small portion of northern Ohio (Fig. 1). The watershed drains an area of 2757 km^2 into Lake Erie. Historically, the River Raisin watershed was dominated by wetlands and hardwood forest (Roth et al., 1996), but is now representative of many agricultural river systems in the Great Lakes region (Allan et al., 1997). Currently, land use/land cover in the watershed is distributed between agricultural row crops (53%), pasture (16%), forest (12%), urban/developed (11%), wetlands (7%), and open water (1%). Topography is hilly and rolling in the western and northwestern regions of the study area, while it is comparatively flat in the southeast. Underlying soils in the watershed are dominated by sandy loams, loams, and clay loams with moderate/high infiltration rates in the northwest, while in the southeast soils are clay, clay loams, and silty clays with lower infiltration rates (Dodge, 1998). The River Raisin watershed is considered to be one of the most biologically rich watersheds in Michigan (Allan et al., 1997).

2.2. Data collection

2.2.1. Physiographic data

Topography was obtained from the United States Geological Survey (USGS) in the form of the 30 m spatial resolution National Elevation Dataset (NED, 2014). Land use/land cover data was obtained from the United States Department of Agriculture (USDA) — National Agricultural Statistics Service (NASS) in the form of the 2012 Cropland Data Layer (CDL), which has a 30 m spatial resolution (NASS, 2012). Tabular and spatial soil data was obtained at the county level from the Natural Resources Conservation Service (NRCS) Soil Survey Geographic (SSURGO) Database (NRCS, 2014). Local crop rotations and management operations in the River Raisin watershed were obtained from Sommerlot et al. (2013). Spatial information in SSURGO, which includes physical and chemical soil properties, was gathered at scales ranging from 1:12,000 to 1:63,360. Climate data was obtained from the National Climatic Data Center (NCDC). Continuous daily precipitation and temperature data were available at five and four locations, respectively from 1988 to 2009.

Predefined streams and subwatersheds layers were obtained from the Great Lakes Regional River Database Classification System prepared by the Michigan Institute for Fisheries Research (IFR). Both datasets are based on the 1:24,000 resolution National Hydrography Dataset Plus. The database divides the stream network into confluence-to-confluence stream reaches, where each subwatershed contains an individual stream reach representing a stretch of homogeneous physicochemical, geomorphological, and biological features (Einheuser et al., 2013a,b). Using this database, the River Raisin watershed contains 1235 individual stream segments and associated subwatersheds.

2.2.2. Biological data

There are several advantages of using macroinvertebrates and fish as indicators of stream health. Macroinvertebrates react to localized impacts and a broad range of pollution tolerances on a relatively short-term basis (Barbour et al., 1999). In contrast, fish are useful because they represent long-term effects and broad habitat conditions (Barbour et al., 1999). Measures of stream health used in this study were: number of Ephemeroptera, Plecoptera, and Trichoptera (EPT)

taxa, Family-level Index of Biotic Integrity (FIBI), the Hilsenhoff Biotic Index (HBI), and fish Index of Biotic Integrity (IBI). EPT taxa is a count that quantifies the taxonomic richness of common pollutant-intolerant macroinvertebrate orders (Sponseller et al., 2001). FIBI is an index of macroinvertebrate composition and richness at a family-level taxonomic resolution. HBI is an organic pollution – tolerance index for macroinvertebrate taxa based on taxon-specific tolerance values (Hilsenhoff, 1988). Fish IBI is a multi-metric index detects divergence from biological integrity (Karr, 1999). Locations of biological sampling (71 macroinvertebrates, 45 fish) are presented in Fig. 1, while histograms of each macroinvertebrate stream health measure within the study area are presented in Fig. S1. Each measure was divided into five stream health classes by splitting the entire Michigan macroinvertebrate monitoring dataset (2634 data points) into quintiles, while for fish IBI classes from Lyons (1992) were used (Table S1).

Macroinvertebrate sampling was performed by the Michigan Department of Environmental Quality from 1996 to 2003 in June through September (MDEQ, 1997). Stream sampling length was between 30 m and 100 m, where the goal was to collect approximately 300 ± 60 organisms over a minimum of 20 min per site (Einheuser et al., 2012). Fish data were obtained from the Michigan River Inventory database (Seelbach and Wiley, 1997) and Michigan Department of Natural Resources monitoring program, where single-pass sampling occurred in wadeable streams along 80–960 m stretches and along lengths of 1610 m in non-wadeable streams between 1982–2007 (Einheuser et al., 2013a).

2.3. Modeling process

A multi-step modeling process was used to simulate the four stream health measures in the River Raisin watershed (Fig. 2). First, we use Soil and Water Assessment Tool (SWAT) to generate daily time-series streamflow, sediment, and nutrient data for every stream in the watershed. The streamflow data was input into the USGS Hydrologic Index Tool (HIT) (Henriksen et al., 2006) to calculate 171 hydrologically significant flow regime variables. Flow regime variables and sediment and nutrient concentration variables were grouped by three methods: all streams (no-grouping), k-means clustering, and by stream order. Variable selection was performed within each group using three methods (Bayesian variable selection, PCA, and Spearman's rank correlation) to produce best variable sets. Each variable set and grouping combination was used to build ANFIS models that predict stream health measures (EPT taxa, FIBI, HBI and IBI). Final best model selection (a combination of stream grouping method, variable selection method, and ANFIS model characteristics) for each stream health measure was based on multiple performance measures.

2.4. Soil and water assessment tool

The Soil and Water Assessment Tool (SWAT version 2012) is a physically-based, semi-distributed watershed/water quality model widely used for water resources planning and decision-making (Neitsch et al., 2005). Developed by the USDA — Agricultural Research Service, SWAT predicts the impact of management practices and climate change on hydrology and sediment, nutrient, pesticide, and bacteria yields on a daily time-step over long simulation periods (Arnold et al., 1998). SWAT simulates various processes of a watershed system, including surface runoff, soil erosion, evapotranspiration, nutrient cycling, crop growth, streamflow routing, sediment deposition and entrainment, bacterial growth and die-off, and pesticide transport (Gassman et al., 2007).

SWAT delineates a watershed into multiple subwatersheds based on topography and stream networks. Subwatersheds are further discretized into unique hydrologic response units (HRUs). An HRU is defined as an area of homogeneous land use, soil type, slope, and management operations. In this study, we used predefined Michigan IFR layers
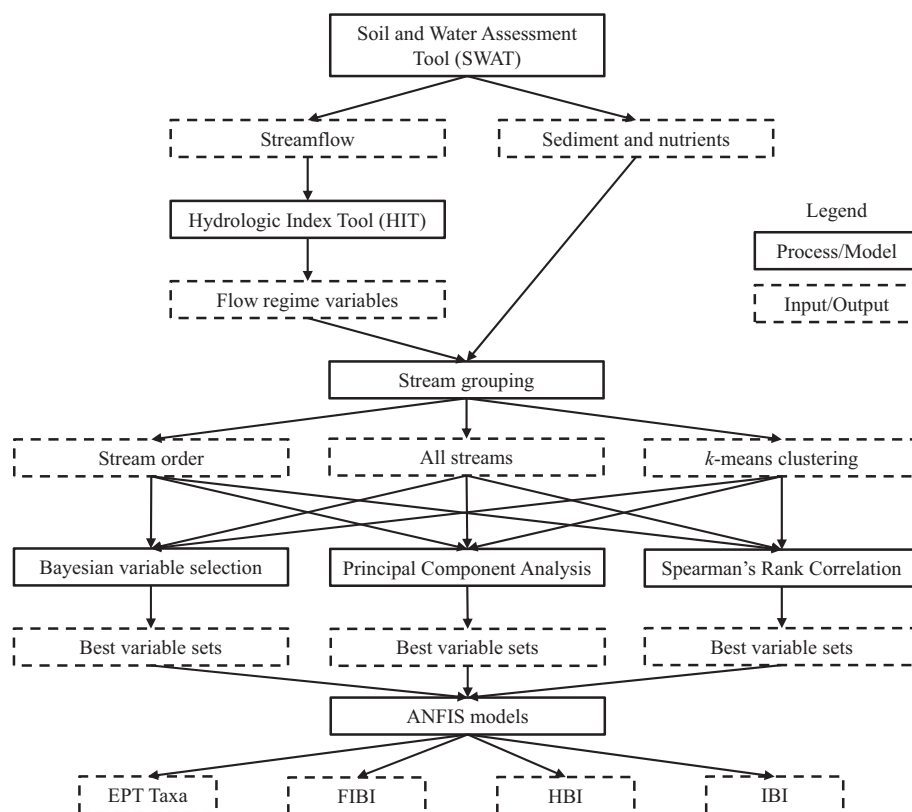
Fig. 2. Stream health variable selection and modeling process.

that divided the watershed into individual reaches and subwatersheds. Each subwatershed represents a stretch of homogeneous physicochemical, geomorphological, and biological features. Therefore, in each subwatershed, the dominant land use, soil type, and slope were selected for HRU definition in SWAT.

2.4.1. SWAT model calibration and validation

Model calibration was performed by modifying input parameter values and comparing model output values (such as time-series stream discharge) with corresponding measured data (White and Chaubey, 2005). Calibration was satisfactory after meeting some statistical criteria, such as minimization of error or optimizing the Nash–Sutcliffe model efficiency coefficient (NSE). Model validation was performed by comparing the calibrated model output with measured data for a time-period independent of the calibration period (Moriasi et al., 2007).

Three statistical criteria were used to ensure the model was calibrated and validated: NSE, root-mean-square error-observations standard deviation ratio (RSR), and percent bias (PBIAS). Moriasi et al. (2007) developed model evaluation guidelines for each parameter with recommended performance ratings on a monthly time-step. Calibration is considered satisfactory for any parameter at NSE > 0.50 and RSR < 0.70. Acceptable PBIAS varies by parameter, where streamflow calibration is satisfactory at PBIAS < |25%|, sediment at PBIAS < |55%|, and nutrients at PBIAS < |70%|.

The River Raisin watershed SWAT model was calibrated and validated for streamflow, sediment load, nitrate (NO_3) load, nitrite (NO_2) load, and total phosphorus (TP) load on a daily time-step. Locations of streamflow gauging stations (USGS 04175600, 04176000, and 04176500) and sediment and nutrient sampling (580046) are presented in Fig. 1. Streamflow was calibrated from 1996–2000 and validated from 2001–2005 at three USGS locations (streamflow gauging stations

04175600, 04176000, and 04176500), with a two year model initialization period (1994–1995). Based on grab sample data availability for sediment and nutrients, these parameters were calibrated from 2000–2002 and validated from 2003–2005 at one location (station 580046).

Following calibration, SWAT was run from 1988–2009 on a daily time-step. The first two years (1988–1989) were used for the model initialization period. Based on the SWAT model output, daily streamflow, sediment, and nutrient concentrations were obtained for all 1235 reaches in the watershed.

2.5. Stream characterization

Flow regime was characterized using the HIT (version 1.48), which calculates 171 biologically relevant hydrologic indices using daily and peak streamflow data. The indices were divided into the five components of flow regime: magnitude (M), frequency (F), duration (D), timing (T), and rate of change (R). Each component is further classified by type of flow event (A, average; L, low; and H, high). Flow regime characterization required continuous daily streamflow data for each stream reach in the River Raisin watershed from the beginning of the 1990 water year (October 1, 1989) through the end of the 2009 water year (September 30, 2009).

In addition to flow, seasonal and annual sediment and nutrient concentrations were also characterized for each reach. The specific nutrients were organic nitrogen (OrgN), nitrate (NO_3), nitrite (NO_2), ammonium (NH_4), organic phosphorus (OrgP), and mineral phosphorus (MinP). Seasons were defined as December–January–February (DJF), March–April–May (MAM), June–July–August (JJA), and September–October–November (SON), and annual (ANN). The seven water quality components defined across five time periods (annually and four seasons) resulted in 35 concentration measures. Combining the

171 hydrologic indices and the 35 sediment and nutrient concentration measures, there were a total of 206 variables from which to develop EPT taxa, FIBI, HBI, and IBI prediction models.

2.6. Stream grouping

Stream size has been reported as a key predictor of species richness and presence (Pont et al., 2009). Due to large variability in flow regime characteristics and composition of the organisms in different sections of the stream network, streams were grouped to improve predictability of stream health measures. In this study, two methods were used to group stream networks within the study area. The first method partitions streams based on the stream order and the second method partitions streams into k clusters based on the nearest mean. The results were compared with a no grouping scenario (all streams) in which one predictive model is used for all stream segments within the watershed.

2.6.1. Stream order

Stream order was calculated for all streams using the Strahler stream order classification system (Strahler, 1957). Based on the River Continuum Concept developed by Vannote et al. (1980), streams were grouped into headwaters (orders 1–3) and medium-sized streams (orders 4–6). This grouping considers broad characteristics of lotic communities that vary according to stream size from headwaters to river mouths.

2.6.2. k-Means clustering

As an alternative to the River Continuum Concept for grouping streams, a data-driven method known as k-means clustering was used. It was hypothesized that grouping streams based on their physicochemical characteristics may approximate ecological behavior as defined in the River Continuum Concept. The original data was standardized by computing each values' z-score and PCA was applied using the Euclidean distance metric. This resulted in 206 new variables (principal components) given by a linear combination of the original data. The data were then grouped into two clusters (C1 and C2) using k-means. This unsupervised learning technique was parameterized with two initial points and the Euclidean metric. A description of the model implementation and pseudocode is presented in the Supplementary materials.

2.7. Variable selection

Given the large number of variables characterizing the flow regime and water quality conditions in each reach, a variable selection procedure was required to eliminate redundant variables. The number of predictor variables used in ANFIS is limited by available stream health data, necessitating a maximum of three variables. This avoids model overfitting and is a function ANFIS model structure, which is discussed in Section 2.8. The variable selection methods are described below.

2.7.1. Spearman's rank correlation

Spearman's rank correlation is a nonparametric measure of statistical dependence between two variables, and has been used for variable selection and redundancy reduction in multiple studies related to predicting stream health measures (Maret et al., 2010; Waite et al., 2010; Einheuser et al., 2012, 2013a). Spearman's rank correlation coefficient (ρ) was calculated for all variable pairings. Predictor variables (flow regime and water quality variables) that exhibited significant correlation ($\alpha < 0.05$) with each of the stream health indices (EPT taxa, FIBI, HBI, and IBI) were identified. Independent variables with the highest ρ were selected. Of selected predictor variables that had high correlations ($\rho > |0.7|$) with each other, the one with the weakest correlation with the stream health measure was removed from consideration (Wang et al., 2008; Waite et al., 2010).

2.7.2. Principal component analysis

PCA orthogonally transforms the dataset of predictor variables (in which the variables may be correlated) into a new set of principal components (PCs) that are linearly uncorrelated (Pearson, 1901). The number of PCs is equal to the number of variables in the original dataset, where the first PC (PC1) explains the largest amount of variance and the remaining PCs explain non-increasing amounts of variance.

Individual variable PC loadings are the correlation coefficients between the PC score and original variables; they indicate the importance of that variable in accounting for the PC's variability. Therefore, the individual variables can be extracted to interpret the PC and the dataset's variation. Variables with the greatest component loadings were extracted from the PC1 to develop one variable set. In addition, the variable with the greatest component loading from each of the top three principal components (PC1–PC3) was extracted to create a new set of three variables. Finally, the coefficients of the first three PCs were used as variables in model development.

2.7.3. Bayesian variable selection

Bayesian variable selection was used to identify a subset of important variables from the independent variable set. The Bayesian framework assumes that the combinations and number of relevant variables that yields the most predictive power are random. Therefore, they can be sampled by combining the data evidence and uniform prior weights on all such possible combinations. Additional information about Bayesian variable selection is provided in the Supplementary material.

A measure of "selectivity" is used here, which directly compares the importance of the variables. The selectivity is defined as the probability of being selected out of all posterior samples for the variable. Variables with the greatest selectivity were chosen for further use for development of stream health predictive models.

2.8. ANFIS and best model selection

Predictive models were created using fuzzy logic for each of the stream health indices. Input variables are defined using graphical membership functions (MFs). An MF is a curve that determines an input value's degree of membership in a particular class. Here, a membership value of zero represents no membership and one represents full membership in a class.

Building membership functions and inference rules is one of the most challenging tasks in modeling with fuzzy logic (Chen and Mynett, 2003; Huang et al., 2010) and is often subjective (Adriaenssens et al., 2004) and time consuming (Huang et al., 2010). Given these limitations, a hybrid approach called adaptive-neuro fuzzy inference system (ANFIS) was used to optimize membership function development. ANFIS uses artificial neural networks (ANNs) to construct and tune membership functions by minimizing output error for use in fuzzy logic (Jang, 1993).

The ANFIS models were built using the Fuzzy Logic Toolbox in MATLAB R2013b (MathWorks, 2013). Five membership function shapes were tested and up to four membership functions were created for each variable. Triangular and trapezoidal membership functions are linear and are commonly used because of simplicity (Adriaenssens et al., 2004; Marchini, 2011). The remaining membership functions, generalized bell (Bell), Gaussian and Gaussian composite (GaussC), are nonlinear and better suited to ecological data (Marchini, 2011).

A limit of three variables was imposed on model creation. The number of variables and membership functions per variable limits are based on the size of the macroinvertebrate and fish datasets, where the number of samples cannot exceed the number of modifiable parameters when building ANFIS membership functions. The number of modifiable parameters in building an ANFIS model is a function of the number of variables, membership functions per variable, and shape of the membership function. If the number of samples exceeds the number of modifiable parameters in the predictive model, there is a danger of

overfitting and noise due to the significant increase in rules and parameters (Cobaner, 2011; Sanikhani and Kisi, 2012).

All possible combinations of number of membership functions (two to four) for two and three variable sets were developed under each membership function type. A total of 180 ANFIS models for each stream health measure were created using these options. Following determination of number of variables, membership function types and number of membership functions per variable, the ANFIS models were trained and tested.

Cross-validation was used to train, test, and select the best ANFIS model. Specifically, k-fold cross-validation was used because it is effective in situations where more data cannot be collected, prevents overfitting during model construction, and helps in best model selection (Mahmood and Khan, 2009). In k-fold cross-validation, the dataset is randomly split into k mutually exclusive subsets (folds) of approximately equal size (Kohavi, 1995). Here, ten folds were used, which is common practice in many fields (Mahmood and Khan, 2009). The ANFIS models (based on MF type and number of MFs per variable) were trained on nine folds (90% of the data) and tested on the remaining fold (10% of the data). This process was repeated ten times, with one fold being removed from model training and used for testing each time. Unique folds were created for each distinct MF type and number of MFs per variable combination. Two performance measures, coefficient of determination (R^2) and root-mean-square error (RMSE), were calculated for each testing dataset and averaged across the ten ANFIS models. The highest average R^2 and lowest average RMSE was used to select the best ANFIS model.

Following selection of the best general ANFIS model (MF type and number of MFs per variable), the best individual model from ten-fold cross validation was selected for each stream health measure and stream grouping. Final best model selection was accomplished by testing each model on all ten sets of test data (10% of the data) (Hamaamin et al., 2013). When predicted values were beyond the minimum or maximum of the stream health measures they were adjusted to the minimum or maximum values (Nejadhashemi et al., 2007).

3. Results and discussion

3.1. SWAT model calibration and validation

Results of the calibration and validation are presented in Table 1. Time-series observed and simulated streamflows are presented in Fig. S2 of the Supplementary material. Based on the criteria described by Moriasi et al. (2007), the model was successfully calibrated and validated for all parameters and locations.

3.2. Stream grouping

Streams were clustered into two groups using the River Continuum Concept (stream orders 1–3 and orders 4–6) and k-means clustering on the hydrologic indices and pollutant concentrations data. Stream order and cluster groupings are presented in Fig. S3. The two clusters were expected to emulate the stream order grouping, where cluster 1 (C1) is similar to stream orders 1–3 and cluster 2 (C2) is similar to stream orders 4–6. Clustering was equivalent to stream order for 80% of the streams in the study area. Streams of orders 1–3 were classified as C1 at a rate of 80%, while streams of orders 4–6 were classified as C2 at a 75% rate. Therefore, there was a 20% difference between orders 1–3 and C1, and a 25% difference between orders 4–6 and C2.

The difference in clustering and stream order methods occurs in two key watershed locations. C2 extends further into the northwestern headwaters of the River Raisin watershed, classifying many order 1–3 streams as C2. This also occurs for the watershed's other major river (the Saline River) in the northeastern headwaters. The clustering procedure identified the two major river systems in the watershed and grouped them based on their flow regime and water quality attributes.

However, because most of the variables are streamflow-driven, the similarity within each cluster is primarily due to similarities in flow parameters rather than water quality parameters.

Using k-means clustering to designate headwaters versus midreaches could be a useful alternative to the River Continuum Concept, because the River Continuum Concept does not take into account stream disturbances such as floods, (Junk et al., 1989) and hypothesizes an ideal ecological system (Statzner and Higler, 1985). By clustering streams based on their hydrologic and pollutant concentration characteristics, a more realistic grouping of similar streams may be produced for later use in variable selection and model development.

3.3. Variable selection

Variable selection was completed for each method (Spearman's rank correlation, Bayesian variable selection, and multiple PCA variable sets), stream health measure (EPT Taxa, FIBI, HBI, and IBI), and stream grouping (all streams, stream orders 1–3, stream orders 4–6, cluster 1, and cluster 2). The top three variables selected using these methods are presented in Table 2 for all streams. Variable selection for the remaining stream groupings are presented in Table S2 through Table S5 of the Supplementary materials.

3.3.1. Variable selection by method

Unique variable sets were selected for each stream health measure using Spearman's rank correlation and Bayesian variable selection in all stream groups. Variable selection using PCA (PC1: top three variables from the first PC, and PC1 to PC3: top variable from each of the first three PCs) is the same regardless of stream health measure because PCA is a non-dependent procedure (a response variable is not specified). Individual variables extracted from PCA are presented with their loading scores, while the percentage of variation explained is presented for the top three principal components. Spearman's ρ (correlation with stream health measure) is presented for Spearman's rank correlation. All selected variables have a significant ρ at α = 0.05. Selectivity (percentage of times variable was selected for model building) is presented for Bayesian variable selection. Component loading (correlation of each variable with its principal component) is presented for both PCA methods.

Diverse variable sets were identified for all variable selection methods (Table 2 and Tables S2 through S5). Spearman's rank correlation and Bayesian variable selection produced multiple unique variable classifications for each stream health index and grouping. For example, under all streams for HBI (Table 2), sediment concentration (SED), duration (DH), and rate of change (RA) variables were selected. Rate of change/flashiness variables are consistently of high importance across all stream groupings and stream heath indices for these two variable selection methods. Among this variable set, RA5 (number of days in which flow is greater than the previous day) and RA9 (variability in flow reversals) are the most common. Individual selectivity of variables in Bayesian variable selection is low in some cases, such as for FIBI (Table 2). Given the low selectivity of the most important variables, it is possible that the resultant ANFIS created from these variables will have low predictive power. However, for other stream health measures (IBI and HBI in Table 2) the highest ranked variables have high selectivity, which may lead to better predictions in the ANFIS models.

While resultant variable sets from Spearman's rank correlation and Bayesian variable selection were diverse, this was not the case for the PCA methods. Individual variables extracted from PCA consistently include related average (MA), low (ML), and high magnitude (MH) hydrologic indices (Table 2). This is not surprising when selecting the variables within PC1, because their high loading indicates high correlation with PC1 and with each other. Magnitude variables (MA, ML, and MH) also had highest loading in PC1–PC3, indicating that magnitude variables explain most of the dataset's variation. Of the principal components themselves, the sum of their explained variance is 58% for all

streams, while it is lower for orders 1–3 and C1 (54% and 55%, respectively) and higher for orders 4–6 and C2 (84% and 74%, respectively). The lower percentage of explained variance for all streams, orders 1–3, and C1 indicates that these variables (PC1, PC2, and PC3) may not adequately predict stream health because they do not capture enough of the dataset's variance.

3.3.2. EPT taxa variable selection

Variables selected for EPT taxa models were diverse, although magnitude and rate of change variables were commonly included in most stream groupings for Spearman's rank correlation and Bayesian variable selection. Pollution concentrations selected through Spearman's rank correlation exhibit negative correlation with EPT taxa: higher concentrations result in reduced presence of sensitive macroinvertebrates, as exhibited for all streams (Table 2), orders 4–6 (Table S3), and C2 (Table S5). The decline in EPT taxa can be attributed to increases in oxygen depletion due to phosphorus enrichment and eutrophication. In an indirect manner, large sediment and phosphorus loads in streams are related due their transport mechanisms. This increase in turbidity reduces light penetration, which is vital to primary producers that are a major food source for macroinvertebrates (Richards et al., 1993). A single frequency variable was also commonly identified using Bayesian variable selection (FH10: number of flows above the median of the annual minima) and was also positively correlated with EPT taxa, demonstrating that consistent flows above annual minimums are beneficial to sensitive macroinvertebrates.

3.3.3. FIBI variable selection

Variables important to FIBI often characterize magnitude and nitrogen concentrations. Magnitude variables are important for all streams (Table 2), orders 1–3 (Table S2), and C1 (Table S4). In the grouping of larger streams such as those in orders 4–6 seasonal nitrogen variables are commonly selected (Table S2). As expected, higher nutrient concentrations decreased macroinvertebrate integrity. Low flow duration (DL) was deemed important for FIBI in orders 1–3 (Table S2) and C1 (Table S4) through Spearman's rank correlation. Here, there is a positive correlation between magnitude of low flow duration (DL2 and DL4) and FIBI; longer periods of extremely low magnitude flow are detrimental to macroinvertebrates. Meanwhile, in streams of orders 4–6 (Table S3), flood frequency (FH7 — average number of flow events above seven times the median flow) is highly negatively correlated with FIBI, indicating that increased frequency of extreme floods impacts macroinvertebrate integrity. In a predominantly agricultural watershed with limited floodplain connectivity, as observed in the study area, these events are expected to occur with greater frequency because water is not able to adequately dissipate onto a floodplain. A river's connection to a floodplain during high flow conditions maintains productivity and diversity (Poff et al., 1997), but this is inhibited by land use modification in the watershed and may result in lower FIBI.

3.3.4. HBI variable selection

Variables selected for HBI models prominently include sediment and nutrient concentrations, which is unsurprising given that the index is a measure of macroinvertebrate pollution tolerance. Of the pollution concentration variables related to HBI, seasonal nitrogen concentrations are the most common. The Spearman ρ values reveal that as pollution concentrations increase, HBI increases, which indicates more degraded stream health. Rate of change variables are also prevalent in most stream groupings for Spearman's rank correlation and Bayesian variable selection. For example, RA5 (number of days when flow is greater than the previous day) is negatively correlated ($\rho = -0.50$, Table 2). When streamflow magnitude increases regularly, pollution concentrations decrease and HBI improves.

3.3.5. IBI variable selection

Variables important for IBI were the most diverse of all stream health measures. For example, magnitude, duration, timing, rate of change, and seasonal nitrate concentration variables are all selected as important for all streams (Table 2), higher order streams (Table S3), and C2 (Table S5). Meanwhile, in orders 1–3 (Table S2) and C1 (Table S4), long-term average magnitude variables are the most commonly selected. DH22 (average of annual median of number of days between floods with a recurrence interval of 1.67 years) is selected by the Bayesian method in multiple instances (all streams, orders 4–6, and C2). Variability in average spring and autumn flow magnitudes (long-term MA variables) are particularly important for smaller streams (orders 1–3 and C1). For Spearman's rank correlation, greater variability in average monthly flows for spring and autumn months correlates with improved stream health. The importance of long-term stream magnitude variability is unsurprising as it often dictates ecosystem function and biodiversity (Poff et al., 1997).

In addition, the prevalence of rate of change (RA) variables in all stream groupings demonstrates that short-term variability in streamflow plays a role in fish community health. In general, larger streams (orders 4–6) with greater rates of change (magnitude over time components), as observed through RA variables, generally have comparatively worse biotic integrity (Table S3). This is consistent with the concept of short-term variability as a determinant of a stream's physical habitat, which itself is a determinant of biotic composition (Naiman et al., 2008).

Overall, different variable sets were selected by these methods, although their input data was the same. This is because each variable selection method is unique in their statistical/mathematical framework. We suggest that incorporating expert knowledge as a first step in reduction of variables (prior to using the variable selection methods) that are known to be unimportant may further improve selection of the most important variables.

3.4. Best model selection

3.4.1. Best models for each variable selection method

Best ANFIS models for each stream health measure and variable selection method are presented in Table 3 for all streams. The best models for other stream groupings are presented in Tables S6 through S9 of the Supplementary materials. Within each table the individual best variable selection method/membership function characteristics are highlighted for each stream health measure. Best models were determined by the lowest average RMSE and highest average R^2 across ten check datasets. ANFIS models are denoted by the MF type, number of variables, and number of membership functions per variable. For example, the best Bayesian variable selection ANFIS model for HBI in Table 3 is GaussC (2/3), which indicates the Gaussian MF type with two variables, where the first variable has two MFs and the second variable has three MFs. This corresponds to the variables NO_3son and NO_3djf in Table 2 for HBI Bayesian variable selection. ANFIS models were not built for IBI under C1 and orders 1–3 because the fish sampling dataset was not large enough to build ANFIS models or perform model checking. In these cases there were nine IBI observations, where the number of modifiable parameters for all possible ANFIS models was always greater than this.

Table 1
River Raisin watershed calibration and validation results.

Parameter (station)	Calibration			Validation		
	NSE	RSR	PBIAS	NSE	RSR	PBIAS
Streamflow (04175600)	0.54	0.68	14.6%	0.47	0.73	1.6%
Streamflow (04176000)	0.64	0.60	14.5%	0.46	0.73	11.1%
Streamflow (04176500)	0.76	0.49	13.4%	0.61	0.62	5.7%
Sediment (580046)	0.57	0.65	24.1%	0.50	0.71	−18.7%
NO_3 (580046)	0.97	0.78	11.5%	0.58	0.64	31.5%
NO_2 (580046)	0.60	0.63	−36.1%	0.58	0.65	13.9%
TP (580046)	0.78	0.47	−0.9%	0.58	0.65	−31.5%

Table 2
Variable selection for all streams.

Method		EPT Taxa	FIBI	HBI	IBI
Spearman (ρ)		MA4 (-0.67)	MH24 (-0.64)	SEDjja (0.64)	TA1 (0.82)
		OrgPson (-0.61)	ML2 (0.54)	DH14 (0.57)	MA34 (-0.81)
		MA26 (-0.58)	DL11 (0.51)	RA5 (-0.50)	NO$_3$djf (-0.72)
Bayesian (selectivity)		FH10 (0.49)	MA3 (0.12)	NO$_3$son (0.65)	DH22 (1.00)
		MA35 (0.28)	MA35 (0.11)	NO$_3$djf (0.16)	TL3 (0.99)
		RA5 (0.20)	MA39 (0.11)	RA8 (0.14)	RA9 (0.98)
PCA	PC1 (loading)	ML2 (0.75)	ML2 (0.75)	ML2 (0.75)	ML2 (0.75)
		MA45 (-0.53)	MA45 (-0.53)	MA45 (-0.53)	MA45 (-0.53)
		ML1 (0.28)	ML1 (0.28)	ML1 (0.28)	ML1 (0.28)
	PC1 to PC3 (loading)	ML2 (0.75)	ML2 (0.75)	ML2 (0.75)	ML2 (0.75)
		MA28 (0.36)	MA28 (0.36)	MA28 (0.36)	MA28 (0.36)
		ML13 (-0.45)	ML13 (-0.45)	ML13 (-0.45)	ML13 (-0.45)
	PC1 to PC3 (variation explained)	PC1 (30.0%)	PC1 (30.0%)	PC1 (30.0%)	PC1 (30.0%)
		PC2 (17.5%)	PC2 (17.5%)	PC2 (17.5%)	PC2 (17.5%)
		PC3 (10.9%)	PC3 (10.9%)	PC3 (10.9%)	PC3 (10.9%)

For detailed variable definitions see the abbreviations in the Supplementary materials.

Each variable selection method produced a best ANFIS model that varied among stream grouping and health measures. Models created using Bayesian variable selection and Spearman's rank correlation were often the best. However, all three PCA methods also produced best models, as in the case of IBI in Table 3. Best variable selection methods also differed between stream health measures. For example, PCA selection methods always produced the best method for IBI. Therefore, when developing IBI prediction models with ANFIS, PCA should be used as the variable selection method. Bayesian variable selection and Spearman's rank correlation were more successful for the macroinvertebrate measures. However, the success of Spearman's rank correlation coupled with the method's simplicity suggests it can be used as alternative to more complex methods such as Bayesian variable selection.

Membership function type selection for best models was generally diverse. All five membership function types were selected at least once under each variable selection method and stream health measure. Most membership function shapes were triangular for the macroinvertebrate measures across all stream groupings, while the shapes for fish IBI were either Gaussian or Gaussian composite (GaussC). The triangular membership function was selected for over half of the models across all stream groupings/variable selection methods for FIBI and HBI. Number of EPT taxa was the only stream health measure where linear (triangular and trapezoidal) and nonlinear (generalized bell, Gaussian, and Gaussian composite) membership function types were equally selected as the best across variable selection methods.

In terms of number of variables used, the two variable models overwhelmingly outperformed those with three variables. Further, these models generally used two membership functions per variable, resulting in relatively simplistic (less parameterized) models. This was true for all four stream health measures.

3.4.2. Overall best variable selection methods and models

Table 4 lists the final best ANFIS models for each stream grouping method. Most of the best models were built with Gaussian-type MFs, supporting the findings of Marchini (2011) that nonlinear MFs are more suitable for ecological problems. Stream health prediction was generally better when splitting the watershed into distinct groups. The best models in stream orders 1–3 performed better than the comparable C1 (lower RMSE and higher R^2), while model performance in C2 was generally superior to stream orders 4–6. Model performance under the all-stream set was typically worse than any of the clustering methods. This indicates the benefit of splitting streams into groups based on their characteristics prior to ANFIS model development. However, because models performed well for all grouping approaches, it is difficult to select one method over another (k-means clustering versus stream order) for grouping streams.

3.4.3. Combined stream grouping model performance

Best model predictions of all stream health measures are presented in Fig. S4 (all streams), Fig. S5 (C1 and C2), and Fig. S6 (orders 1–3 and orders 4–6). The result of combining the best performing clustering models and stream order models is a discrete modeling system comprised of two models. Each model in the system uses unique variables, MF shape, and number of MFs, and is applicable for different streams (based on stream order or cluster) in the watershed.

The stream grouping methods reveal distinct differences in the prediction models. For example, models built on clusters C1 and C2 separately for EPT taxa, FIBI, and HBI are generally distinct in stream health. Here, C1 corresponds to relatively worse stream health (lower EPT taxa and FIBI, higher HBI). This association is also present in the stream order models for macroinvertebrate measures, although it is not as pronounced. Given that the k-means clustered stream model performances are satisfactory in all cases, C1 models are good predictors of poorer stream health and C2 models generally correspond to better stream health.

Although IBI models were only built on the larger stream groups and for all streams, each model performed exceptionally well, with validation R^2 values reaching 0.99. Overall, the models built on all streams

Table 3
ANFIS model average performance across 10-folds, no stream grouping, for best model (MF type and number of variables) under each variable selection method.

Health measure	Method	MF type	Variables (MFs)	RMSE (train)	R^2 (train)	RMSE (check)	R^2 (check)
EPT Taxa	Spearman	Trapezoid	2 (2/2)	1.234	0.695	1.493	0.633
	Bayesian	Triangle	2 (2/2)	1.332	0.655	1.438	0.622
	PCA	Triangle	2 (2/2)	1.409	0.625	1.494	0.534
	PCA PC1	Triangle	2 (2/2)	1.528	0.534	1.596	0.473
	PCA PC1–3	GaussC[a]	2 (2/2)	1.304	0.662	1.428	0.644
FIBI	Spearman	Triangle	3 (2/2/2)	3.794	0.617	5.161	0.452
	Bayesian	Triangle	2 (2/2)	4.655	0.461	5.017	0.452
	PCA	Triangle	2 (2/2)	4.579	0.479	5.235	0.356
	PCA PC1	Bell[b]	2 (4/3)	3.859	0.584	4.945	0.463
	PCA PC1–3	Gaussian	2 (2/3)	3.797	0.630	5.390	0.451
HBI	Spearman	Gaussian	2 (2/2)	0.338	0.521	0.380	0.467
	Bayesian	Gaussian	2 (2/3)	0.294	0.628	0.412	0.374
	PCA	Bell[b]	2 (2/3)	0.378	0.242	0.408	0.447
	PCA PC1	Triangle	2 (2/2)	0.396	0.346	0.440	0.335
	PCA PC1–3	Triangle	2 (2/3)	0.365	0.454	0.437	0.369
IBI	Spearman	Gaussian	2 (3/3)	4.131	0.945	7.094	0.839
	Bayesian	GaussC[a]	2 (2/2)	10.612	0.742	12.41	0.621
	PCA	Gaussian	3 (3/3)	3.082	0.764	3.470	0.938
	PCA PC1	Gaussian	2 (2/4)	1.763	0.984	3.926	0.939
	PCA PC1–3	GaussC[a]	2 (2/3)	5.045	0.936	6.917	0.857

[a] GaussC: composite Gaussian MF.
[b] Bell: generalized bell MF.

Table 4
Final best models for each stream health measure and stream grouping.

Health measure	Group	Method	MF type	Vbls (MFs)	RMSE (train)	R^2 (train)	RMSE (check)	R^2 (check)
EPT Taxa	All	PC 1-3	GaussC[a]	2 (2/2)	1.327	0.678	1.266	0.673
	C1	Bayesian	GaussC[a]	2 (2/2)	0.990	0.660	0.959	0.677
	C2	PC 1	Triangle	2 (2/2)	0.905	0.810	0.765	0.921
	Orders 1–3	Spearman	Gaussian	2 (3/2)	0.877	0.850	0.760	0.824
	Orders 4–6	Bayesian	Triangle	2 (2/3)	0.836	0.836	0.776	0.697
FIBI	All	Bayesian	Triangle	2 (2/2)	4.699	0.474	4.529	0.485
	C1	Bayesian	Gaussian	2 (2/4)	2.539	0.720	2.423	0.728
	C2	PC 1-3	Gaussian	2 (2/2)	3.204	0.719	2.781	0.854
	Orders 1–3	Bayesian	Triangle	2 (2/2)	4.125	0.520	3.616	0.654
	Orders 4–6	PCA PC1	Gaussian	2 (2/3)	3.322	0.694	2.838	0.798
HBI	All	Spearman	Gaussian	2 (2/2)	0.342	0.543	0.327	0.563
	C1	Spearman	Gaussian	2 (2/2)	0.317	0.492	0.298	0.601
	C2	Bayesian	Triangle	2 (2/2)	0.245	0.702	0.228	0.751
	Orders 1–3	Bayesian	Triangle	2 (2/2)	0.321	0.655	0.304	0.620
	Orders 4–6	Bayesian	Gaussian	2 (2/2)	0.271	0.601	0.235	0.598
IBI	All	PC 1	Gaussian	2 (2/4)	1.477	0.996	1.209	0.996
	C2	PCA	Gaussian	2 (3/2)	1.676	0.994	1.579	0.993
	Orders 4–6	PC 1	GaussC[a]	2 (2/2)	1.912	0.993	1.834	0.995

[a] GaussC: composite Gaussian MF.

generally displayed more dispersion from the 1:1 prediction line than models built on stream order or *k*-means clustering.

3.4.4. Statistical differences

Wilcoxon signed-rank tests showed that there were no statistically significant differences ($\alpha = 0.05$) between observed datasets and any of the stream health measure/stream grouping combinations. Therefore, the ability to correctly predict stream health class was used to determine which stream grouping method was superior. Stream health classes are defined as excellent, good, fair, poor, and very poor (Table S1).

3.4.5. Stream health class prediction

Correct prediction of stream health class is important for effective modeling of stream health because these qualitative descriptors are often used to communicate with watershed stakeholders and natural resources managers. The number of correct stream health class
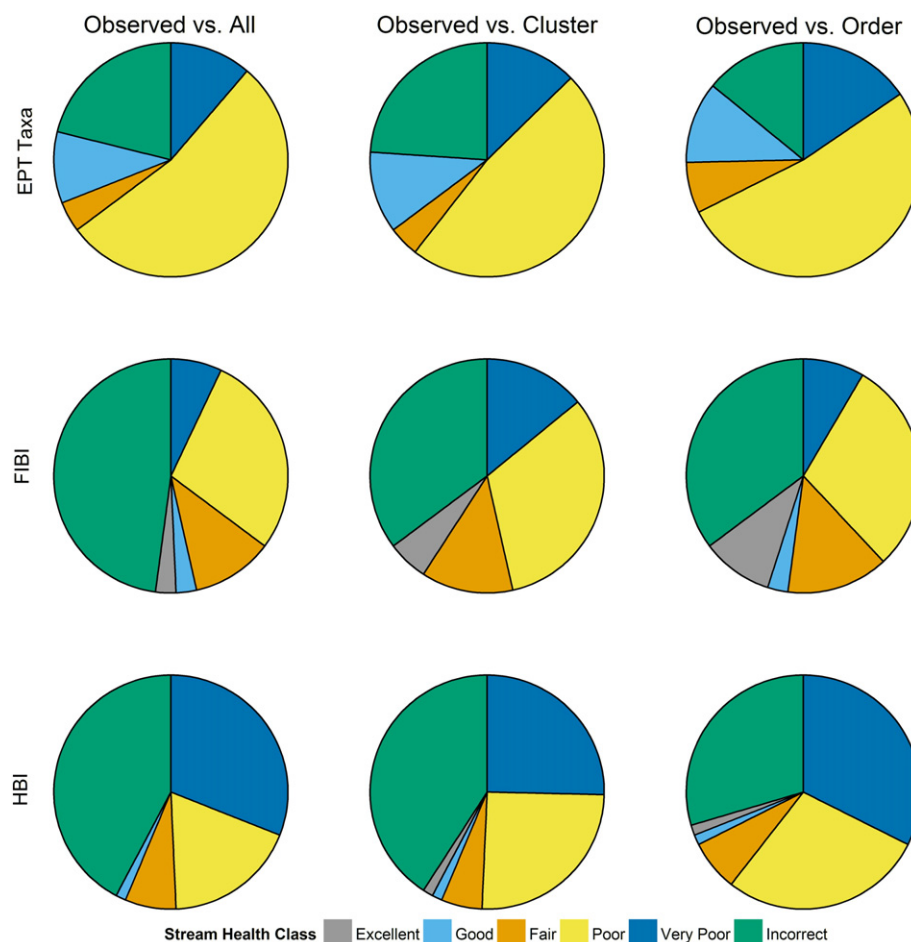

Fig. 3. Model predictions compared with observed data. Each pie chart indicates the proportion of model predictions that correctly predict the stream health class (excellent, good, fair, poor, and very poor) and the number of incorrect class predictions. Rows represent macroinvertebrate stream health measures and columns represent stream groupings.
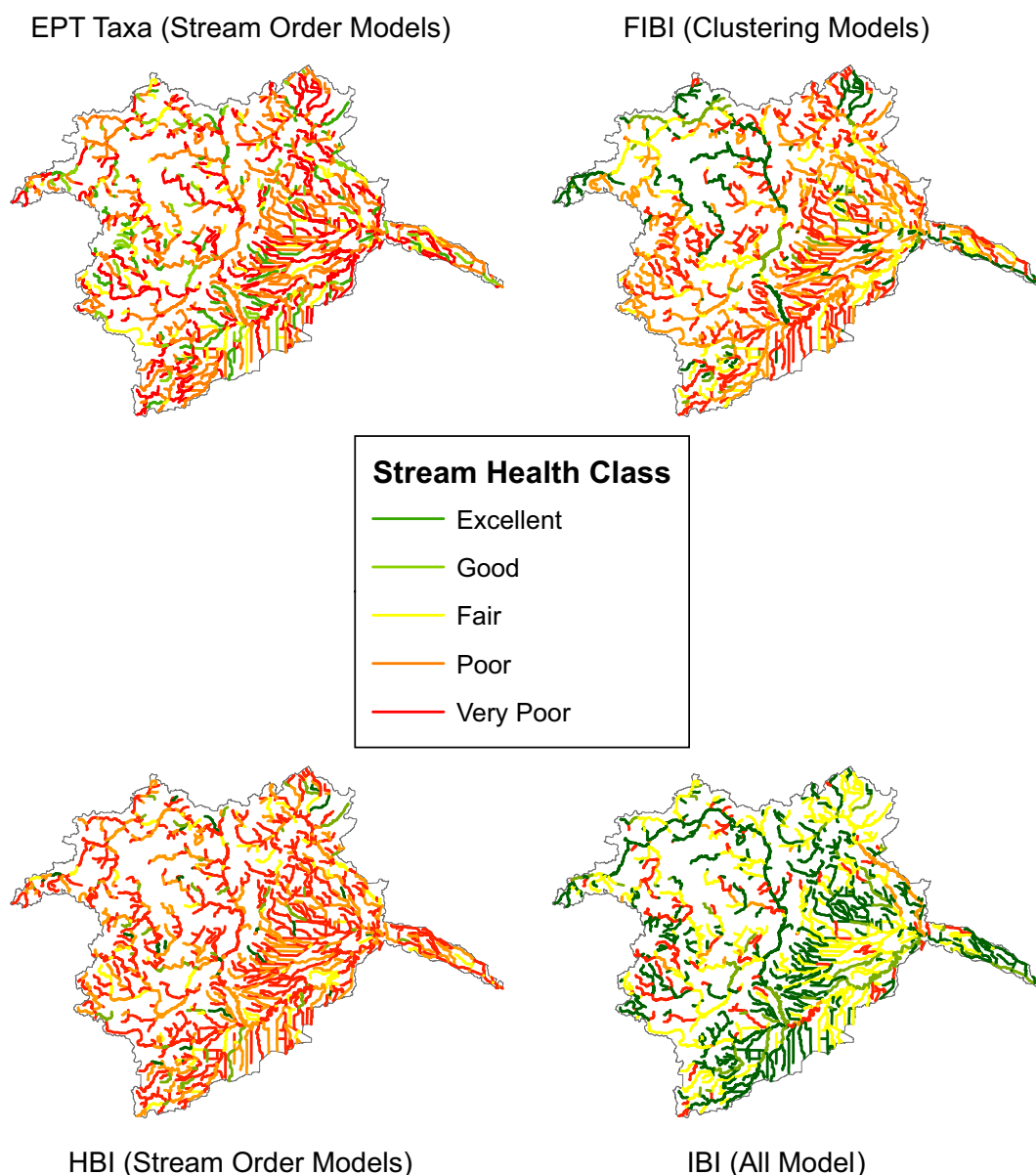
EPT Taxa (Stream Order Models)   FIBI (Clustering Models)

Stream Health Class
—— Excellent
—— Good
—— Fair
—— Poor
—— Very Poor

HBI (Stream Order Models)   IBI (All Model)

Fig. 4. River Raisin watershed stream health classes for each measure and best model.

predictions and total number of incorrect class predictions is presented for each macroinvertebrate measure and stream grouping in Fig. 3. EPT taxa models had the lowest number of incorrect class predictions. In contrast, the non-grouped models of FIBI, HBI, and HBI incorrectly predicted the stream health class for about 40% of the observed data, which corresponds with the relatively worse RMSE and R^2 of these models (Table 4). Both grouping methods correctly predicted class more frequently than the models built without stream grouping. Based on the number of incorrect classifications, stream order grouping models should be selected for EPT taxa and HBI modeling. Stream order and k-means clustering classify the same number of data point incorrectly, but the clustered models have superior performance measures. Therefore, clustering models were selected to predict FIBI for all streams in the study area.

3.5. Watershed stream health

Using the best models for EPT taxa, FIBI, HBI, and IBI, health class was predicted for all streams in the River Raisin watershed (Fig. 4).

Percentages of each stream health class are presented in Fig. S7. The macroinvertebrate measures indicate that a majority of the watershed has "poor" or "very poor" stream health, especially in the southeastern part of the watershed where most land has been converted to agriculture. This is apparent in the organic pollution-sensitive HBI where most of the watershed is classified as "very poor", due to nitrogen and phosphorus sourced from agriculture. The outlet of the River Raisin extending to Lake Erie is an EPA Area of Concern with nonpoint source pollution control issues, so these conditions are expected. Most occurrences of "fair" to "excellent" stream health are in the northern headwaters of the study area, where much of the land is still forested.

Conversely, much of the watershed is classified as "excellent" to "fair" when examined using fish IBI. The variables selected through PCA for developing the IBI model without stream grouping were based on average and low flow magnitudes. Although land use change has altered the flow regime in the past 150 years, the main impact of the land use change are more pronounced in high magnitude events. For example, higher peak streamflows occur due to agricultural expansion and urbanization (Niehoff et al., 2002; Hundecha and Bardossy,

2004; Nejadhashemi et al., 2012). Therefore, the results may be misleading due to lack of inclusion of high magnitude flow variables in the IBI models. Meanwhile, Einheuser et al. (2013b) found that fish measures were more sensitive to changes in long-term flow magnitudes, while macroinvertebrate measures were more sensitive to smaller changes in water quality on shorter time-scales. The difference in sensitivity is likely why IBI scores are better than the macroinvertebrate measures.

4. Conclusions

The goal of this research was to test multiple selection techniques to identify influential variables for use in the development of stream health models. This research lays the foundation for using data-driven methods to select influential in-stream variables for stream health prediction models. Although variables identified for model development differed between variable selection methods, stream health measures, and stream groupings, the Bayesian method consistently performed the best. When developing ANFIS models, grouping streams by k-means clustering or stream order improved model predictability compared to the watershed-level stream health models. However, the k-means clustering method commonly used in data mining is recommended because it forms stream groups based on the data characteristics of each stream while reproducing general stream order trends.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2014.12.066.

References

Adriaenssens, V., Baets, B.D., Goethals, P.L.M., Pauw, N.D., 2004. Fuzzy rule-based models for decision support in ecosystem management. Sci. Total Environ. 319 (1–3), 1–12.

Adriaenssens, V., Goethals, P.L.M., De Pauw, N., 2006. Fuzzy knowledge-based models for prediction of Asellus and Gammarus in watercourses in Flanders (Belgium). Ecol. Model. 195 (1–2), 3–10.

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. 35, 257–284.

Allan, D., Erickson, D., Fay, J., 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshw. Biol. 37 (1), 149–161.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: Model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish. USEPA, Washington, DC (Report No.: EPA 841-B-99-002).

Bryce, S.A., Lomnicky, G.A., Kaufmann, P.R., McAllister, L.S., Ernst, T.L., 2008. Development of biologically based sediment criteria in mountain streams of the western United States. N. Am. J. Fish Manag. 28 (6), 1714–1724.

Chen, Q., Mynett, A.E., 2003. Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. Ecol. Model. 162 (1), 55–67.

Clapcott, J.E., Collier, K.J., Death, R.G., Goodwin, E.O., Harding, J.S., Kelly, D., et al., 2012. Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity: stream integrity along land-use gradients. Freshw. Biol. 57 (1), 74–90.

Cobaner, M., 2011. Evapotranspiration estimation by two different neuro-fuzzy inference systems. J. Hydrol. 398, 292–302.

Compin, A., Céréghino, R., 2007. Spatial patterns of macroinvertebrate functional feeding groups in streams in relation to physical variables and land-cover in Southwestern France. Landsc. Ecol. 22 (8), 1215–1225.

Dodge, K.E., 1998. River Raisin Assessment. Michigan Department of Natural Resources, Fisheries Division, Ann Arbor, MI (Report No.: 23).

Einheuser, M.D., Nejadhashemi, A.P., Sowa, S.P., Wang, L., Hamaamin, Y.A., Woznicki, S.A., 2012. Modeling the effects of conservation practices on stream health. Sci. Total Environ. 435–436, 380–391.

Einheuser, M.D., Nejadhashemi, A.P., Wang, L., Sowa, S.P., Woznicki, S.A., 2013a. Linking biological integrity and watershed models to assess the impacts of historical land use and climate changes on stream health. Environ. Manag. 51 (6), 1147–1163.

Einheuser, M.D., Nejadhashemi, A.P., Woznicki, S.A., 2013b. Simulating stream health sensitivity to landscape changes due to bioenergy crops expansion. Biomass Bioenergy 58, 198–209.

Fellows, C.S., Clapcott, J.E., Udy, J.W., Bunn, S.E., Harch, B.D., Smith, M.J., et al., 2006. Benthic metabolism as an indicator of stream ecosystem health. Hydrobiologia 572 (1), 71–87.

Frimpong, E.A., Sutton, T.M., Engel, B.A., Simon, T.P., 2005. Spatial-scale effects on relative importance of physical habitat predictors of stream health. Environ. Manag. 36 (6), 899–917.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Trans ASABE. 50, 1211–1250.

Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160 (3), 249–264.

Hamaamin, Y.A., Nejadhashemi, A.P., Einheuser, M.D., 2013. Application of fuzzy logic techniques in estimating the regional index flow for Michigan. Trans. ASABE 56 (1), 103–115.

Hawkins, C.P., Norris, R.H., Hogue, J.N., Feminella, J.W., 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol. Appl. 10 (5), 1456–1477.

Henriksen, J.A., Heasley, J., Kennen, J.G., Nieswand, S., 2006. Users' Manual for the Hydroecological Integrity Assessment Process Software (Including the New Jersey Assessment Tools). United States Geological Survey, Biological Resources Discipline, p. 71 (Report No.: 2006-1093).

Hilsenhoff, W.L., 1988. Rapid field assessment of organic pollution with a family-level biotic index. J. N. Am. Benthol. Soc. 7 (1), 65–68.

Huang, Y., Lan, Y., Thomson, S.J., Fang, A., Hoffmann, W.C., Lacey, R.E., 2010. Development of soft computing and applications in agricultural and biological engineering. Comput. Electron. Agric. 71 (2), 107–127.

Hundecha, Y., Bardossy, A., 2004. Modeling the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. J. Hydrol. 292 (1–4), 281–295.

Jang, J.-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685.

Johnson, L.B., Host, G.E., 2010. Recent developments in landscape approaches for the study of aquatic ecosystems. J. N. Am. Benthol. Soc. 29 (1), 41–66.

Junk, W.J., Bayley, P.B., Sparks, R.E., 1989. The Flood Pulse Concept in river–floodplain systems. Can. Spec. Publ. Fish. Aquat. Sci. 106 (1), 110–127.

Karr, J.R., 1981. Assessment of biotic integrity using fish communities. Fisheries 6 (6), 21–27.

Karr, J.R., 1999. Defining and measuring river health. Freshw. Biol. 41 (2), 221–234.

Karr, J.R., Dudley, D.R., 1981. Ecological perspective on water quality goals. Environ. Manag. 5 (1), 55–68.

Karr, J., Yoder, C., 2004. Biological assessment and criteria improve Total Maximum Daily Load decision making. J. Environ. Eng. 130 (6), 594–604.

Kennard, M.J., Arthington, A.H., Pusey, B.J., Harch, B.D., 2005. Are alien fish a reliable indicator of river health? Freshw. Biol. 50 (1), 174–193.

Kennard, M.J., Pusey, B.J., Arthington, A.H., Harch, B.D., Mackay, S.J., 2006. Development and application of a predictive model of freshwater fish assemblage composition to evaluate river health in eastern Australia. Hydrobiologia 572 (1), 33–57.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence, pp. 1137–1143.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. 120 (2–3), 65–73.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (1), 39–52.

Lencioni, V., Maiolini, B., Marziali, L., Lek, S., Rossaro, B., 2007. Macroinvertebrate assemblages in glacial stream systems: a comparison of linear multivariate methods with artificial neural networks. Ecol. Model. 203 (1–2), 119–131.

Lermontov, A., Yokoyama, L., Lermontov, M., Machado, M.A.S., 2009. River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. Ecol. Indic. 9 (6), 1188–1197.

Lücke, J.D., Johnson, R.K., 2009. Detection of ecological change in stream macroinvertebrate assemblages using single metric, multimetric or multivariate approaches. Ecol. Indic. 9 (4), 659–669.

Lyons, J., 1992. Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin. United States Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN (Report No.: NC-149).

Mah, D.Y.S., Bustami, R.A., 2012. Conserving the land: the resilience of riparian wetlands and river channels by a fuzzy inference system. Sustain. Sci. 7 (2), 267–272.

Mahabir, C., Hicks, F., Fayek, A.R., 2006. Neuro-fuzzy river ice breakup forecasting system. Cold Reg. Sci. Technol. 46 (2), 100–112.

Mahmood, Z., Khan, S., 2009. On the use of k-fold cross-validation to choose cutoff values and assess the performance of predictive models in stepwise regression. Int. J. Biostat. 5 (1).

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Softw. 15 (1), 101–124.

Marchini, A., 2011. Modelling ecological processes with fuzzy logic approaches. In: Jopp, F., Reuter, H., Breckling, B. (Eds.), Modelling Complex Ecological Dynamics. Springer, Berlin Heidelberg, pp. 133–145.

Marchini, A., Facchinetti, T., Mistri, M., 2009. F-IND: a framework to design fuzzy indices of environmental conditions. Ecol. Indic. 9 (3), 485–496.

Maret, T.R., Konrad, C.P., Tranmer, A.W., 2010. Influence of environmental factors on biotic responses to nutrient enrichment in agricultural streams. J. Am. Water Resour. Assoc. 46 (3), 498–513.

Mathon, B.R., Rizzo, D.M., Kline, M., Alexander, G., Fiske, S., Langdon, R., et al., 2013. Assessing Linkages in stream habitat, geomorphic condition, and biological integrity using a generalized regression neural network. J. Am. Water Resour. Assoc. 49 (2), 415–430.

MDEQ, 1997. GLEAS Procedure #51 Survey Protocols for Wadable Rivers. Report No.: Fisheries Special Report 25. Michigan Department of Environmental Quality, Surface Water Quality Division, Ann Arbor, MI.

Metternicht, G., 2001. Assessing temporal and spatial changes of salinity using fuzzy logic, remote sensing and GIS. Foundations of an expert system. Ecol. Model. 144 (2–3), 163–179.

Mondy, C.P., Usseglio-Polatera, P., 2014. Using fuzzy-coded traits to elucidate the non-random role of anthropogenic stress in the functional homogenisation of invertebrate assemblages. Freshw. Biol. 59 (3), 584–600.

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.

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2009. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environ. Model. Softw. 24 (8), 982–993.

Mouton, A.M., Dedecker, A.P., Lek, S., Goethals, P.L.M., 2010. Selecting variables for habitat suitability of Asellus (Crustacea, Isopoda) by applying input variable contribution methods to artificial neural network models. Environ. Model. Assess. 15 (1), 65–79.

Moya, N., Hughes, R.M., Domínguez, E., Gibon, F.-M., Goitia, E., Oberdorff, T., 2011. Macroinvertebrate-based multimetric predictive models for evaluating the human impact on biotic condition of Bolivian streams. Ecol. Indic. 11 (3), 840–847.

Naiman, R.J., Latterell, J.J., Pettit, N.E., Olden, J.D., 2008. Flow variability and the biophysical vitality of river systems. Compt. Rendus Geosci. 340 (9–10), 629–643.

NASS, 2012. USDA National Agricultural Statistics Service Cropland Data Layer. Published Crop-Specific Data Layer [Internet]. (cited 2014 Feb 2, Available from:). http://nassgeodata.gmu.edu/CropScape/.

NED, 2014. USGS National Elevation Dataset [Internet]. (cited 2014 Feb 2, Available from:). http://ned.usgs.gov/.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2005. Soil and Water Assessment Tool Theoretical Documentation, Version 2005. USDA Agricultural Research Service, Temple, TX.

Nejadhashemi, A.P., Sheridan, J.M., Shirmohammadi, A., Montas, H.J., 2007. Hydrograph separation by incorporating climatological factors: application to small experimental watersheds. J. Am. Water Resour. Assoc. 43 (3), 744–756.

Nejadhashemi, A.P., Wardynski, B.J., Munoz, J.D., 2012. Large-scale hydrologic modeling of the Michigan and Wisconsin agricultural regions to study impacts of land use changes. Trans. ASABE 55 (3), 821–838.

Niehoff, D., Fritsch, U., Bronstert, A., 2002. Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. J. Hydrol. 267, 80–93.

Niemi, G.J., McDonald, M.E., 2004. Application of ecological indicators. Annu. Rev. Ecol. Evol. Syst. 35, 89–111.

NRCS, 2014. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey [Internet]. (cited 2014 Feb 2, Available from:). http://websoilsurvey.nrcs.usda.gov/.

Ocampo-Duque, W., Ferré-Huguet, N., Domingo, J.L., Schuhmacher, M., 2006. Assessing water quality in rivers with fuzzy inference systems: a case study. Environ. Int. 32 (6), 733–742.

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19 (2), 101–121.

Park, Y.S., Verdonschot, P.F.M., Chon, T.S., Lek, S., 2003. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. Water Res. 37, 1749–1758.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin Philos. Mag. J. Sci. 2 (11), 559–572.

Pinto, U., Maheshwari, B.L., 2011. River health assessment in peri-urban landscapes: an application of multivariate analysis to identify the key variables. Water Res. 45 (13), 3915–3924 (Jul).

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., et al., 1997. The natural flow regime. Bioscience 47 (11), 769–784.

Pont, D., Hughes, R.M., Whittier, T.R., Schmutz, S., 2009. A predictive index of biotic integrity model for aquatic-vertebrate assemblages of western U.S. Streams. Trans. Am. Fish. Soc. 138 (2), 292–305.

Richards, C., Host, G.E., Arthur, J.W., 1993. Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. Freshw. Biol. 29, 285–294.

Roth, N.E., Allan, J.D., Erickson, D.L., 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landsc. Ecol. 11 (3), 141–156.

Sanikhani, H., Kisi, O., 2012. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. Water Resour. Manag. 26, 1715–1729.

Seelbach, P.W., Wiley, M.J., 1997. Overview of the Michigan Rivers Inventory (MRI) project. Report No.: Fisheries Technical Report 97-3. Michigan Department of Natural Resources, Fisheries Division, Lansing, MI.

Sommerlot, A.R., Nejadhashemi, A.P., Woznicki, S.A., Giri, S., Prohaska, M.D., 2013. Evaluating the capabilities of watershed-scale models in estimating sediment yield at field-scale. J. Environ. Manag. 127, 228–236.

Sponseller, R.A., Benfield, E.F., Valett, H.M., 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshw. Biol. 46 (10), 1409–1424.

Statzner, B., Higler, B., 1985. Questions and comments on the River Continuum Concept. Can. J. Fish. Aquat. Sci. 42, 1038–1044.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Civ. Eng. 101, 1258–1262.

USEPA, 2009. National Water Quality Inventory: Report to Congress, 2004 Reporting Cycle. United States Environmental Protection Agency, Office of Science and Technology, Office of Water, Washington, DC (Report No.: EPA 841-R-08-001).

USEPA, 2011. A Primer on Using Biological Assessments to Support Water Quality Management. United States Environmental Protection Agency, Office of Science and Technology, Office of Water, Washington, DC (Report No.: EPA 810-R-11-01).

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37 (1), 130–137.

Waite, I.R., Brown, L.R., Kennen, J.G., May, J.T., Cuffney, T.F., Orlando, J.L., et al., 2010. Comparison of watershed disturbance predictive models for stream benthic macroinvertebrates for three distinct ecoregions in western US. Ecol. Indic. 10 (6), 1125–1136.

Wang, L., Brenden, T., Seelbach, P., Cooper, A., Allan, D., Clark Jr., R., et al., 2008. Landscape based identification of human disturbance gradients and reference conditions for Michigan streams. Environ. Monit. Assess. 141 (1–3), 1–17.

White, K.L., Chaubey, I., 2005. Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. J. Am. Water Resour. Assoc. 41, 1077–1089.

Whittaker, G., Confesor, R., Di Luzio, M., Arnold, J.G., 2010. Detection of overparameterization and overfitting in an automatic calibration of SWAT. Trans. ASABE 53 (5), 1487–1499.

Zadeh, L.A., 1965. Fuzzy sets. Inf. Control. 8 (3), 338–353.