Statistical Methods for Ecological Breakpoints and Prediction Intervals

Jabed H. Tomal\textsuperscript{a,}\textsuperscript{*}, Jan J.H. Ciborowski\textsuperscript{b}

\textsuperscript{a}Department of Computer & Mathematical Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario, Canada M1C 1A4.

\textsuperscript{b}Department of Biological Sciences, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4.

Abstract

The relationships among ecological variables are usually obtained by fitting statistical models that go through the conditional means of the variables. For example, the nonparametric \textit{loess} regression model and the parametric piecewise linear regression model - that go through the conditional mean of the response variable given the predictor - are used to analyze simple to complex relationships among variables. In this article, we have proposed to use \textit{loess} to identify the number and positions of ecological breakpoints, and the piecewise linear regression model to estimate the breakpoints. In contrast, the piecewise linear quantile regression model - which goes through the quantiles of the conditional distribution of the response variable given the predictor - provides much richer information in terms of estimating relationships and breakpoints. We have proposed to use the piecewise linear quantile regression to estimate the breakpoints and thus to obtain prediction intervals of the breakpoints. Three statistical methods have been proposed to construct confidence bands for the ecological response given the human induced disturbances to the nature. The methods are illustrated with two examples from the ecological literature – relating an index of wetlands' fish community ‘health’ to the amount of human activity in wetlands' adjacent watersheds; and relating the biomass of Cyanobacteria to the Total Phosphorus concentration in Canadian lakes.

Keywords: Ecological Breakpoints, Bootstrap, Loess, Piecewise Linear Regression, Quantile Regression

1. Introduction

Maintaining the health of aquatic habitat (biotic integrity) is one of the primary objectives set forth by the US Clean Water Act of 1972 (PL 92 – 500). According to Frey (1977), “biotic integrity” is defined as the “ability of a...
habitat to support and maintain a balanced, integrated, adaptive community of
organism having a composition, diversity and functional organism comparable to
that of a natural habitat.” A natural habitat, also known as reference condition,
refers to an area with minimum level of anthropogenic stress (Host et al., 2005).
In this article, we have proposed several statistical methods and applied them
to two data sets relating to aquatic health of lakes across U.S. and Canada.

The first application is about assessing/estimating threshold effects of agri-
cultural stress, the primary source of human induced disturbance affecting nat-
ural habitat loss for fish in lakes and rivers (Brazner and Beals, 1997; Crosbie
and Chow-Fraser, 1999), on fish index of biotic integrity (Fish IBI), a metric
representing health of an ecoregion or watershed. Danz et al. (2005) derived agri-
cultural stress gradient to characterize degradation of nature using GIS based
data. Uzarski et al. (2005) and Bhagat et al. (2007) developed the multimet-
ric index of biotic integrity through assessing fish community composition with
dominant scirpus vegetation across U.S. and Canada Great Lakes Coastline.

The second application is concerned with measuring threshold effects of total
phosphorus - an environmental contaminant of lakes and rivers through sewage
disposal, agricultural runoff and other forms of pollution (Reynolds and Walsby,
1975) - on the rise of toxigenic cyanobacteria biomass which causes oxygen
depletion, impacts recreation, ecosystem integrity, human and animal health,
and prevents the use of water bodies from drinking water (Downing et al., 2001;
Marieke et al., 2014). Marieke et al. (2014) collected the data from 149 lakes
spreading across three provinces in Canada, where the original sources are the
Ministries of the Environment of Alberta, British Columbia, and Ontario.

Ecological threshold is defined as the point of abrupt change of the response
of an ecological process such as the index of ecological health against habitat
loss such as the human induced disturbance affecting natural habitat (Fahrig,
2001; Francesco Ficetola and Denoël, 2009). Ecological thresholds may appear
handy in understanding the health of a natural habitat, and provide guideline
for conservation targets. Hence, it is very important to understand the point
when human induced disturbance would lead to sharp decline of the health of
natural habitat.

Nonparametric regression is a good choice for studying simple to complex
relationships between a dependent and an independent variable. Trexler and
Travis (1993) discussed the application of locally weighted regression scatterplot
smoothing (loess) in ecology. Following the application of loess in ecology by
Toms and Lesperance (2003), we have suggested to use loess to identify the
number and positions of the ecological breakpoints.

Piecewise linear regression model is very popular in estimating ecological
thresholds. Shea and Vecchione (2002) developed methodology of determining
discontinuities (thresholds) in measurements of ecological variables using a
piecewise linear regression model. Toms and Lesperance (2003) demonstrated
the use of piecewise regression as a statistical technique to model ecological
thresholds, and compared a sharp-transition model with three models incorpo-
rating smooth transitions: the hyperbolic-tangent, bent-hyperbola, and bent-
cable models. Francesco Ficetola and Denoël (2009) compared a wide variety
of statistical models, and suggested that the piecewise linear regression model is a smart choice of estimating ecological thresholds. In all of the mentioned applications, the authors estimated one breakpoint using the piecewise linear regression model. In this article, we extended the model to incorporate two breakpoints and showed two relevant applications.

Quantile regression is a method of estimating relationships between variables in ecological processes defined through different quantiles of the conditional distribution of the response variable. As a result, the quantile regression models provide a more complete view of possible causal relationships between variables. Cade and Noon (2003) gave a gentle introduction to quantile regression for ecologists. They covered linear and non-linear regression models which homogeneous and heterogeneous error variances. McClain and Rex (2001) studied the relationship between dissolved oxygen concentration and maximum size in deep-sea turrid gastropods using a linear quantile regression model. Austin (2007) reviewed some statistical methods for their applications in ecology. Among the methods, linear and nonlinear quantile regressions were incorporated into species response models used in conservation. Bissinger et al. (2008) predicted marine phytoplankton maximum growth rates from temperature using a non-linear quantile regression model. Planque and Buffaz (2008) used a linear quantile regression model to study fish recruitment-environment relationships in marine ecology. Cade et al. (2005) used linear and non-linear quantile regression models to reveal hidden bias and uncertainty in habitat models in ecology. In this article, we have used piecewise linear quantile regression model to estimate ecological thresholds. The estimates of threshold for different quantiles provide a good prediction intervals for the breakpoints. Recall that the confidence interval of threshold provides an idea of the accuracy in estimation of the threshold - not the prediction interval.

Construction of prediction intervals of the response variable given the predictor could be as interesting as the following. Karanth et al. (2004) constructed prediction interval for tigers (carnivores) densities using their prey densities. In this article, we have proposed three prediction intervals for the ecological variables (Fish IBI, Cyanobacteria Biomass) given the human induced disturbances to the aquatic environment (Agricultural Stress, Total Phosphorus) based on: (i) nonparametric regression, (ii) piecewise linear regression, and (iii) quantile regression.

The rest of the article is organized as follows. Section 2 describes the statistical methods used in this article: (i) Loess and Bootstrap Prediction Band, (ii) Piecewise Linear Regression Model, and (iii) Quantile Regression Model. Section 3 presents the results of the statistical methods applied to the (i) Fish Index of Biotic Integrity vs Agricultural Stress, and (ii) Cyanobacteria Biomass vs Total Phosphorus data. Finally, Section 4 concludes with the Discussion and Conclusion.
2. Materials and Methods

The following are the statistical methods/models that we propose to estimate ecological thresholds and to construct prediction intervals.

2.1. Loess and Bootstrap Prediction Band

Let $y$ and $x$ be the response and predictor variables, respectively. A non-parametric smoothed regression model \textit{loess} (Cleveland, 1979) is defined as

$$y = m(x; h) + \epsilon,$$

where $m(x; h)$ is the smoothed function of interest with smoothing parameter $h$, and $\epsilon$ is an independent error term with mean 0 and constant scale $\sigma^2$. The smoothed function $m(x; h)$ is obtained by fitting a polynomial using weighted least squares with varying large and small weights for the close and distant observations, respectively, in a neighborhood of $x$. The \textit{loess} model is robust against a few outliers as it prevents unusual observations exerting large influence in the fitting procedure.

The non-parametric smoothed regression model nicely captures both linear and nonlinear relationships among variables. When the linear models fit poorly due to intrinsic non-linearity in the data, the non-parametric smoothed regression model appears as a savior and shows paths to learn relationships among variables. In this article, our goal of fitting a nonparametric smoothed regression model, more specifically \textit{loess}, is to gain knowledge about the number and position(s) of the possible breakpoints(s) in ecological relationships.

When the purpose of a nonparametric regression model is of providing visual information about the relationships between variables, the goal of prediction band is of providing an impression of the variability of the response variable $y$ given the predictor variable $x$. The bootstrap method (Efron and Tibshirani, 1994) can become useful in obtaining the nonparametric prediction band for the ecological response. Following the methods presented in Härdle and Bowman (1988) and Davison and Hinkley (1997), we propose the following algorithm (Algorithm 1) to construct prediction band for the ecological response using the smoothed regression model. The beauty of this algorithm is that it can also be applied to any parametric method to construct confidence band for the response. For which one need to replace $m(x; h)$ by the appropriate parametric regression model.

2.2. Piecewise Linear Regression Model

A simple linear regression model defines the relationship between a response and a predictor variable - which goes through the conditional mean of the response given the predictor - is linear in parameters and variables. A piecewise linear regression model with one breakpoint, which goes through the conditional mean of the response, connects two linear lines at the breakpoint. Similarly, a piecewise linear regression model with two breakpoints connects three linear lines, where two linear lines are connected to each breakpoint. Piecewise linear
The slopes (independent and identically distributed) iid where $y$ and $\alpha$ are the values for the $i$th response and predictor variables, and $\alpha_1$ and $\alpha_2$ are the two breakpoints. We assume that the errors $\epsilon_i$ are iid (independent and identically distributed) normal with zero mean, constant variance, and finite absolute moment for some order greater than 2. The slopes for the first, second, and third lines of this model are $\beta_1$, $\beta_1 + \beta_2$, and $\beta_1 + \beta_2 + \beta_3$, respectively. The above parametrization of the model forces continuity and abrupt transitions at the breakpoints. The parameters are estimated using non-linear least squares method.
2.3. Piecewise Linear Quantile Regression Model

Classical regression methods focus on estimating parameter vector \( \theta \) of a regression model \( m(x; \theta) \) defined at the conditional mean of the response variable \( y \) as a function of the vector of predictor variables \( x \). In mathematical notation, we write \( E(y|x, \theta) = m(x; \theta) \), where \( m \) is the model of interest. However, it is possible to construct regression models that are defined through different quantiles of the conditional distribution of the response variable given the predictors (Cade and Noon, 2003). Just as mean alone is not enough to characterize an entire distribution, the classical regression method gives an incomplete picture of the conditional distribution of the response variable given the predictors (Mosteller and Tukey, 1977).

The regression models which go through different quantiles of the conditional distribution of the response variable are known as quantile regression models. Such models allowed us studying relationships among variables through conditional median of the response variable, and the full range of other conditional quantile functions. By supplementing the classical regression model which is defined at the conditional mean only, quantile regression models provide a more complete statistical analysis of the relationships among variables. For example, in a simple linear regression model with heterogeneous variance linearly increasing/decreasing with the values of the predictor implies that there is not a single slope that characterizes changes in the conditional distribution of the response given the predictor. In such a situation, by focusing exclusively on changes in the conditional mean, a classical linear regression model may underestimate/overestimate the slopes in the heterogeneous conditional distribution (Cade et al., 1999; Terrell et al., 1996). Unequal variance in the conditional distribution implies that there are many slopes describing the relationship between the response and predictor, and quantile regression models estimate multiple slopes and provides a more informative picture of the relationships.

Let \( Q_\tau(y|x) \) be the \( \tau \)th quantile of the conditional distribution of the response \( y \) given the predictor \( x \). Then the piecewise linear quantile regression model with two breakpoints is defined as:

\[
Q_\tau(y|x) = \begin{cases} 
\beta_0(\tau) + \beta_1(\tau)x_i + \epsilon_i & \text{for } x_i \leq \alpha_1(\tau) \\
\beta_0(\tau) + \beta_1(\tau)x_i + \beta_2(\tau)(x_i - \alpha_1(\tau)) + \epsilon_i & \text{for } \alpha_1(\tau) < x_i \leq \alpha_2(\tau) \\
\beta_0(\tau) + \beta_1(\tau)x_i + \beta_2(\tau)(x_i - \alpha_1(\tau)) + \beta_3(\tau)(x_i - \alpha_2(\tau)) + \epsilon_i & \text{for } x_i > \alpha_2(\tau)
\end{cases}
\]

where \( \alpha_1(\tau) \) and \( \alpha_2(\tau) \) are the first and second breakpoints defined at the \( \tau \)th quantile of the conditional distribution.

To estimate the parameters of the piecewise linear quantile regression model the following objective function is optimized:

\[
\min_{\theta(\tau) \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau (g_i(\theta(\tau))) ,
\]

where \( g_i(\theta(\tau)) = (y_i - Q_\tau(y|x)) \), assumed to be differentiable with respect to \( \theta(\tau) \), and \( \rho_\tau \) is the loss function defined as \( \rho(u) = u(\tau - I(u < 0)) \), for some
\( \tau \in (0,1) \) and \( I \) being the indicator function. Here, \( \theta(\tau) \) is the vector of parameters, with dimension \( p \), for the regression model at \( \tau \)th quantile. The above optimization problem is solved using linear programming method (Koenker, 2005).

The advantage of quantile regression is that there is no restriction for any distribution of the error term nor there is any restriction in specifying the structure for the error variance. In this article, we aim to show that the quantile regression estimates are good choices to construct prediction intervals for the breakpoints without assuming any parametric distribution for the estimates.

3. Results

In this section, we present results of the statistical methods, presented in Section 2, applied to two data sets from ecological literature.

3.1. Fish Index of Biotic Integrity vs Agricultural Stress

The index of biotic integrity (IBI) reflects the health of an ecoregion or watershed. Uzarski et al. (2005) and Bhagat et al. (2007) have developed a multimetric index of biotic integrity by assessing fish community composition at 30 sites with dominant Schoenoplectus (formerly Scirpus) vegetation across the US Great Lakes Coastal Wetlands. Researchers (Brazner and Beals, 1997; Crosbie and Chow-Fraser, 1999) have quantified agricultural and land use as the major cause of natural habitat loss for fish population in lakes and rivers. Danz et al. (2005), as part of Great Lakes Environmental Indicators (GLEI) program, have derived agricultural stress gradient to characterize the degradation of nature using GIS based data across the US Great Lakes coastline. In this study, we regress the response variable fish index of biotic integrity using the predictor variable agricultural stress gradient. Theoretically, the fish IBI scores vary from 0 to 100, with larger scores representing better ecological health of watersheds. The predictor variable agricultural stress gradient, a scaled principal component analysis (PCA) score, varies from 0 to 1, with larger numbers representing higher agricultural activities. Table 1 shows the means and 95% confidence intervals for the means of the two variables for the 13 GLEI (Danz et al., 2005) sites and 17 Uzarski (Uzarski et al., 2005) sites. The Uzarski sites have slightly larger Fish IBI and agricultural stress than that of the GLEI sites. On the other hand, GLEI sites have smaller mean but wider confidence interval of mean for agricultural stress. Bhagat et al. (2007) examined the relationship between Fish IBI for Schoenoplectus vegetation and agricultural stress and concluded that the Fish IBI scores exhibited a threshold response with respect to agricultural stress. The following subsections 3.1.1, 3.1.2, and 3.1.3 contain results from statistical methods applied to the Fish IBI vs Agricultural Stress data.
**Table 1**: Means and 95% confidence intervals of means of the two variables *Fish IBI* and *agricultural stress* for the GLEI and Uzarski studies.

| Study     | Variables | n  | Mean | 95% CI of mean lower | 95% CI of mean upper |
|-----------|-----------|----|------|----------------------|----------------------|
| GLEI      | Fish IBI  | 13 | 42.23| 36.57                | 47.90                |
|           | Agg. Stress | 0.30 |     | 0.20                  | 0.40                |
| Uzarski   | Fish IBI  | 17 | 46.06| 38.04                | 54.08                |
|           | Agg. Stress | 0.34 |     | 0.27                  | 0.40                |
| Combined  | Fish IBI  | 30 | 44.40| 39.47                | 49.33                |
|           | Agg. Stress | 0.32 |     | 0.27                  | 0.38                |

**3.1.1. Results: Loess and Bootstrap Confidence Band**

We have used the *R* (*R Core Team, 2017*) statement `loess` to fit the model and to construct the confidence band represented by equation 1 and algorithm 1, respectively. The goals are 3 folds: getting an idea about the (i) types of relationship between Fish IBI and agricultural stress, (ii) possible locations of the breakpoints, and (iii) prediction band for the response Fish IBI given the predictor agricultural stress.

Figure 1 shows the fitted model `loess` and its 80% and 95% prediction bands superimposed to the data. The relationship between Fish IBI and agricultural stress is negative: as the agricultural stress increases the Fish IBI decreases. The nonparametric model `loess` is showing that there are possibly two breakpoints: one is around agricultural stress 0.22 and the other is around 0.45. The Fish IBI score remains steady for agricultural stress level up to 0.22, decreases sharply from 0.22 to 0.45, and reaches to its minimum after 0.45. The 80% and 95% bootstrap prediction bands for the response Fish IBI are obtained from 10,000 bootstrap samples. The prediction bands provide us an idea about possible locations and spread of future data points.

**3.1.2. Results: Piecewise Linear Regression Model**

We have fitted the piecewise linear regression model (equation 2) to the Fish IBI vs Agricultural Stress data. The piecewise linear regression model helps us estimating the relationship among variables and the breakpoints as well. It also helps us making inferences to check if the relationship is statistically significant. The model also gives us the confidence intervals for the breakpoints. We have used the *R* package “segmented” (*Muggeo, 2015*) to fit the piecewise linear regression model.

Table 2 shows the estimates and standard errors of the regression coefficients and breakpoints of the piecewise linear regression model fitted to the Fish IBI vs Agricultural Stress data. The first and second breakpoints $\alpha_1$ and $\alpha_2$ are 0.263 and 0.488, respectively. The four regression coefficients $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are 51.924, 4.405, $-166.138$ and 274.001, respectively.

Table 3 shows the estimates, standard errors, and 95% confidence intervals of the breakpoints and slopes of the piecewise linear regression model fitted to
the Fish IBI vs Agricultural Stress data. The 95% confidence intervals for the first and second breakpoints $\alpha_1$ and $\alpha_2$ are (0.196, 0.331) and (0.391, 0.585), respectively. The confidence interval tells us the accuracy of the estimates: the smaller the width of the interval the better the estimates. Recall that the width of the confidence interval depends on the size of the sample. The larger the sample the narrower the width of the confidence interval. As there is no overlap between the two confidence intervals, the two breakpoints are significantly different from each other.

The 95% confidence interval of slope in the first segment of the regression model is $(-70.110, 78.920)$. As the confidence interval contains $\beta_1 = 0$ inside, the relationship between Fish IBI and Agricultural Stress is statistically insignificant in this segment. That is, Fish IBI does not reflect the changes as Agricultural Stress changes from 0 to 0.263. The 95% confidence interval of slope in the second segment of the regression model is $(-272.800, -50.670)$. As the confidence interval does not contain $\beta_1 + \beta_2 = 0$ inside, the relationship between Fish IBI and Agricultural Stress is statistically significant in this segment. That is, Fish IBI decreases significantly with the increase of Agricultural Stress between 0.263 to 0.488. The 95% confidence interval of slope in the third
Table 2: Estimates and standard errors of the regression coefficients and breakpoints of the piecewise linear regression models fitted to the Fish IBI Scores vs Agricultural Stress data.

| Names | Parameters | Fish IBI vs Agricultural Stress |
|-------|------------|---------------------------------|
|       |            | Estimates | Standard Errors |
| Breakpoints | $\alpha_1$ | 0.263 | 0.033 |
|           | $\alpha_2$ | 0.488 | 0.047 |
| Coefficients | $\beta_0$ | 51.924 | 7.805 |
|           | $\beta_1$ | 4.405 | 36.104 |
|           | $\beta_2$ | -166.138 | 64.801 |
|           | $\beta_3$ | 274.001 | 96.975 |

The segment of the regression model is $(-54.240, 278.800)$. As the confidence interval contains $\beta_1 + \beta_2 + \beta_3 = 0$ inside, the relationship between Fish IBI and Agricultural Stress is statistically insignificant in this segment. That is, Fish IBI does not reflect the changes as Agricultural Stress changes from 0.488 and 1.00.

Table 3: Estimates, standard errors (SE) and 95% confidence intervals (95% CI) of the breakpoints (breaks) and slopes of the piecewise-linear regression lines fitted to the Fish IBI Scores against Agricultural Stress data. The statistically significant slopes are highlighted by light gray color.

| Names | Parameters | Fish IBI vs Agricultural Stress |
|-------|------------|---------------------------------|
|       |            | Estimates | SE | 95% CI |
|       |            |            | Lower | Upper |
| Breaks | $\alpha_1$ | 0.263 | 0.033 | 0.196 | 0.331 |
|         | $\alpha_2$ | 0.488 | 0.047 | 0.391 | 0.585 |
| Slopes | $\beta_1$ | 4.405 | 36.100 | -70.110 | 81.920 |
|         | $\beta_1 + \beta_2$ | -161.700 | 53.810 | -272.800 | -50.670 |
|         | $\beta_1 + \beta_2 + \beta_3$ | 112.300 | 80.680 | -54.240 | 278.800 |

Figure 2 shows the piecewise linear regression model fitted to the Fish IBI and Agricultural Stress data. This figure visualizes the two breakpoints 0.263 and 0.488 along with the 95% confidence intervals (see the two lines along the horizontal axis). This plot also shows the 80% and 95% prediction bands for the response variable Fish IBI given the predictor agricultural stress which provide us an idea about the positions of the future data points. The prediction intervals are wider in this figure, in contrast to Figure 1, as the model contains 6 parameters and uses only 30 observations for estimation.

3.1.3. Results: Piecewise Linear Quantile Regression

Piecewise linear quantile regression model provides much richer information in terms of estimating relationship and breakpoints than the piecewise linear regression model which goes through the mean. The collection of breakpoints from different quantile regression curves provides prediction intervals for the
breakpoints. Moreover, the upper and lower quantile regression curves provide a nice prediction band for the data points.

Figure 3 shows the fitted piecewise linear quantile regression model (equation 3) for quantiles $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ applied to the Fish IBI vs Agricultural Stress data. The quantile regression curves at 10th and 90th quantiles provide a very nice prediction band for the data points.

The fitted quantile regression curves provide a collection of estimates for the first and second breakpoints. Table 4 shows the estimates of the breakpoints $\alpha_1$ and $\alpha_2$ of the piecewise linear regression models for quantiles $\tau = (0.10, 0.25, 0.40, 0.50, 0.60, 0.75, 0.90)$ applied to the Fish IBI vs Agricultural Stress data. The smallest and the largest breakpoints are highlighted by light and dark gray, respectively. The collection of estimates for the first breakpoint ranges from 0.233 to 0.283 and can be considered as the 80% prediction band for $\alpha_1$. The collection of estimates for the second breakpoint ranges from 0.448 to 0.552 and can be considered as the 80% prediction band for $\alpha_2$. So the first breakpoint can occur between 0.233 to 0.283 and the second breakpoint can occur between 0.448 to 0.552. In contrast the confidence intervals for the
breakpoints provide accuracy in the estimates - not the prediction interval.

3.2. Cyanobacteria Biomass vs Total Phosphorus

Cyanobacteria biomass is often used as an index of the quality of aquatic health. The bloom of cyanobacteria biomass degrades aquatic ecosystem and thus is directly related to human and animal health (Downing et al., 2001). Cyanobacteria poisoning produces various types of toxins which are linked to diseases such as carcinoma and others (Falconer and Humpage, 1996). The solution of these problems requires understanding of the issues that cause the bloom of cyanobacteria in lakes and river. It has been shown that one of the main nutrients that causes cyanobacteria bloom is total phosphorus (McCauley et al., 1989; Downing et al., 2001).

The increase of cyanobacterial biomass is a global phenomenon and has been observed in various parts of the world. Moreover, the reports of algal bloom are on rise across Canada (Pick, 2016). Marieke et al. (2014) have used data from 149 lakes spread across three region in Canada and fitted a linear regression model to predict log of cyanobacterial biomass (µg/L) using log of total phosphorus (µg/L). The data were provided by the Ministries of the Environment of Alberta, British Columbia and Ontario.
We have observed that the relationships between log of cyanobacterial biomass and log of total phosphorus is better reflected by a broken stick model with abrupt changes at two position defined by the log of total phosphorus. Hence, we have used the piecewise linear regression model to estimate the breakpoints and confidence bands for the response variable given the predictor. The data comprised of 43 sites from Alberta, 10 from British Columbia, and 97 from Ontario. Table 5 shows the means and 95% confidence intervals of the means for the two variables log of cyanobacterial biomass (Cyan) and log of total phosphorus (TP) for the lakes in Alberta, British Columbia (BC) and Ontario. The overall mean of the log of cyanobacterial biomass is 4.88 with 95% confidence interval (4.51, 5.25). The overall mean of the log of total phosphorus is 2.77 with 95% confidence interval (2.61, 2.93). The means of log of cyanobacterial biomass and total phosphorus are the largest in Alberta followed by British Columbia and Ontario. Both Cyanobacteria biomass and total phosphorus are significantly higher in Alberta than British Columbia and Ontario. The mean of Cyanobacteria biomass and total phosphorus in British Columbia and Ontario are not significantly different.

The following subsections 3.2.1, 3.2.2, and 3.2.3 contain results from statis-

---

Table 4: Estimates of the breakpoints $\alpha_1$ and $\alpha_2$ of the piecewise linear quantile regression models for quantiles $\tau = (0.10, 0.25, 0.40, 0.50, 0.60, 0.75, 0.90)$ applied to the Fish IBI vs Agricultural Stress data. The smallest and largest breakpoints are highlighted by light and dark gray, respectively.

| Quantiles | Breakpoints | Quantiles | Breakpoints |
|-----------|-------------|-----------|-------------|
| $\tau$   | $\alpha_1(\tau)$ | $\alpha_2(\tau)$ | $\tau$   | $\alpha_1(\tau)$ | $\alpha_2(\tau)$ |
| 0.10      | 0.233       | 0.452     | 0.60       | 0.255     | 0.476     |
| 0.25      | 0.264       | 0.469     | 0.75       | 0.284     | 0.528     |
| 0.40      | 0.270       | 0.448     | 0.90       | 0.264     | 0.552     |
| 0.50      | 0.264       | 0.466     | -          | -         | -         |

Table 5: Means and 95% confidence intervals of means of the two variables log of cyanobacteria biomass (Cyan) and log of total phosphorus (TP) for the lakes in Alberta, British Columbia (BC) and Ontario.

| Study    | Variables | $n$ | Mean | 95% CI of mean |
|----------|-----------|-----|------|----------------|
|          |           |     |      | lower | upper |
| Alberta  | Cyan      | 43  | 6.59 | 6.03  | 7.14  |
|          | TP        | 3.54|      | 3.32  | 3.76  |
| BC       | Cyan      | 10  | 5.52 | 4.01  | 7.04  |
|          | TP        | 2.57|      | 1.87  | 3.27  |
| Ontario  | Cyan      | 97  | 4.05 | 3.64  | 4.47  |
|          | TP        | 2.45|      | 2.27  | 2.63  |
| Combined | Cyan      | 150 | 4.88 | 4.51  | 5.25  |
|          | TP        | 2.77|      | 2.61  | 2.93  |
tical methods applied to the cyanobacteria biomass vs total phosphorus data.

3.2.1. Results: Loess and Bootstrap Prediction Band

Figure 4 shows the fitted model loess and its 80% and 95% prediction bands superimposed to the data. The relationship between log of cyanobacteria biomass and log of total phosphorus is positive: as the log of total phosphorus increases the log of cyanobacteria biomass increases. The nonparametric model loess is showing that there are possibly two breakpoints: one is around log of total phosphorus 2.5 and the other is around 4.0. The cyanobacteria biomass increases steadily for the log of total phosphorus level up to 2.5, sharply from 2.5 to 4.0, and slowly after 4.0. The 80% and 95% bootstrap prediction bands for the data points are obtained from 10,000 bootstrap samples. Having seen the prediction bands, one can easily predict the possible range of log of cyanobacteria biomass for a given value of the log of total phosphorus.

3.2.2. Results: Piecewise Linear Regression Model

We have fitted the piecewise linear regression model (equation 2) to the cyanobacteria biomass vs total phosphorus data. Table 6 shows the estimates
and standard errors of the regression coefficients and breakpoints of the piecewise linear regression model fitted to the cyanobacteria biomass and total phosphorus data. The first and second breakpoints $\alpha_1$ and $\alpha_2$ are 2.798 and 3.737, respectively. The four regression coefficients $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are 1.050, 1.176, 2.181 and $-2.526$, respectively.

| Names       | Parameters | Cyanobacteria vs Phosphorus |
|-------------|------------|-----------------------------|
|             |            | Estimates       | Standard Errors |
| Breakpoints | $\alpha_1$ | 2.798           | 0.219           |
|             | $\alpha_2$ | 3.737           | 0.237           |
| Coefficients| $\beta_0$  | 1.050           | 0.678           |
|             | $\beta_1$  | 1.176           | 0.326           |
|             | $\beta_2$  | 2.181           | 0.815           |
|             | $\beta_3$  | $-2.526$        | 0.858           |

Table 6: Estimates and standard errors of the regression coefficients and breakpoints of the piecewise linear regression models fitted to the Cyanobacteria Biomass vs Total Phosphorus data.

Table 7 shows the estimates, standard errors and 95% confidence intervals of the breakpoints and slopes of the piecewise linear regression model fitted to the cyanobacteria biomass vs total phosphorus data. The statistically significant slopes are highlighted by light gray color. The 95% confidence intervals for the breakpoints $\alpha_1$ and $\alpha_2$ are (2.366, 3.230) and (3.269, 4.206), respectively. The two breakpoints are statistically different as there is no overlap between the two confidence intervals.

The estimated slope of the first segment of the regression line is $\hat{\beta}_1 = 1.176$ with 95% confidence interval (0.533, 1.820). As the confidence interval does not contain $\beta_1 = 0$ inside, the relationship between log of cyanobacteria biomass and log of total phosphorus is statistically significant in this segment. This indicates that the log of cyanobacteria biomass increases steadily at a rate of 1.176 with the increase of the log of total phosphorus. The estimated slope in the second segment of the regression line is $\hat{\beta}_1 + \hat{\beta}_2 = 3.357$ with 95% confidence interval (1.880, 4.833). As the confidence interval does not contain $\beta_1 + \beta_2 = 0$ inside, the relationship between log of cyanobacteria biomass and log of total phosphorus is statistically significant. In this segment, the log of cyanobacteria biomass increases sharply/rapidly and significantly at a rate of 3.357 with the increase of the log of total phosphorus. The estimated slope in the third segment of the regression line is $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 0.831$ with 95% confidence interval ($-0.003$, 1.665). As the confidence interval contains $\beta_1 + \beta_2 + \beta_3 = 0$ inside, the relationship between log of cyanobacteria biomass and log of total phosphorus is statistically insignificant.

Figure 5 shows the scatter plot of log of cyanobacteria biomass against log of total phosphorus highlighting the labels for the lakes in Alberta, British Columbia and Ontario. The scatter plot displays the fitted piecewise linear regression model along with 80% and 95% prediction bands for the response.
variable given the predictor. The two lines along the horizontal axis are the marginal 95% confidence intervals of the breakpoints $\alpha_1$ and $\alpha_2$. The two non-overlapping confidence intervals shows that the two breakpoints are statistically significant. The prediction intervals for the data points are nice and compact as there are many data points (150, here) in this application. The prediction intervals give us an idea about the range of future values of log of cyanobacteria biomass for a given value of log of log of total phosphorus.

### 3.2.3. Results: Piecewise Linear Quantile Regression

Figure 6 shows the fitted piecewise linear quantile regression model (equation 3) for quantiles $\tau = 0.10, 0.25, 0.50, 0.75,$ and 0.90 applied to the log of cyanobacteria biomass and log of total phosphorus data. The piecewise linear regression curve at the 90th quantile provided one breakpoint only along the second threshold. The quantile regression curves at 10th and 90th quantiles provide a prediction band for the data points.

The quantile regression models provide a collection of estimates for the first and second break points. Table 8 shows the values of the first and second breakpoints taken by the fitted quantile regression models at quantiles $\tau = 0.10, 0.25, 0.40, 0.50, 0.60, 0.75,$ and 0.90. The value for the first break point at quantile $\tau = 0.90$ is not available for sparsity in the data points. The values for the first break point $\alpha_1$ range from the smallest 2.553 to the largest 2.833. That means the first threshold can occur as early as 2.553 and last until 2.833. The values for the second break point $\alpha_2$ range from the smallest 3.627 to the largest 3.957. That means the second threshold can occur as early as 3.627 and can last until 3.957. The 80% prediction bands for $\alpha_1$ and $\alpha_2$ are (2.553, 2.833) and (3.627, 3.957), respectively. Recall that the Table 7 shows the confidence intervals of $\alpha_1$ and $\alpha_2$ which measure the accuracy of estimation - NOT prediction interval.
4. Discussion and Conclusion

A comprehensive collection of statistical methods have been proposed to estimate ecological thresholds in two applications. We proposed to use the nonparametric regression model \textit{loess} to have an idea about the number and positions of the breakpoints. Having obtained initial values of the breakpoints from \textit{loess}, the piecewise linear regression model was used to estimate the breakpoints. Finally, a near complete collection of estimates of the breakpoints was obtained using the piecewise linear quantile regression method. While the piecewise linear regression model allows us obtaining confidence intervals for the breakpoints providing the accuracy in estimation, the piecewise linear quantile regression model allows us obtaining prediction intervals for the breakpoints. The prediction intervals tell us the potential start and end of the breakpoints based on the human induced disturbance to the nature. Furthermore, we have presented three methods for obtaining the prediction intervals for the response variable indicating the index of ecological health given the predictor variable representing human induced disturbance to the nature: (i) nonparametric bootstrap pre-
Figure 6: Piecewise linear quantile regression models for quantiles $\tau = 0.10, 0.25, 0.50, 0.75, \text{ and } 0.90$ applied to the log of Cyanobacteria Biomass and log of Total Phosphorus data. The models with quantiles 0.10 and 0.90 provide 80% prediction band for the data points.

diction band using loess, (ii) parametric prediction band using the parametric piecewise linear regression model, and (iii) parametric prediction band using the piecewise linear quantile regression model.

The methods have been applied to the Fish Index of Biotic Integrity vs Agricultural Stress data. While the Fish IBI reflects the health of an ecoregion or watershed, the Agricultural Stress reflects the human induced agricultural disturbances to the nature. From the estimates of the piecewise linear regression model, the first and second breakpoints occur at Agricultural Stress PCAs 0.263 and 0.488, respectively. The two breakpoints are significantly different from each other. The Fish IBI does not reflect to the changes as Agricultural Stress changes from 0 to 0.263, decreases significantly with the increase of Agricultural Stress from 0.263 to 0.488, and does not reflect to the changes as Agricultural Stress changes from 0.488 to 1.00. The 80% prediction interval for the first and second breakpoints are (0.233, 0.284) and (0.448, 0.552), respectively. That is, we are 80% confident that the first breakpoint can occur from 0.233 to 0.283 and the second breakpoint can occur from 0.448 to 0.552.

The other application is of finding ecological breakpoints relating cyanobacterial biomass ($\mu g/L$), which causes degradation of aquatic ecosystem, and to-
Table 8: Estimates of the breakpoints $\alpha_1$ and $\alpha_2$ of the piecewise linear quantile regression models for quantiles $\tau = 0.10, 0.25, 0.40, 0.50, 0.60, 0.75,$ and $0.90$ applied to the log of Cyanobacteria Biomass vs log of Total Phosphorus data. The smallest and largest breakpoints are highlighted by light and dark gray, respectively.

| Quantiles $\tau$ | Breakpoints $\alpha_1(\tau)$ | Quantiles $\tau$ | Breakpoints $\alpha_2(\tau)$ |
|-----------------|-------------------------------|-----------------|-------------------------------|
| 0.10            | 2.580                         | 0.60            | 2.553                         |
|                 | 3.871                         |                 | 3.759                         |
| 0.25            | 2.794                         | 0.75            | 2.639                         |
|                 | 3.738                         |                 | 3.627                         |
| 0.40            | 2.833                         | 0.90            | NA                            |
|                 | 3.807                         |                 | 3.957                         |
| 0.50            | 2.773                         | -               | -                             |
|                 | 3.756                         | -               | -                             |

Total phosphorus ($\mu g/L$), a human induced contaminant of lakes and rivers. The piecewise linear regression model is fitted to the log of cyanobacterial biomass and log of total phosphorus. The first and second breakpoints are 2.798 and 3.737, respectively, and are significantly different from each other. The log of cyanobacteria biomass increases slowly before 2.798, sharply from 2.798 to 3.734, and slowly after 3.737 of the log of total phosphorus. The 80% prediction intervals for the first and second breakpoints are (2.553, 2.833) and (3.627, 3.957), respectively. That is, we are 80% confident that the first breakpoint can occur between 2.553 and 2.833, and the second breakpoint can occur between 3.627 and 3.957.

The 80% and 95% bootstrap prediction bands using loess – for the Fish Index of Biotic Integrity vs Agricultural Stress data (Figure 1) and Cyanobacteria Biomass vs Total Phosphorus data (Figure 4) – are smooth and rough, respectively. The 95% prediction band is rough as there are less amount of data towards the edges. Overall, the prediction intervals using the proposed bootstrap method are nice. However, the methodology is not much sensitive to the sizes of the data sets. Recall that the Fish data and the Cyanobacteria data contain 30 and 150 observations, respectively. The width of the prediction band using the parametric piecewise linear regression model is sensitive to the size of the data. Figure 2 shows a wide prediction band as the Fish data contain only 30 observations when the model contains 6 parameters. Figure 5 shows a nice and compact prediction band as the Cyanobacteria data contain 150 observations which is much larger comparing to the number of parameters in the model. The prediction bands (Figures 3 & Figure 6) from the piecewise linear quantile regression is a bit erratic as the model specific to different quantiles uses even much less data points. As a result, the prediction bands do not look as nice as the other two, one from loess and the other from piecewise linear regression model. However, as like all statistical methods, large amount of data points would make the prediction bands smooth and nice.

In order to maintain the health of an ecoregion or watershed, our recommendation is to limit agricultural activities of surrounding wetlands by keeping stress level at most 0.233. To maintain the health of an aquatic ecosystem in lakes and rivers, we recommend to keep the amount of log of total phosphorus...
level at most 2.553. Recall that 0.233 and 2.553 are the lower limits of the prediction intervals of the first breakpoint for the fish and cyanobacteria data, respectively.

Acknowledgment

The present study forms part of the projects [2012GS430203; 2014BAC09B00] supported by the Ministry of Science and Technology, P.R. China.

References

Mike Austin. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200(1):1 – 19, 2007. ISSN 0304-3800. doi: http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005. URL http://www.sciencedirect.com/science/article/pii/S0304380006003140.

Yakuta Bhagat, Jan J.H. Clborowski, Lucinda B. Johnson, Donald G. Uzarski, Thomas M. Burton, Steven T.A. Timmermans, and Matthew J. Cooper. Testing a fish index of biotic integrity for responses to different stressors in great lakes coastal wetlands. *Journal of Great Lakes Research*, 33:224 – 235, 2007. doi: 10.3394/0380-1330(2007)33[224:TAFIOB]2.0.CO;2.

Jan E Bissinger, David JS Montagnes, Jonathan Sharples, and David Atkinson. Predicting marine phytoplankton maximum growth rates from temperature: Improving on the eppley curve using quantile regression. *Limnology and Oceanography*, 53(2):487, 2008.

J C Brazner and E W Beals. Patterns in fish assemblages from coastal wetland and beach habitats in green bay, lake michigan: a multivariate analysis of abiotic and biotic forcing factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(8):1743–1761, 1997. doi: 10.1139/f97-079.

Brian S. Cade and Barry R. Noon. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8):412–420, 2003. ISSN 1540-9309. doi: 10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2.

Brian S. Cade, James W. Terrell, and Richard L. Schroeder. Estimating effects of limiting factors with regression quantiles. *Ecology*, 80:311–323, 1999.

Brian S. Cade, Barry R. Noon, and Curtis H. Flather. Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, 86(3):786–800, 2005. ISSN 1939-9170. doi: 10.1890/04-0785. URL http://dx.doi.org/10.1890/04-0785.
William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979. doi: 10.1080/01621459.1979.10481038.

Barb Crosbie and Patricia Chow-Fraser. Percentage land use in the watershed determines the water and sediment quality of 22 marshes in the great lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(10):1781–1791, 1999. doi: 10.1139/f99-109.

Nicholas P. Danz, Ronald R. Regal, Gerald J. Niemi, Valerie J. Brady, Tom Hollenhorst, Lucinda B. Johnson, George E. Host, Joann M. Hanowski, Carol A. Johnston, Terry Brown, John Kingston, and John R. Kelly. Environmentally stratified sampling design for the development of great lakes environmental indicators. *Environmental Monitoring and Assessment*, 102(1):41–65, 2005. ISSN 1573-2959. doi: 10.1007/s10661-005-1594-8.

A.C. Davison and D.V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. ISBN 9780521574716.

John A Downing, Susan B Watson, and Edward McCauley. Predicting cyanobacteria dominance in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(10):1905–1908, 2001. doi: 10.1139/f01-143. URL http://www.nrcresearchpress.com/doi/abs/10.1139/f01-143.

B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994. ISBN 9780412042317.

L. Fahrig. How much habitat is enough? *Biological Conservation*, 100(1):65 – 74, 2001. ISSN 0006-3207. doi: https://doi.org/10.1016/S0006-3207(00)00208-1. URL http://www.sciencedirect.com/science/article/pii/S0006320700002081.

Ian R. Falconer and Andrew R. Humpage. Tumour promotion by cyanobacterial toxins. *Phycologia*, 35(6S):74–79, 1996. doi: 10.2216/i0031-8884-35-6S-74.1. URL http://dx.doi.org/10.2216/i0031-8884-35-6S-74.1.

Gentile Francesco Ficetola and Mathieu Denoël. Ecological thresholds: an assessment of methods to identify abrupt changes in species–habitat relationships. *Ecography*, 32(6):1075–1084, 2009. ISSN 1600-0587. doi: 10.1111/j.1600-0587.2009.05571.x. URL http://dx.doi.org/10.1111/j.1600-0587.2009.05571.x.

D.G. Frey. Biological integrity of water - an historic approach. In R.K. Ballantine and L.J. Guarraia, editors, *The Integrity of Water*, pages 127–140. U.S. Environmental Protection Agency, 1977.

Wolfgang Härdle and Adrian W. Bowman. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401):102–110, 1988. ISSN 01621459.
G. E. Host, J. Schuldt, J. J. H. Ciborowski, L. B. Johnson, T. Hollenhorst, and C. Richards. Use of gis and remotely sensed data for a priori identification of reference areas for great lakes coastal ecosystems. *International Journal of Remote Sensing*, 26(23):5325–5342, 2005. doi: 10.1080/01431160500219364. URL http://dx.doi.org/10.1080/01431160500219364.

K Ullas Karanth, James D Nichols, N Samba Kumar, William A Link, and James E Hines. Tigers and their prey: predicting carnivore densities from prey abundance. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4854–4858, 2004.

R. Koenker. *Quantile Regression*. Cambridge: Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

Beaulieu Marieke, Pick Frances, Palmer Michelle, Watson Sue, Winter Jenny, Zurawell Ron, and Gregory-Eaves Irene. Comparing predictive cyanobacterial models from temperate regions. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(12):1830–1839, 2014. doi: 10.1139/cjfas-2014-0168. URL http://dx.doi.org/10.1139/cjfas-2014-0168.

Edward McCauley, John A. Downing, and Susan Watson. Sigmoid relationships between nutrients and chlorophyll among lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(7):1171–1175, 1989. doi: 10.1139/f89-152. URL http://dx.doi.org/10.1139/f89-152.

C. McClain and M. Rex. The relationship between dissolved oxygen concentration and maximum size in deep-sea turrid gastropods: an application of quantile regression. *Marine Biology*, 139(4):681–685, 2001. ISSN 1432-1793. doi: 10.1007/s002270100617. URL http://dx.doi.org/10.1007/s002270100617.

F. Mosteller and J.W. Tukey. *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company, 1977. ISBN 9780201048544.

Vito M. R. Muggeo. *segmented: Regression Models with Breakpoints/Changepoints Estimation*, 2015. R package version 0.5-1.4.

Frances R. Pick. Blooming algae: a canadian perspective on the rise of toxic cyanobacteria. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(7):1149–1158, 2016. doi: 10.1139/cjfas-2015-0470. URL http://dx.doi.org/10.1139/cjfas-2015-0470.

Benjamin Planque and Laure Buffaz. Quantile regression models for fish recruitment–environment relationships: four case studies. *Marine Ecology Progress Series*, 357:213–223, 2008. doi: https://doi.org/10.3354/meps07274. URL http://www.int-res.com/articles/meps2008/357/m357p213.pdf.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org.
C.S. Reynolds and A.E. Walsby. Water-blooms. *Biological Reviews*, 50(4):437–481, 1975. ISSN 1469-185X. doi: 10.1111/j.1469-185X.1975.tb01060.x. URL http://dx.doi.org/10.1111/j.1469-185X.1975.tb01060.x.

G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003. ISBN 9780471471356. doi: 10.1002/0471725315.

E. Shea and M. Vecchione. Quantification of ontogenetic discontinuities in three species of oegopsid squids using model ii piecewise linear regression. *Marine Biology*, 140(5):971–979, 2002. ISSN 1432-1793. doi: 10.1007/s00227-001-0772-7. URL http://dx.doi.org/10.1007/s00227-001-0772-7.

James W. Terrell, Brian S. Cade, Jeanette Carpenter, and Jay M. Thompson. Modeling stream fish habitat limitations from wedge-shaped patterns of variation in standing stock. *Transactions of the American Fisheries Society*, 125(1):104–117, 1996. doi: 10.1577/1548-8659(1996)125⟨0104:MSFHLF⟩2.3.CO;2.

Judith D. Toms and Mary L. Lesperance. Piecewise regression: A tool for identifying ecological thresholds. *Ecology*, 84(8):2034–2041, 2003. ISSN 1939-9170. doi: 10.1890/02-0472.

Joel C. Trexler and Joseph Travis. Nontraditional regression analyses. *Ecology*, 74(6):1629–1637, 1993. ISSN 1939-9170. doi: 10.2307/1939921. URL http://dx.doi.org/10.2307/1939921.

Donald G. Uzarski, Thomas M. Burton, Matthew J. Cooper, Joel W. Ingram, and Steven T.A. Timmermans. Fish habitat use within and across wetland classes in coastal wetlands of the five great lakes: Development of a fish-based index of biotic integrity. *Journal of Great Lakes Research*, 31:171 – 187, 2005. doi: 10.1016/S0380-1330(05)70297-5.