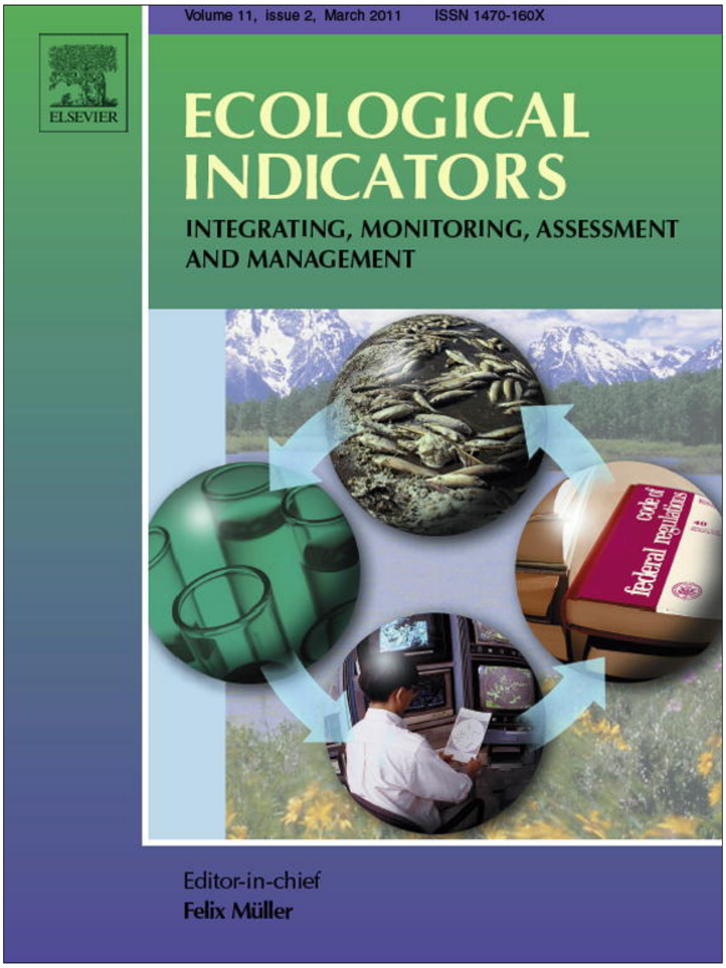
Volume 11, issue 2, March 2011
ISSN 1470-160X
ELSEVIER
ECOLOGICAL INDICATORS
INTEGRATING, MONITORING, ASSESSMENT AND MANAGEMENT
Editor-in-chief
Felix Müller

Ecological Indicators 15 (2012) 1–9

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

To threshold or not to threshold? That's the question

Song S. Qian[a,*], Thomas F. Cuffney[b]

[a] Nicholas School of the Environment, Duke University, Durham, NC 27708, USA.
[b] US Geological Survey, 3916 Sunset Ridge Road, Raleigh, NC 27607, USA

A R T I C L E   I N F O

Article history:
Received 2 March 2011
Received in revised form 11 August 2011
Accepted 12 August 2011

Keywords:
Bayesian statistics
Benthicmacroinvertebrates
Changepoint
Management threshold
Simulation
Urbanization

A B S T R A C T

The concept of evaluating multiple alternative models to determine ecological response form is over a century old and is ever more relevant as modern computing power allows ever more complicated models to be routinely used but often without a reasonable model verification process, particularly in fields where the ecological conceptual model is still developing. The emphasis for developing a statistical model is to test the validity of the hypothesis represented by the model. We present a framework of model identification and evaluation that includes exploratory data analysis and model diagnostics and evaluation. This framework emphasizes the importance of evaluating multiple alternative models when evaluating the validity of the model. This process is illustrated by using a model-building problem for quantifying the stream ecological response to urbanization using a data set from a large ecological study designed to understand how stream ecosystems respond to urbanization. The paper focuses on the question of whether a threshold model is appropriate, and demonstrates the importance of evaluating multiple alternative models in the detection of ecological thresholds, and illustrates how choosing an inappropriate model can lead to erroneous conclusion regarding the existence of thresholds.

1. Introduction

Statistics is a tool for inductive reasoning, and the emphasis of statistical modeling is on the discovery of the underlying causal relation that resulted in the observed data. A typical statistical modeling problem, however, is solved using a hypothetical deductive reasoning process in three steps: (1) model formulation – defining the probabilistic distribution of the response variable and characterizing the distribution by modeling the mean variable as a function of one or more predictor variables, (2) parameter estimation – estimating model parameters using available data, and (3) model interpretation – examining whether the fitted model can be interpreted using subject matter knowledge and justified based on the goodness-of-fit. These three steps correspond to the three tasks that Fisher defined for addressing statistical modeling problems (Fisher, 1922). The first step poses the hypothesis of the study and the subsequent steps test the hypothesis. Model formulation requires the interaction between ecological and statistical knowledge. Because the chosen model is assumed to be the true model, the parameter estimation process always leads to the optimal fit of the chosen model to the data. Consequently, assessing the validity of the model based solely on a model's fit can be ambiguous. Ambiguity can be reduced by proposing multiple models as competing alternatives. Comparison of these multiple alternative models can expose the weakness of models that would otherwise be masked by focusing only on the optimal model fit of a single model. This process is consistent with the multiple alternative working hypotheses approach recommended by Chamberlin (1890) when explaining new phenomena. This paper illustrates the importance of the multiple alternatives in ecological data analysis and modeling through the process of identifying the appropriate model form for describing the response of stream ecosystems to urbanization, with a focus on identifying a threshold response.

Although the theoretical value of the ecological threshold concept is still a topic of debate, its practical value in environmental management is attractive especially to managers of natural resources. As defined by Groffman et al. (2006), "an ecological threshold is the point at which there is an abrupt change in ecosystem quality, property or phenomenon, or where small changes in an environmental driver produce large responses in the ecosystem." Because of the complicated nonlinear dynamics of a threshold change and the multiple factors that can affect ecosystems, the detection and quantification of ecological thresholds is challenging. The question of detecting a threshold response is centered on determining whether or not a threshold exists. This can be addressed either through ecological theories or through empirical evidence in the form of statistical data analysis and modeling. This paper focuses on the process of assembling empirical evidence

* Corresponding author. Present address: Cardno ENTRIX, Inc., 5400 Glenwood Avenue, Suite G-03, Raleigh, NC 27612. Tel.: +1 919 239 8906.
E-mail addresses: mdqian@gmail.com (S.S. Qian), tcuffney@usgs.gov (T.F. Cuffney).

1470-160X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2011.08.019

for deciding the existence of a threshold. A statistical definition of ecological threshold is proposed (Section 2.1) and used to discuss general principles of model identification as applied to a threshold problem. Data from the US Geological Survey's (USGS) National Water Quality Assessment (NAWQA) program is used to illustrate the process of model selection and evaluation. These data were collected to study the effects of urbanization on stream ecosystems in nine metropolitan areas across the conterminous United States.

2. Methods

A model identification process is problem-specific. In this section, we present a class of simple threshold models and discuss a potential model diagnostic process.

2.1. Statistical definition of a threshold problem

When presenting a statistical model (e.g., a linear regression model), we often use the familiar form of the mathematical equation:

$$y = \beta_0 + \beta_1 x + \varepsilon, \tag{1}$$

where y is the response variable, x is the predictor variable, and β_0, and β_1 are model coefficients. The error term ε is assumed to follow a normal distribution with mean 0 and a constant variance, or $\varepsilon \sim N(0, \sigma^2)$. This equation is equivalent to a normality assumption of the response variable: $y \sim N(\beta_0 + \beta_1 x, \sigma^2)$. In general, a statistical model can be expressed as a probabilistic distribution about the response variable of interest (y) and the distribution is characterized by a parameter vector θ. For example, in the linear regression model of Eq. (1), the distribution of the response is assumed to be normal, with the mean modeled by a linear function of the predictor (or environmental stressor) x: $\mu = \beta_0 + \beta_1 x$. For this problem, $\theta = \{\beta_0, \beta_1, \sigma^2\}$. A general notation is $y \sim \pi(\theta, x)$ where π represent a generic distribution function. A statistical threshold exists when the distribution parameters change as the environmental stressor crosses a specific value (ϕ):

$$y \sim \begin{cases} \pi(\theta_1, x) & \text{if } x < \phi \\ \pi(\theta_2, x) & \text{if } x \geq \phi \end{cases} \tag{2}$$

where π is a probability distribution function parameterized by θ and environmental predictor variable x. Most (if not all) existing threshold models in the literature can be summarized in terms of Eq. (2). Quantitative options for estimating the threshold lie in the selection of the response variable distribution (π) and the determination of the dependency of the mean variable on one or more predictor variables (x). Different distributions often require very different computational methods, leading to numerous models in the literature. Consequently, when selecting a model, it is important to know the assumptions and conditions of the problem at hand. Both Bayesian and classical approaches can be used for parameter estimation. Model parameter estimation and model diagnostics for the class of linear threshold models, where the response variable y is assumed to have a normal distribution and the normal distribution mean is modeled by a linear function of the predictor, are discussed in the online supplementary materials. The linear class of threshold models includes the simple linear regression model, the piecewise linear (or the hockey stick) model, and the step function model as special cases.

2.2. The generalized linear threshold models

A frequently used benthic macroinvertebrate community indicator of stream ecosystem condition is EPT taxa richness (EPTr), which is the number of mayfly (*Ephemeroptera*), stonefly (*Plecoptera*), and caddisfly (*Trichoptera*) taxa in a sample. In the United States, EPTr is used by states (NCDE, 2006) and Federal agencies (Barbour et al., 1999) as a bio-indicator for evaluating water quality conditions. Because EPTr is a count variable, the Poisson distribution is often used to approximate its distribution:

$$\begin{aligned} y &\sim Pois(\lambda) \\ \log(\lambda) &= \beta_0 + \beta_1 x + \epsilon \end{aligned} \tag{3}$$

The error term $\epsilon \sim N(0, \sigma^2)$ is used to account for possible overdispersion. Based on the definition of Eq. (2), a threshold exists if model coefficients β_0, β_1 change along the gradient of predictor x. As in the linear class of threshold model, the Poisson threshold model can have three specific forms:

- The step function model

$$\log(\lambda) = \begin{cases} \beta_0 + \epsilon_1 & \text{if } x < \phi \\ \beta_0 + \delta + \epsilon_2 & \text{if } x \geq \phi \end{cases} \tag{4}$$

- The piecewise linear (or hockey stick) model:

$$\log(\lambda) = \beta_0 + (\beta_1 + \delta I(x - \phi))(x - \phi) + \epsilon \tag{5}$$

where $I(a)$ is a unit step function ($I(a) = 0$ when $a < 0$ and $I(a) = 1$ when $a \geq 0$), also known as the indicator function.

- The general model

$$\log(\lambda) = \begin{cases} \beta_0 + \beta_1 x + \epsilon_1 & \text{if } x < \phi \\ (\beta_0 + \delta_0) + (\beta_1 + \delta_1)x + \epsilon_2 & \text{if } x \geq \phi \end{cases} \tag{6}$$

The first question to address in the analysis is whether a threshold response is appropriate from an ecological perspective. If the answer is affirmative, the follow-up question is how to specify the model form. Selecting the correct model form is critical because a wrong model is likely to result in a threshold that is meaningless. If the underlying relationship is nonlinear, proper transformation of the predictor variable is necessary for the linear model to be useful.

2.3. Considerations in model evaluation

In the course of the investigation, a threshold response was initially hypothesized before data analysis. The study was, hence, designed to uncover the proper model form for the threshold response (Eqs. (4), (5), or (6)). A generalized linear model (without a change point) was also included as a contrast to the threshold response models.

Threshold analyses were conducted by fitting all three alternative models (general, hockey stick, and step) to the same data, and comparing the resulting threshold distributions (see online supplementary materials). The estimated threshold distributions are often useful for determining whether the model provides evidence for or against the existence of a threshold. Furthermore, comparison of model predictions with the data can be used to validate the model by determining if the predictions are consistent with the data. Comparisons of the estimated threshold distributions and model prediction form the basis for model selection and refinement. Some forms of the general model may be ecologically unrealistic, but the inclusion of the general model provides a mathematical check for the other two models. The step function and hockey stick models are special cases of the general model, if one of them fits the data well, the general model should also fit the data well with a similar change point estimate. If one of the two special case models fits the data well, while the general model does not or yields a different change point estimate, we have reasons to re-evaluate the model.

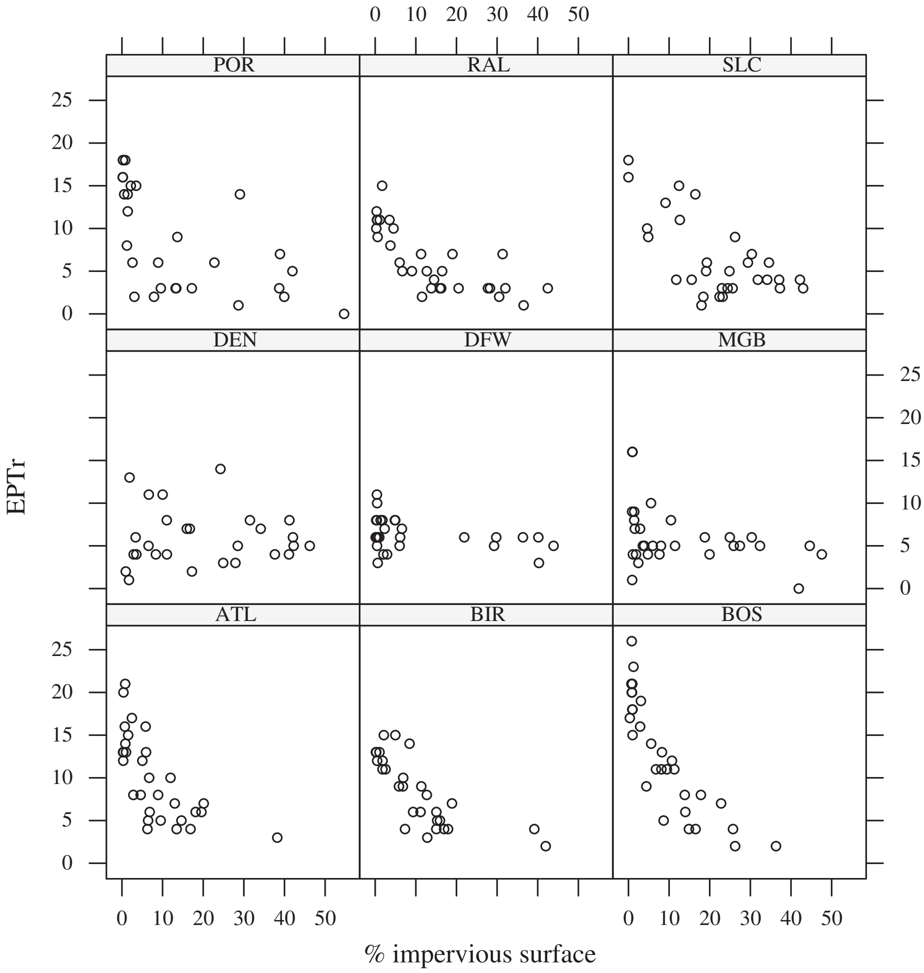

Fig. 1. Scatter plots of EPTr versus % impervious surface.

An unsettling feature of the step function model is that the model will "detect" a change point whether one exists or not, as long as the underlying function is monotonic. This problem can be partially addressed by applying a transformation of the predictor and/or the response variable. If the underlying relation can be approximated by the step function, a transformation of the predictor and/or the response variable will not affect the estimated change point. Otherwise, transforming the predictor may result in a different change point. As a result, transforming the predictor and/or the response variable should always be considered when the step function model is used. In general, we transform the predictor so that the transformed predictor values are distributed more less symmetric to avoid potential leverage data points. In this study, the predictor is a fraction variable (% impervious surface), which is often transformed using the logit function (the log ratio of % impervious surface over % pervious surface). The logit transformation is preferred over the other commonly used transformation (i.e., square root of arcsine) because the logit transformation can be easily interpreted. Although transforming the predictor cannot be used as a definite test for the validity of the step function, the use of transformation will likely provide useful information for model selection.

2.4. Computation

Parameters of the three threshold models can be estimated using classical statistics methods such as the maximum likelihood estimator. The R package `segmented` (Muggeo, 2003) or the hockey stick model described in Qian (2010) can be used for the piecewise linear model. The step function model can be analyzed using the two statistical methods described in Qian et al. (2003). However, to facilitate model comparisons, a consistent Bayesian computational

Table 1
Estimated change points (posterior modes) (% impervious surface).

Models	ATL	BIR	BOS	DEN	DFW	MGB	POR	RAL	SLC
Without logit transforming the predictor									
Step	6.35	9.11	3.59	1.93	37.83	1.38	2.48	5.8	16.84
Hockey	37.83	0.28	0.28	1.38	43.9	1.93	4.69	42.24	42.8
General	0.83	41.69	35.62	1.93	43.35	2.48	28.99	2.48	17.39
With logit transforming the predictor									
Step	6.35	9.11	3.58	1.92	0.26	1.37	2.48	5.24	16.84
Hockey	0.82	0.82	1.37	1.92	0.26	1.92	6.35	0.26	0.26
General	0.82	0.82	26.23	1.92	0.26	2.48	28.99	0.82	16.84

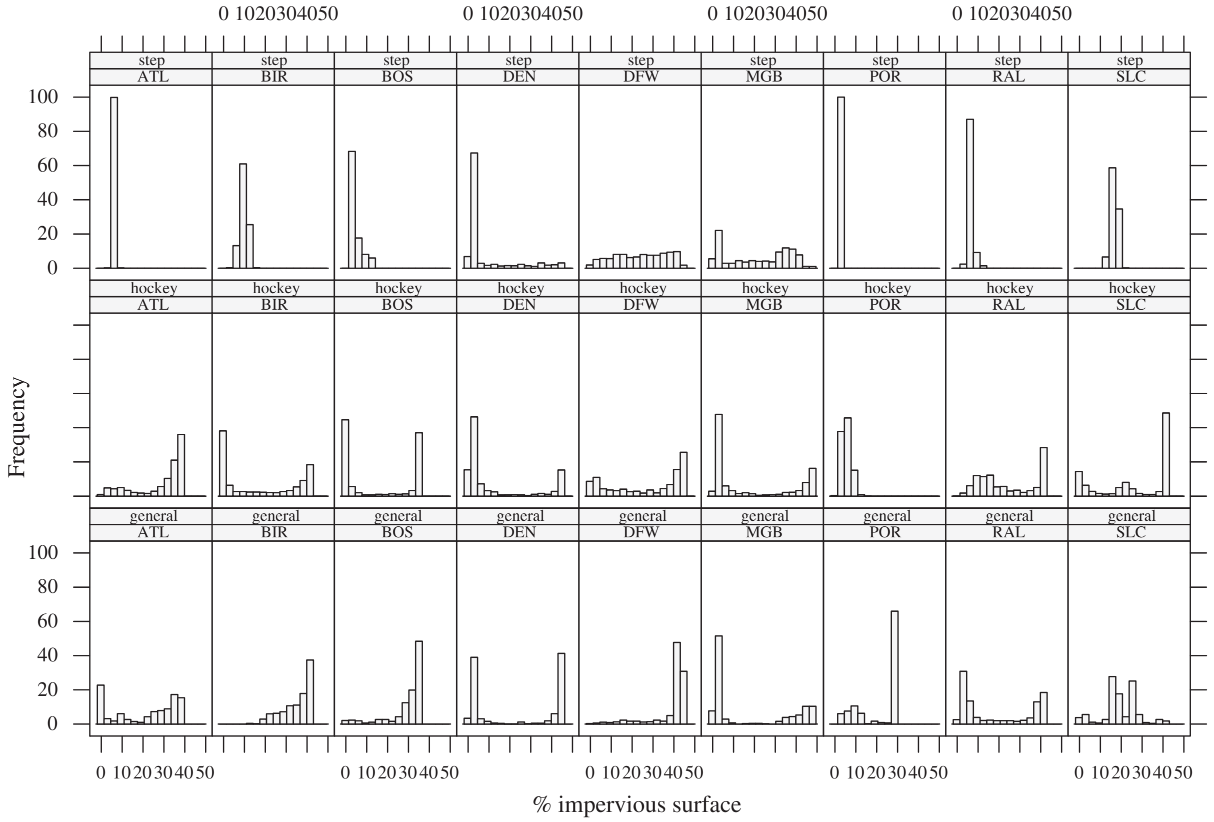

Fig. 2. Estimated threshold distributions from 9 metropolitan areas using 3 change point models: step function (step), hockey stick (hockey), and general (general).

method was used. This method used Markov chain Monte Carlo (MCMC) simulation to estimate model parameters. Details of the computational method are presented in the online supplementary materials.

3. Data

Data used in this paper were collected as part of the USGS NAWQA Program's study on the effects of urbanization on stream ecosystem (Couch and Hamilton, 2002). This study included 9 metropolitan regions in the continental US (Atlanta, GA [ATL]; Birmingham, AL [BIR]; Boston, MA [BOS]; Dallas-Fortworth, TX [DFW]; Denver, CO [DEN]; Milwaukee-Green Bay, WI [MGB]; Portland, OR [POR]; Raleigh, NC [RAL]; Salt Lake City, UT [SLC]) that represent a wide range of climate and geological conditions. Study watersheds in each region were selected along an urbanization gradient defined by a multimetric urban intensity index (Brown et al., 2009; Cuffney et al., 2010). However, for this study, the fraction of impervious surface is used as a measure of urban disturbance to facilitate comparisons with other urban studies. EPT taxa richness is used as a measure of ecological response.

4. Results

4.1. Exploratory data analysis

Scatter plots of EPTr against % impervious surface show a strong nonlinear decreasing pattern (Fig. 1). In addition, the variation of EPTr tends to decrease as % impervious surface increases. These scatter plots show a typical "wedge" shaped data cloud that is a common feature in many ecological data sets (Paul et al., 2009; Carter and Fend, 2005). The wedge shape is characteristic of count data because the variance of a count variable is usually proportional to its mean. This feature can also be interpreted from an ecological perspective, as a change in the influence of urban and non-urban variables that affect EPTr across the urbanization gradient. The variation of EPTr is high at low end of the urbanization continuum where a variety of other factors (e.g., background land cover condition) are more important than urbanization. Near the high end of the urbanization spectrum, the effect of urban development on EPTr dominates all other factors. These scatter plots also show that the predictor variable (% impervious surface) distribution is skewed to the left—more watersheds are clustered around the lower end of the urbanization continuum. This distribution reflects the difficulty in finding streams in areas with high imperviousness that met the criteria for inclusion in these studies: flowing above ground and connected to the riparian areas (e.g., no concrete lined channels). A highly skewed predictor may lead to a model that is unduly influenced by data points with large predictor variable values. The logit transformed % impervious surface is close to normality. As a result, we will present models with both transformed and untransformed predictors.

4.2. Threshold models and their fit to data

4.2.1. Untransformed predictor

The three threshold models (Eqs. 4 - 6) were fit to the data for each of the nine metropolitan areas. For each metropolitan area, the estimated posterior distributions of the threshold from the three models are compared (Fig. 2). The threshold posterior distribution is an important part of the model diagnostics (as illustrated in the online supplementary materials). A widespread (DFW, step) or a U- (BOS, hockey stick), L- (BOS, step), or J-shaped (BOS, general) posterior distribution suggests that either a threshold does not exist or a wrong model was used. These distributions are not indicative of a strong threshold because they allocate the posterior

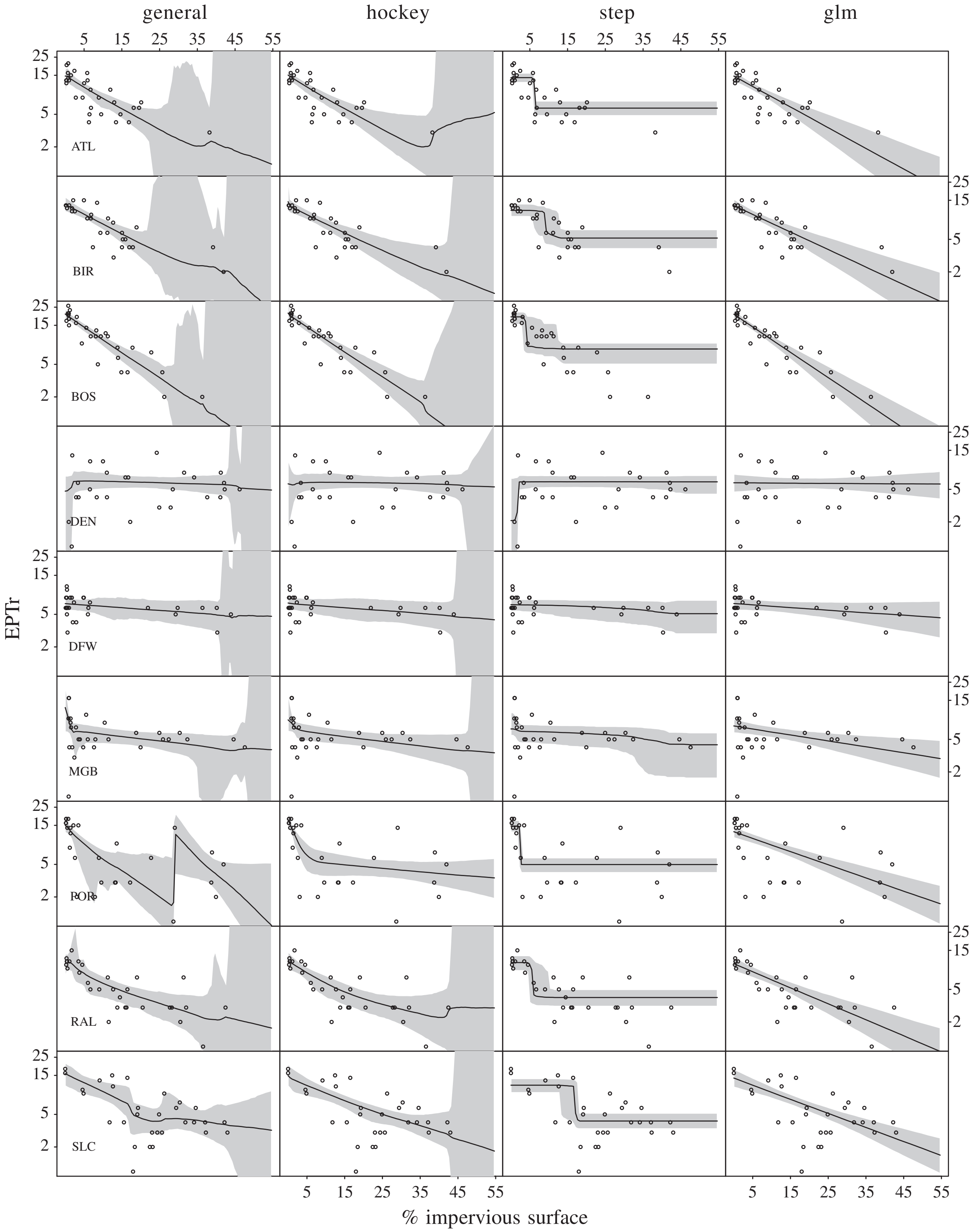

Fig. 3. Fitted change point models (general, hockey, and step function) are superimposed on the observed data and compared with the generalized linear model. The *y*-axis is EPTr (in logarithmic scale) and the *x*-axis is % impervious surface. Rows represent metropolitan areas (labeled). Columns represent models (left to right: general model, hockey stick model, step function model, and simple linear model).

change point mass at one (L-, or J-shaped) or both (U-shaped) ends of the gradient or uniformly across the gradient (widespread). A concentrated threshold distribution (e.g., RAL, step) often suggests the existence of a threshold, but additional evidence must be evaluated before the existence of a threshold can be confirmed (e.g., the posterior change point distribution estimated using the general model). Although the posterior distributions using the step function model show consistent concentrated peaks (8 of the 9 regions), locations of these peaks (i.e., the estimated change points) rarely coincide with the change points estimated using the general model

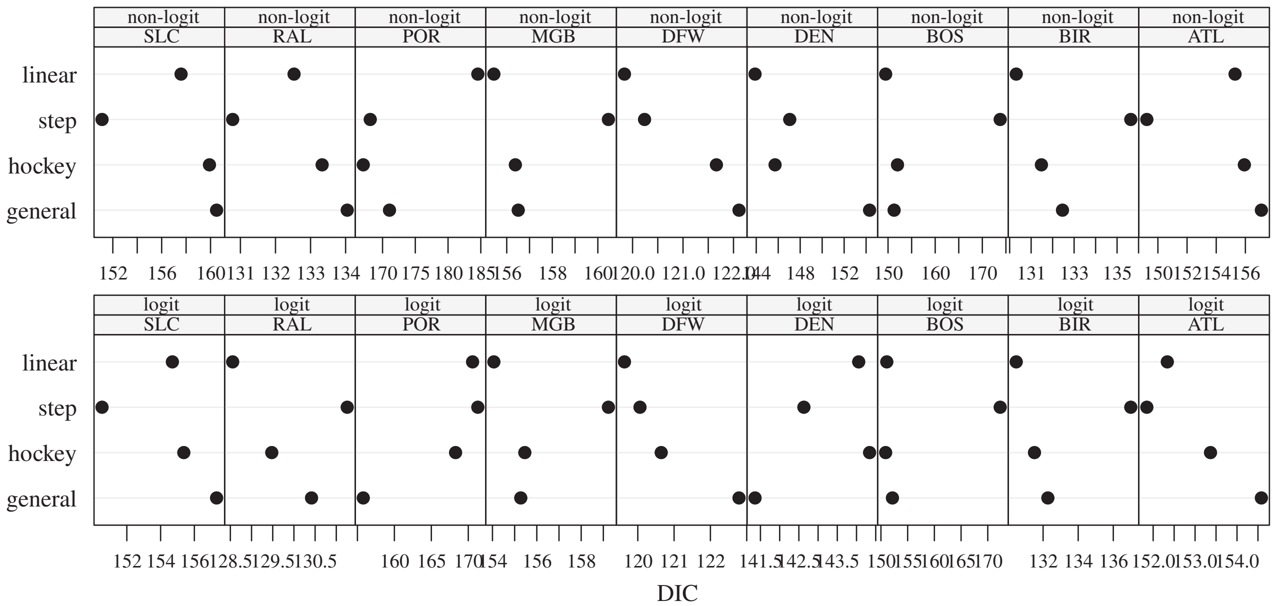

Fig. 4. Deviance information criterion (DIC) for the four competing models (top row, untransformed predictor; bottom row, logit transformed predictor).

(Table 1). The most effective means for checking a model's fit is the residual plot when the response variable is normally distributed. However, the residual distribution is not specified for a generalized linear model. For this reason, model fit was evaluated by comparing model predictions and the observations by superimposing model predicted inter-quartile range on the data plots (Fig. 3). The *y*-axes of plots in Fig. 3 (and in Fig. 6) are in the logarithmic scale as the Poisson model predicts the log-mean EPTr. In the MCMC simulation, the posterior model is characterized by the joint distribution of model coefficients represented by many sets (2500 in this analysis) of random samples of model coefficients drawn from the joint posterior distribution. Each set of random samples defines a possible model, and each possible model is used to make predictions of the mean EPTr along the urban gradient. Consequently, for each

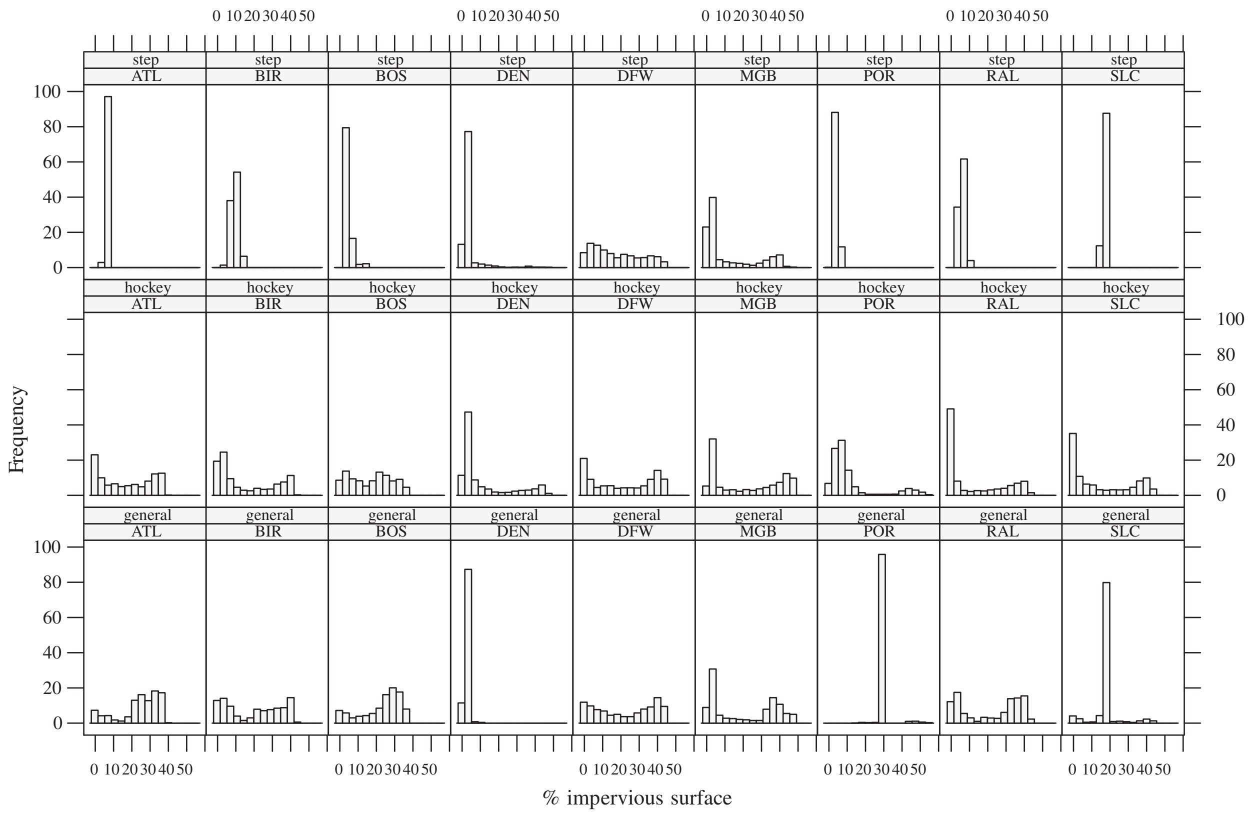

Fig. 5. Estimated threshold distributions from 9 metropolitan areas using 3 change point models: step function (step), hockey stick (hockey), and general (general). The predictor (% impervious surface) is logit transformed.

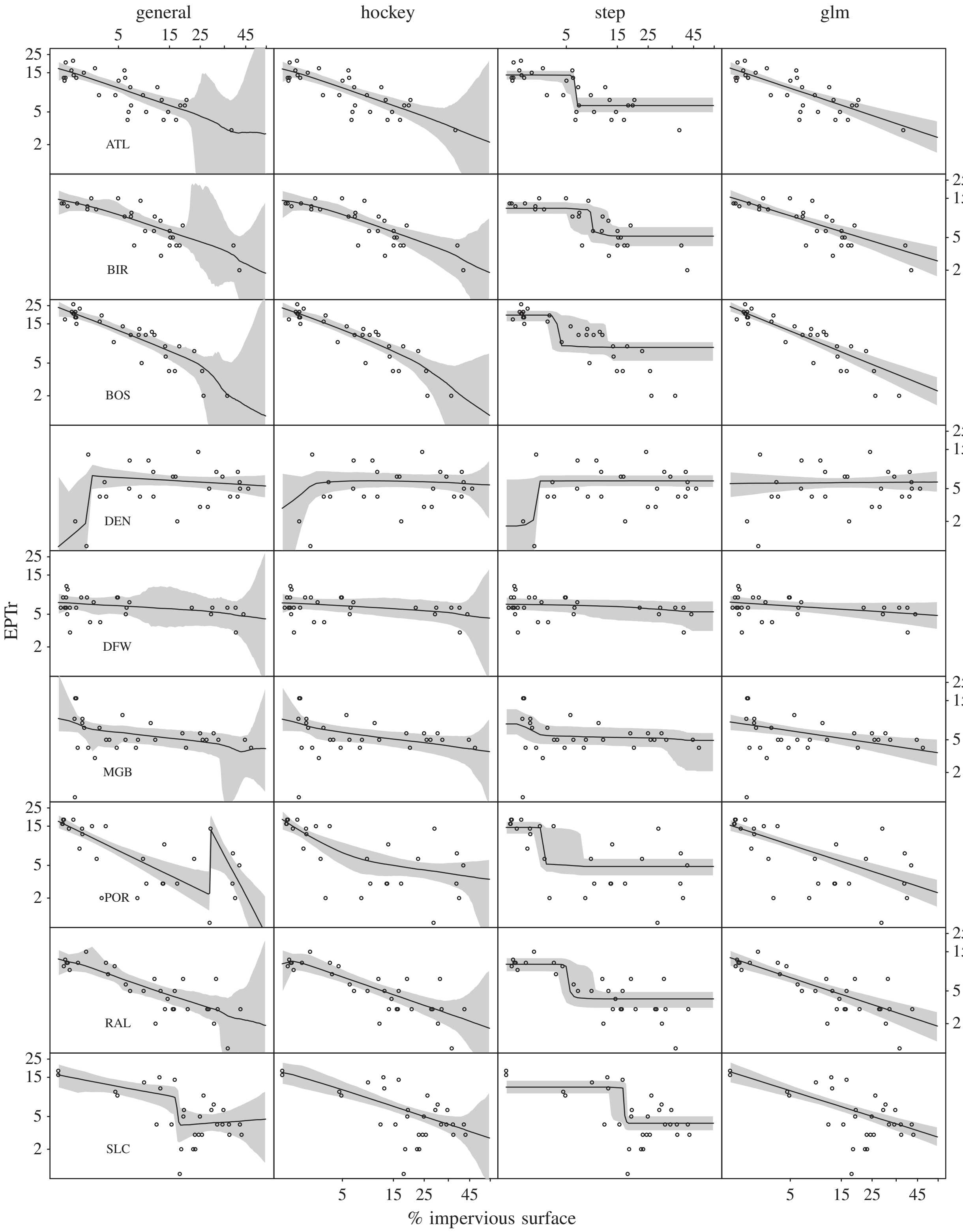

Fig. 6. Fitted change point models (general, hockey, and step function) are superimposed on the observed data and compared with the generalized linear model. The *y*-axis is EPTr (in logarithmic scale) and the *x*-axis is the logit transformed % impervious surface. Rows represent metropolitan areas (labeled). Columns represent models (left to right: general model, hockey stick model, step function model, and simple linear model).

predictor variable (% impervious surface) value, there are 2500 predictions. The shaded polygons in Fig. 3 are the middle 95% range of the predicted means and the solid line is the median of these 2500 predictions.

Fig. 2 suggests that thresholds may exist for some metropolitan areas (e.g., step function models for BIR and RAL) and that the threshold value varies from region to region. Although all three models resulted in a median line going through the middle of the data cloud (Fig. 3), interpretation of the resulting models can be challenging. For example, what would be the ecological explanation for the sudden increase in EPTr at ca. 30% impervious surface as indicated by the general model for POR? Visual assessment of the fit of the three threshold models and the generalized linear model offers no consensus on the appropriate model forms. Neither does the commonly used deviance information criterion (DIC) (Fig. 4) (Spiegelhalter et al., 2002; Qian et al., 2005). Consequently, we must be open to alternative, perhaps nonlinear, models.

4.2.2. Transformed predictor

The logit transformation implies a proportional relationship between the ratio of impervious over pervious surface and the mean EPTr. The slope (β_1 in Eq. (3)) is the proportional constant, that is, for every 1% increase in the ratio, we expect β_1% change in λ (the mean value of EPTr) (see Qian, 2010, pp. 255–257).

The estimated threshold posterior distributions after logit transforming the predictor (Fig. 5) are quite different from the distributions without the transformation (Fig. 2), although the estimated modes did not change dramatically because the departure from the linear model without the transformation is not very strong. The step function model indicates a potential threshold in most (8) of the metropolitan areas while the hockey stick and general models show a much lower number (1 or 2) of possible thresholds. However, comparison of the fitted models with the data does not support the thresholds identified by the step function models and suggests that a generalized linear model (with a logit transformation of the % impervious surface) would be sufficient (Fig. 6) for describing the EPTr response to urbanization. This generalized model implies a log–log linear relation between the mean EPTr and the ratio of impervious surface over pervious surface. That is, for every unit (1%) increase in the ratio, a fixed percentage decrease in EPTr mean is expected. This simple model fits the data well (Fig. 6, right column), providing evidence against the existence of a threshold.

5. Conclusions and discussion

This study revealed characteristics of the commonly used statistical change point methods that have important implications not only for detecting ecological thresholds, but more generally for ecological data analysis and modeling. Statistical inference is a form of hypothetical deduction. The quote that "all models are wrong" (Box, 1976) implies that models are only "correct" when the underlying assumptions about the data are correct. In other words, statistical inference is conditional on the proposed model. Verifying a model is often difficult because the estimated model is an optimal fit to the data. When alternative models are proposed, potential weakness of each candidate model can be identified. Whether a model is adequate or not must be carefully assessed before identifying and interpreting change points. Evaluating multiple alternative models is critical to finding a model that is appropriate for the data. We advocate evaluating multiple alternative models in the spirit of T.C. Chamberlin, who advocated the comparison of multiple working hypotheses as a means of developing rational explanations of new phenomena (Chamberlin, 1890).

In a threshold modeling problem, the multiple model approach is especially important because the three alternative threshold models can often produce a change point even when a threshold does not exist. This problem can be attributed either to random sampling error, unaccounted for confounding factors, and/or the use of an inappropriate model. For example, the step function model will always detect a distinct threshold as long as the underlying function is monotonic. Consequently, when the step function model and its variations are used, users must carefully evaluate

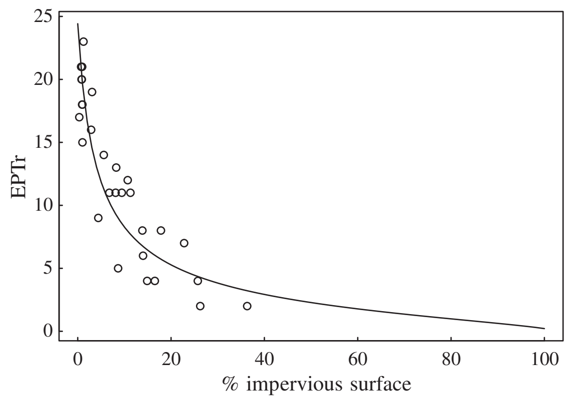

Fig. 7. The fitted GLM (solid line) is compared to the observed data from BOS.

their results against alternative models to avoid misleading results. Furthermore, Chiu et al. (2006) suggest that statistical evidence is often unavailable for distinguishing a hockey stick model from its gradual change counterpart without auxiliary information. The best approach for justifying the use of a specific model is still Chamberlin's method of multiple working hypotheses, where the intended model is one of many alternative models. Only by comparing among multiple alternative models with the original data can the appropriate model and method of threshold detection be determined with a reasonable certainty. In the urbanization example, data support the generalized linear model more than the threshold models. This result is also supported by many ecological studies summarized in Groffman et al. (2006). Our results would have been much different if we had considered only a single model (e.g., the step function model). The generalized linear model with logit transformation of the % impervious surface as the predictor captures the variation in the data well. Based on this model, we expect a fixed fractional change in mean EPTr for every 1% change in the ratio of impervious surface over pervious surface. A 1% change in the ratio can mean a very small or large change in the impervious surface depending on where on the gradient of imperviousness the change is occurring. For example, near the low end of the urban gradient a 1% increase in the ratio can be the result of a small increase in the impervious surface area, but as we move up the urbanization gradient, the same 1% change in the ratio represents an increasingly larger change in impervious surface area. In other words, a small increase in impervious surface in a watershed with no or very little development will lead to the same fractional change in EPTr as a large increase in a watershed that already has significant urban development. Near the high end of the urban gradient, because the EPTr values tend to be very low, even though a small percentage change in impervious surface (e.g., from 98% to 99%) would lead to a large change in the logit transformed predictor (from 3.89 to 4.59), the relative amount of decreasing in EPTr is actually very small. (For example, using the BOS model $\log(\lambda) = 0.82 - 0.65\text{logit}(x)$, a change of 1% in impervious surface from 1% to 2% leads to a change in λ from 45 to 28.5, while the same 1% change from 98% to 99% leads to a change in λ from 0.18 to 0.11.) This model explains the pattern in the data (rapid changes near the low end of the urban gradient and very slow change near the high end of the urban gradient) very well (Fig. 7), even extrapolating outside the data range.

Another important question in threshold analysis is how we apply various statistical models to management problems. The term "threshold" has different meanings when used in different contexts and often conveys a sense of urgency and a need for action when used in a management context. In fitting a statistical model, the term threshold is more accurately described as the change point, that is, the point at which model coefficients change.

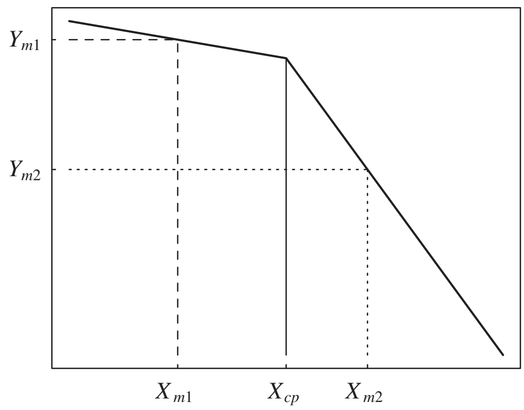

Fig. 8. An hypothetical hockey stick model (thick solid line segments) illustrates the difference between a mathematical change point and management threshold. A management threshold should ensure, with a certain degree of confidence, that the value of the desired endpoint will be preserved. The two dashed lines are two hypothetical desired endpoints (Y_{m1}, Y_{m2}) and their respective thresholds (X_{m1}, X_{m2}). A mathematical change point (X_{cp}) is the predictor variable value at which model coefficients changed.

In a management problem, however, the concept of a threshold response should be discussed in the context of the management endpoint, i.e., the desired outcome. For example, suppose that we want to maintain an EPTr of at least 10, the question now becomes what is the maximum level of impervious surface that can occur in the watershed while keeping EPTr $\geq$ 10. The statistical change point and the ecological threshold can be the same if the response of EPTr to urbanization can be approximated by the step function if the lower and higher values span the management endpoint for the response. However, the change point can differ from the ecological threshold if the underlying model is something other than the step function. For example, if the underlying model is a hockey stick model (Fig. 8), the mathematical change point should not be confused with the management threshold. In this example, the statistical change point (X_{cp}) would not be protective if the management threshold was Y_{m1} and would be overly protective if the management threshold was Y_{m2}. In other words, users of these models should have a clear understanding of the management objective in order to correctly interpret the significance of the change point relative to management thresholds.

Acknowledgement

The author thanks Roxalana Kashuba, Ibrahim Alameddine, and Gerald McMahon for their insightful discussions and comments. Craig Stow, Brian Cade, and Travis Schmidt reviewed an early version of the manuscript and provided valuable comments and suggestions. Comments and suggestions from two anonymous reviewers and the editor are greatly appreciated. The research was completed while SSQ was supported by the USGS through a USGS-Duke University corporative agreement (08HQAG0121).

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolind.2011.08.019.

References

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, 2nd edition, EPA 841-B-99/002. Technical Report, US Environmental Protection Agency; Office of Water; Washington, DC, 1999.

Box, G.E.P., 1976. Science and statistics. Journal of the American Statistical Association 71 (356), 791–799.

Brown, L.R., Cuffney, T.F., Coles, J.F., Fitzpatrick, F., McMahon, G., Steuer, J., Bell, A.H., May, J.T., 2009. Urban streams across the USA: lessons learned from studies in 9 metropolitan areas. Journal of the North American Benthological Society 28 (4), 1051–1069.

Carter, J.L., Fend, S.V., 2005. Setting limits: the development and use of factor-ceiling distributions for an urban assessment using macroinvertebrates. In: Brown, L.R., Gray, R.H., Hughes, R.M., Meador, M.R. (Eds.), Effects of Urbanization on Stream Ecosystems, Bethesda, Maryland. , pp. 179–191 (American Fisheries Society Symposium 47).

Chamberlin, T.C., 1890. The method of multiple working hypotheses. Science 15, 92 (old series).

Chiu, G., Lockhart, R., Routledge, R., 2006. Bent-cable regression theory and applications. Journal of the American Statistical Association 101 (474), 542–553.

Couch, C.A., Hamilton, P.A., 2002. Effects of Urbanization on Stream Ecosystems. Technical Report, US Geological Survey Fact Sheet 042-02.

Cuffney, T.F., Brightbill, R.A., May, J.T., Waite, I.R., 2010. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. Ecological Applications 20 (5), 1384–1401.

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A 222, 309–368.

Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H., Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D., Poff, N.L., Rejeski, D.W., Reynolds, J.F., Turner, M.G., Weathers, K.G., Wiens, J., 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? Ecosystems 9, 1–13.

Muggeo, V.M.R., 2003. Estimating regression models with unknown break-points. Statistics in Medicine 22 (19), 3055–3071.

NCDENR, 2006. Standard Operating Procedures for Benthic Macroinvertebrates. Technical Report, Biological Assessment Unit, North Carolina Department of Environmental and Natural Resources, Raleigh, North Carolina.

Paul, M.J., Bressler, D.W., Purcell, A.H., Barour, M.T., Rankin, E.T., Resh, V.H., 2009. Assessment tools for urban catchments: defining observable biological potential. Journal of the American Water Resources Association 45, 320–330.

Qian, S.S., 2010. Environmental and Ecological Statistics with *R*. Chapman and Hall/CRC Press.

Qian, S.S., King, R.S., Richardson, C.J., 2003. Two statistical methods for the detection of environmental thresholds. Ecological Modelling 166, 87–97.

Qian, S.S., Reckhow, K.H., Zhai, J., McMahon, G., 2005. Nonlinear regression modeling of nutrient loads in streams: a Bayesian approach. Water Resources Research 41, W07012.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of Royal Statistical Society (B) 64, 583–639.