A new method for detecting and interpreting biodiversity and ecological community thresholds

Matthew E. Baker[1]* and Ryan S. King[2]*

[1]*Department of Geography and Environmental Systems, University of Maryland-Baltimore County, Baltimore, MD 21205, USA; and* [2]*Center for Reservoir and Aquatic Systems Research, Department of Biology, Baylor University, One Bear Place #97388, Waco, TX 76798, USA*

Summary

1. Existing methods for identifying ecological community thresholds are designed for univariate indicators or multivariate dimension-reduction of community structure. Most are insensitive to responses of individual taxa with low occurrence frequencies or highly variable abundances, properties of the vast majority of taxa in community data sets. We introduce Threshold Indicator Taxa ANalysis (TITAN) to detect changes in taxa distributions along an environmental gradient over space or time, and assess synchrony among taxa change points as evidence for community thresholds.

2. TITAN uses indicator species scores to integrate occurrence, abundance and directionality of taxa responses. It identifies the optimum value of a continuous variable, x, that partitions sample units while maximizing taxon-specific scores. Indicator z scores standardize original scores relative to the mean and SD of permuted samples along x, thereby emphasizing the relative magnitude of change and increasing the contributions of taxa with low occurrence frequencies but high sensitivity to the gradient. TITAN distinguishes negative ($z-$) and positive ($z+$) taxa responses and tracks cumulative responses of declining [sum($z-$)] and increasing [sum($z+$)] taxa in the community. Bootstrapping is used to estimate indicator reliability and purity as well as uncertainty around the location of individual taxa and community change points.

3. Using two simulated data sets, TITAN correctly identified taxon and community thresholds in more than 99% of 500 unique versions of each simulation. In contrast, multivariate change-point analysis did not distinguish directional taxa responses, resulting in much wider confidence intervals that in one instance failed to capture thresholds in 38% of the iterations.

4. Retrospective analysis of macroinvertebrate community response to a phosphorus gradient supported previous threshold estimates, although TITAN produced narrower confidence limits and revealed that several taxa declined at lower levels of phosphorus. Re-analysis of macroinvertebrate responses to an urbanization gradient illustrated disparate change points for declining (0·81–3·3% urban land) and increasing (6·8–26·6%) taxa, whereas the published threshold estimate (20–30%) missed the declining-taxa threshold because it could not distinguish their synchronous decline from the gradual increase in ubiquitous taxa.

5. *Synthesis and applications*. By deconstructing communities to assess synchrony of taxon-specific change points, TITAN provides a sensitive and precise alternative to existing methods for assessing community thresholds. TITAN has tremendous potential to inform conservation of rare or threatened species, develop species sensitivity models, identify reference conditions and to support development of numerical regulatory criteria.

Key-words: biodiversity conservation, change-point analysis, community analysis, constrained classification, ecological thresholds, indicator species, regime shift, species sensitivity distribution

Introduction

Ecologists have become increasingly interested in analytical methods for detecting and quantifying ecological thresholds (Brenden, Wang, & Su 2008; Andersen *et al.* 2009; Sonderegger *et al.* 2009). Ecological thresholds may be defined as transition points or zones of relatively rapid change between alternate ecosystem states or ecological condition, often in response to small, continuous changes in one or more causal variables (Toms & Lesperance 2003; Huggett 2005; Groffman *et al.* 2006).

Ecological thresholds may be particularly relevant in the context of anthropogenic environmental gradients because such gradients can represent novel physical and chemical conditions falling outside those experienced by species in evolutionary time. Coevolved communities of interacting species possess unique morphological, behavioural and physiological adaptations, often corresponding to a narrow range of environmental conditions (*sensu* Shelford 1913). Species distributions under otherwise unaltered conditions may simultaneously and abruptly change at a critical level of a novel environmental gradient (May 1977; King & Richardson 2003; Sasaki *et al.* 2008). Detection and description of such ecological *community* thresholds has important implications both for ecological theory and application (Townsend, Uhlmann, & Matthaei 2008; Martin & Kirkman 2009).

Ecological community thresholds may be distinct from ecosystem-level, univariate thresholds described by current models (Groffman *et al.* 2006; Andersen *et al.* 2009; Suding & Hobbs 2009) because community responses are multivariate (i.e., one dimension for each taxon). Such thresholds are theoretically relevant to ecologists because of the evolutionary implications of a synchronous response of species to environmental pressures (Huggett 2005; Okland, Skarpaas, & Kausrud 2009). Potential applications of community thresholds include supporting the development of numerical environmental criteria to prevent loss of biodiversity and ecosystem function (King & Richardson 2003) or identification of biological reference conditions to better characterize community dynamics in the absence of disturbance (Wang, Robertson, & Garrison 2007; Brenden *et al.* 2008; Utz, Hildebrand, & Boward 2009).

Current statistical methods used for identifying thresholds were not developed for multivariate species abundance data (Brenden *et al.* 2008; Andersen *et al.* 2009). The vast majority of taxa in community data sets have low occurrence frequencies (i.e., do not occur in a large proportion of the sample units) and highly variable abundances (McCune & Grace 2002). Consequently, most investigators aggregate community data into univariate responses, selecting *a priori* attributes that presumably represent an important facet of community structure, such as the number of taxa or deriving synthetic variables from multivariate analysis of taxa composition among sites (e.g., dissimilarity metrics, ordination axes; King and Richardson 2003, Walsh *et al.* 2005). While aggregating taxa into one or more response variables may, in some instances, increase community signal in response to anthropogenic gradients,

it also likely obscures nonlinear changes in one or more taxa, potentially underestimating or misrepresenting the effect of an anthropogenic gradient on ecological communities. Thus, evaluating ecological community thresholds with existing approaches often involves undesirable generalities, loss of information or assumptions regarding taxon-specific responses.

We introduce a new analytical approach, Threshold Indicator Taxa ANalysis (TITAN), with the goals of (i) exploring and identifying abrupt changes in both the occurrence frequency and relative abundance of individual taxa along an environmental, spatial or temporal gradient; (ii) quantifying uncertainty around locations of abrupt change; and (iii) estimating the relative synchrony and uncertainty of those changes as a nonparametric indicator of a community threshold. We contend that a flexible, taxon-specific approach can yield insight informing and augmenting understanding obtained from existing analytical methods. We illustrate accuracy and diagnostic advantages of TITAN using simulated community data and retrospective analysis of two aquatic macroinvertebrate data sets spanning different types of anthropogenic environmental gradients.

Materials and methods

BACKGROUND AND COMPONENT CALCULATIONS

TITAN combines and extends change-point analysis and indicator species analysis. Change-point analysis (nCPA; King & Richardson 2003; Qian, King, & Richardson 2003) is a nonparametric technique that orders and partitions observations along an environmental gradient, identical to a single-split, multivariate regression tree analysis (MRT; De'Ath 2002). MRT replaces the univariate response of typical regression tree (RT; Breiman *et al.* 1984; De'Ath & Fabricius 2000) with a measure of multivariate dissimilarity among sample pairs. In both nCPA and MRT, optimal partitioning is obtained by maximizing a deviance reduction statistic that compares within-group vs. between-group dissimilarity determined by a user-selected ecological distance metric. As optimal partitioning can be sensitive to sample distribution along the environmental gradient, nCPA adds a bootstrap resampling procedure to assess uncertainty associated with the observed change-point value (King & Richardson 2003; King *et al.* 2005).

TITAN replaces the aggregate, community-level, dissimilarity response of nCPA with taxon-specific, indicator value (IndVal) scores from indicator species analysis (Dufrêne & Legendre 1997). Indicator species analysis is a widely accepted method for identifying indicator taxa in noisy biological data, pruning dendrograms from a hierarchical cluster analysis to an optimal number of groups, or evaluating how sampling unit groupings explain species distributions. IndVal scores are a simple and intuitive measure designed specifically to capture strength-of-association between any particular taxon and an external sample grouping (e.g., an *a priori* cluster analysis typology).

Indicator species analysis produces an IndVal score estimating the association of each taxon with each group. Thus, two IndVal scores are computed for a single taxon in a two-group classification. IndVal scores are the product of cross-group relative abundance (proportion of abundance among all sample units belonging to group *i*) and within-group occurrence frequency (proportion of sample units in group *i* with a positive abundance value). IndVal uses occurrence

frequency within each group to weight a taxon's relative abundance by how consistently it is observed in each group. A large abundance within one sample group results in a greater IndVal score only if the taxon also occurs with great regularity in that same group. IndVal scores are superior to simple abundance as a measure of association because integration of occurrence frequency and abundance leads to a measure of association *that is unbiased by group size* (Dufrêne & Legendre 1997). IndVal scores are scaled from 0% to 100% with a value of 100 indicating that a taxon was collected in every sample within a group and not in any other group. Conversely, in a two-group classification, a value of 50 can mean that a taxon occurred in half the samples within only one group, or in equal abundances in all samples of both groups (e.g., Fig. 2). The probability of obtaining an equal or larger IndVal score from random data (P) is estimated by comparing the magnitude of each observed IndVal score with those generated when group membership is assigned via ≥ 250 randomized permutations (Dufrêne & Legendre 1997).

CHANGE-POINT IDENTIFICATION AND UNCERTAINTY FOR A SINGLE TAXON

TITAN uses IndVal scores instead of deviance reduction (as in nCPA or MRT) to identify change points across a continuous environmental gradient (x; Fig. 1, Step 1). Midpoints between observed values of x are candidate change points (x_i) used to iteratively split observations into two groups, and thus produce two IndVal scores at each split (Fig. 1, Step 2·1). The relative magnitude of IndVal scores for groups on each side of a candidate change point reflects whether a taxon shows greater association with the left (negative response with respect to x) or the right (positive response) side of each split (Fig. 1, Step 2·1C, D). The greater the difference in taxon fidelity (association) created by a specific split, the greater the IndVal score for one of the two groups. The greatest IndVal score at each split and the side of the split on which it occurs are retained for comparison with those at other candidate change points (Fig. 2). In practice, we use a minimum group size of five observations (De'Ath & Fabricius 2000), so for any sample of n observations and depending on the number of unique x-values observed, TITAN compares up to $2n-20$ IndVal scores for each taxon (i.e., $2n$ for IndVals at each split, less the 10 samples and splits needed to satisfy the minimum group size). Any value of x resulting in an IndVal maximum among candidate splits is identified as the observed change point or the optimal partition for that taxon (Fig. 2).

TITAN estimates uncertainty surrounding taxon-specific responses using the distribution of change-point values across a series of bootstrap replicates of the entire data set (Fig. 1, Step 3; Manly 1997; Toms & Lesperance 2003). The bootstrap procedure is necessary because unlike *a priori* group classification required by indicator species analysis, optimal group partitioning along x is initially unknown in TITAN, and is in fact the objective of the analysis. Whereas the permutation procedure is used to estimate the probability that an equal or larger IndVal could be obtained from random data, the bootstrap procedure estimates uncertainty around change-point locations (optimal partitioning along x), as well as consistency in the response direction of each taxon (negative or positive). Variability in change-point location, directionality (positive or negative with respect to x) and magnitude (relative to the absence of structure along x) constitute the information content of the *indicator response* for each taxon in TITAN.
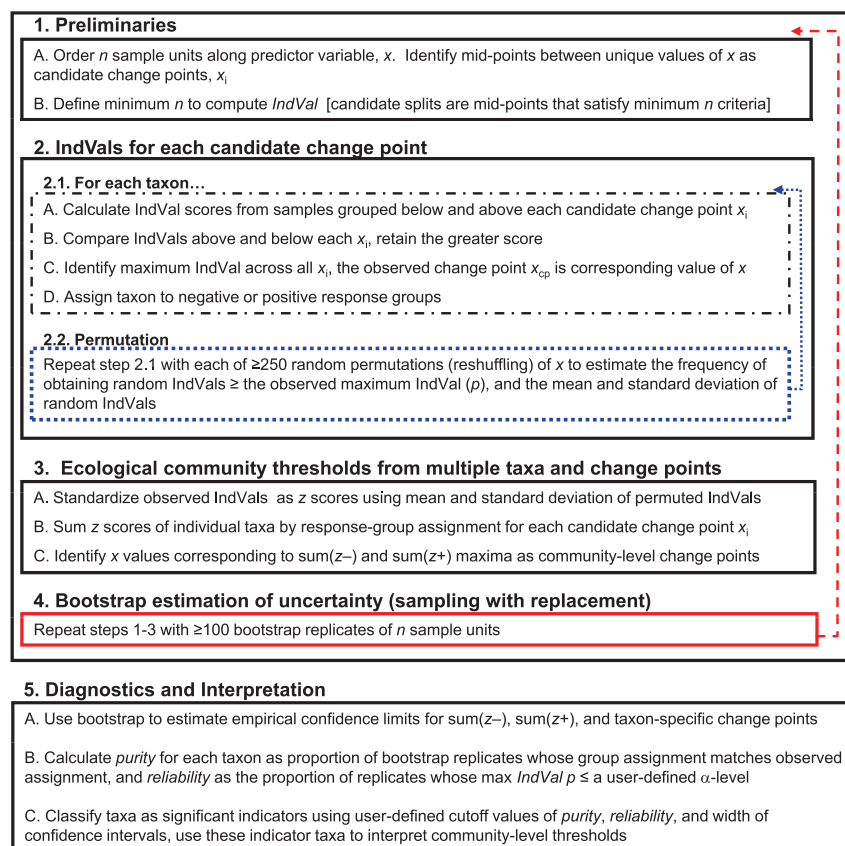
---

1. Preliminaries

A. Order n sample units along predictor variable, x. Identify mid-points between unique values of x as candidate change points, x_i

B. Define minimum n to compute *IndVal* [candidate splits are mid-points that satisfy minimum n criteria]

2. IndVals for each candidate change point

2.1. For each taxon…

A. Calculate IndVal scores from samples grouped below and above each candidate change point x_i

B. Compare IndVals above and below each x_i, retain the greater score

C. Identify maximum IndVal across all x_i, the observed change point x_{cp} is corresponding value of x

D. Assign taxon to negative or positive response groups

2.2. Permutation

Repeat step 2.1 with each of ≥ 250 random permutations (reshuffling) of x to estimate the frequency of obtaining random IndVals $\geq$ the observed maximum IndVal (p), and the mean and standard deviation of random IndVals

3. Ecological community thresholds from multiple taxa and change points

A. Standardize observed IndVals as z scores using mean and standard deviation of permuted IndVals

B. Sum z scores of individual taxa by response-group assignment for each candidate change point x_i

C. Identify x values corresponding to sum(z–) and sum(z+) maxima as community-level change points

4. Bootstrap estimation of uncertainty (sampling with replacement)

Repeat steps 1-3 with ≥ 100 bootstrap replicates of n sample units

5. Diagnostics and Interpretation

A. Use bootstrap to estimate empirical confidence limits for sum(z–), sum(z+), and taxon-specific change points

B. Calculate *purity* for each taxon as proportion of bootstrap replicates whose group assignment matches observed assignment, and *reliability* as the proportion of replicates whose max *IndVal* $p \leq$ a user-defined α-level

C. Classify taxa as significant indicators using user-defined cutoff values of *purity*, *reliability*, and width of confidence intervals, use these indicator taxa to interpret community-level thresholds

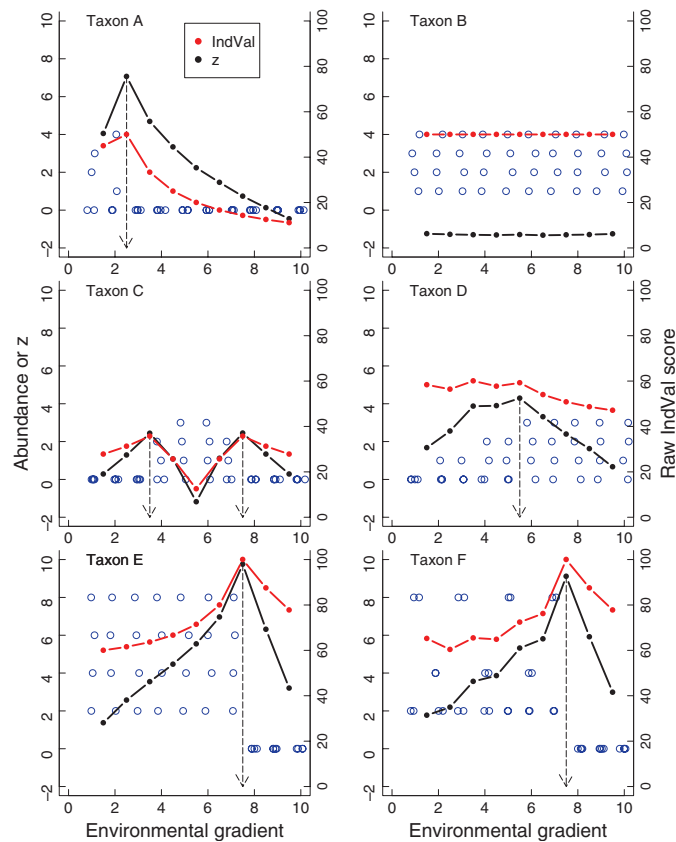**Fig. 1.** Flow chart of Threshold Indicator Taxa ANalysis (TITAN).

Fig. 2. Response of indicator value (IndVal) and *z* scores for six hypothetical taxa abundances along a uniform environmental gradient. Dashed vertical arrows illustrate the maximum *z* score and corresponding environmental change point.

Two important diagnostic indices measuring the quality of the indicator response for any taxon are obtained from bootstrap resampling: purity and reliability. Indicator *purity* is the proportion of change-point response directions (positive or negative) among bootstrap replicates that agree with the observed response. Pure indicators (e.g., purity ≥0·95) are consistently assigned the same response direction, regardless of abundance and frequency distributions generated by resampling the original data.

If bootstrap resampling substantially alters the probability of obtaining an equal or larger IndVal based on 250 random permutations of the data, then that particular taxon is not a reliable indicator. Indicator *reliability* is estimated by the proportion of bootstrap change points whose IndVal scores consistently result in *P*-values below one or more user-determined probability levels (e.g., $P \leq 0.05$). Reliable indicators (e.g., ≥0·95 of the bootstrap replicates achieving $P \leq 0.05$, or some other user-defined proportion of replicates) are those with repeatable and consistently large IndVal maxima.

For each pure indicator taxon, TITAN uses bootstrap replicates to estimate empirical quantiles of the change-point distribution. Variation in change-point estimates highlights uncertainty in the location of the maximum IndVal with respect to *x*. Sharp, nonlinear responses in taxon abundances are reflected by relatively narrow intervals between upper and lower change-point quantiles (e.g., 5%, 95%), whereas taxa with linear or more gradual responses will have broad quantile intervals spanning most of the range of *x*. If the gradient is long enough to produce a Gaussian-like response, the interval will likely encompass the mode because of reliably strong, but impure bootstrap change points. However, we discourage strict-interpretation of these quantiles as confidence limits because any method of

computing confidence limits will be unreliable for taxa with low occurrence frequencies (Manly 1997).

INDENTIFYING ECOLOGICAL COMMUNITY THRESHOLDS FROM MULTIPLE CHANGE POINTS

Once IndVals for each candidate change point and taxon are classified according to response direction, the aggregate response of all indicator taxa at each candidate change point may be used as evidence for a community-level threshold. As a dendrogram assessment tool, Dufrêne & Legendre (1997) recommended that the number of groups resulting in the largest sum of IndVal scores (scores significant at $P < 0.05$ or other criteria) be considered optimal for dendrogram pruning. This approach uses the greatest cumulative IndVal signal to distinguish groups, whether from a single taxon or many taxa, because a primary goal is to facilitate accurate classification of new observations using taxa most characteristic of each group. In contrast, evidence for a community threshold in TITAN requires substantial change across more than the most predominant taxa. As rare taxa are often highly sensitive to environmental alterations and the focus of biodiversity conservation, changes in their distribution are of great interest, although often more difficult to detect. Because the absolute value of IndVal scores is influenced by a taxon's overall abundance, it is less important in TITAN than the magnitude of change *relative to each taxon's abundance distribution* (see below). All taxa are not required to have identical IndVal maxima to produce a community threshold, rather only that large IndVal scores for many taxa occur close together along *x*. Integrating information about taxon-specific changes across groups, we apply the additive indicator

score concept in a novel way, using rescaled indicator responses to partition observations while preserving the information provided by rare taxa.

IndVal scores are rescaled as z scores within TITAN according to degree of departure from expected values by subtracting the mean of randomized permutations from the observed IndVal, and dividing by its permuted SD. Exploratory simulations suggest that permuted IndVal z scores are similar to nonparametric alternatives (i.e., substituting median and interquartile range for mean and standard deviation). We use z scores because the central focus of TITAN is somewhat distinct from the original purpose of IndVal. IndVal was developed to interpret a pre-existing sample typology (Dufrêne & Legendre 1997), whereas TITAN seeks to use IndVal scores to select among candidate groupings. Rather than raw IndVal magnitudes, which would favour the most widely distributed or abundant taxa, standardization facilitates cross-taxa comparison by emphasizing change in IndVals across candidate splits *given a specific pattern of abundance and occurrence* (Fig. 2). Rescaling makes little difference during determination of taxon-specific change points, but it can make a substantial difference during interpretation of their relative information content. Rare or infrequently occurring taxa with smaller IndVal magnitudes can have a very strong z score if their response to environmental change is dramatic. Standardized taxa responses increasing at the change point ($z+$) are distinguished from those decreasing ($z-$) and those showing no response.

Evidence for community-level thresholds among negative and positive taxa is assessed separately by tabulating and summing all $z-$ and $z+$ scores for each value of x. The value(s) of x resulting in the largest cumulative z scores for negative [sum($z-$)] and positive [sum($z+$)] responses correspond to the maximum aggregate change in the frequency and abundance of their respective taxa. Large values of sum(z) scores occur when many taxa have strong responses at a similar value of the environmental gradient, whereas weak or variable responses result in lower sum(z) values without a distinctive maximum. TITAN community-level change points may be assessed by plotting sum(z) scores vs. x, and are easily interpreted via tabular and graphical summaries of change-point distributions of individual taxa.

If sum(z) maxima involve synchronous change in many taxa, including overlap from bootstrap distributions of taxon-specific change points, then these values of x may be interpreted as evidence for observed community thresholds. Bootstrap replicates used to evaluate taxon-specific change points are also summarized to develop distributions of sum(z) responses. Variation in the bootstrapped values of x that produce the greatest sum($z-$) or sum($z+$) values is used to estimate uncertainty associated with community change points, and quantiles (e.g., 0·05, 0·95) of these distributions serve as empirical confidence limits (e.g., Qian *et al.* 2003; Toms & Lesperance 2003). Narrow confidence limits represent further evidence for a community threshold, whereas wide confidence bands suggest other response (e.g., linear, modal or random) dynamics are more likely.

Case studies

SIMULATED COMMUNITY DATA

We developed two simulations to evaluate how TITAN deals with taxa with distinct distributions, classifies responses, and detects change points. By controlling the statistical properties of the response and predictor variables across 500 unique data sets generated by each simulation, we demonstrate TITAN's efficacy and flexibility.

The first simulation involved a shift in respective taxa at two distinct values of x. Eight taxa abundances were simulated along a uniformly distributed environmental gradient (runif in R 2·9·2; $n = 100$, range = 0–100). Abundances were generated using a negative binomial distribution (rnbinom in R 2·9·2) to simulate noisy, heteroscedastic and sparse site-by-taxa matrices typical of community data (McCune & Grace 2002). Two taxa (Sp1–2) were thus assigned frequency and abundance values that differed below and above a value of 40 along the environmental gradient to simulate samples from populations with a threshold decline (Fig. 3). Likewise, three taxa (Sp6–8) were assigned increasing frequency and abundance values above 60 to assess TITAN's ability to distinguish disparate change points and different response directions. One taxon (Sp3) was assigned an increasing change point at 40 and a decreasing change point at 60 to approximate a unimodal distribution. Finally, two taxa (Sp4–5) differed in frequency and abundance but varied randomly with respect to the environment. The entire simulation was repeated 500 times to generate different data sets, obtain diagnostic statistics, and evaluate the ability of nCPA (Bray-Curtis distance among sample units) and TITAN to correctly identify and assess thresholds in the data.

The second scenario involved similar distributions of the environmental gradient to contrast threshold responses with noisy data and generalized, wedge-shaped distributions typical of complex taxon responses to multiple limiting factors (e.g., Cade & Noon 2003; Brenden *et al.* 2008; Konrad *et al.* 2008). In this scenario, two taxa (Sp1–2) were assigned decreasing frequency and abundance values to simulate samples from populations with a threshold decline at 20 (Fig. 4). Three taxa (Sp6–8) were assigned probabilities of frequency and abundance that increased in proportion to x (i.e., generating wedge-shaped distributions). Sp3 was assigned an increasing change point at 20 and a decreasing change point at 60 to approximate a broad, unimodal distribution, whereas Sp4–5 varied randomly with respect to the environment. This simulation was also repeated 500 times to obtain diagnostic statistics.

EVERGLADES DATA

These data were taken from a previous study designed to identify a concentration of surface-water total phosphorus (TP) that corresponded to abrupt changes in macroinvertebrate species composition in the Florida Everglades, USA (King & Richardson 2003). Macroinvertebrate densities (no/m^2, 164 taxa, species or morphospecies-level taxonomy) were measured from 126 marsh sampling stations along a 10-km TP gradient. Concentrations of TP in the data set ranged from < 10 µg/L to > 100 µg/L. The authors used several community variables and estimated TP change points using univariate nCPA. The resulting change points ranged from *c.* 10 µg/L to 25 µg/L TP, and authors concluded that TP > 12–15 µg/L likely corresponded to ecologically significant changes in taxonomic composition.

MARYLAND STREAM DATA

These data were the subject of a previous study on analytical considerations for linking watershed land cover to ecological communities in streams (King *et al.* 2005). In the previous analysis, we used axis scores from non-metric multidimensional scaling (nMDS) of Bray-Curtis dissimilarity in nCPA to identify a level of watershed percent developed land corresponding to an abrupt change in macroinvertebrate community composition in wadeable streams (295 sites, 177 taxa abundances, mostly genus-level identification). Our previous analysis identified a relatively sharp change
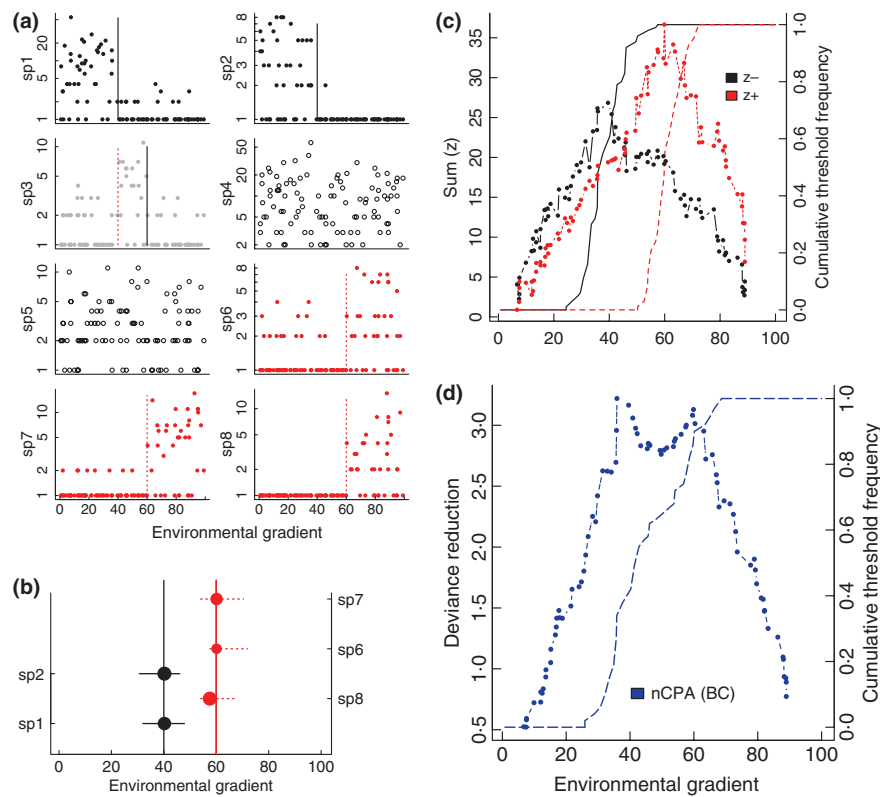
Fig. 3. Threshold Indicator Taxa ANalysis (TITAN) and change-point analysis (nCPA, Bray-Curtis distance) of a two-threshold community response to a simulated environmental gradient (Simulation 1). (a) Simulated abundances of eight taxa in response to the environmental gradient (*x*-axis). According to negative binomial probability distributions used in simulating frequency and abundance, Sp1–2 should decline at 40 (black vertical line), Sp3 should increase at 40 (red vertical line) and decrease at 60 (black vertical line), Sp4–5 should not vary predictably with the environment, and Sp6−8 should increase at 60 (red vertical line). (b) Pure (≥0·95) indicator taxa are plotted in increasing order with respect to their observed environmental change point. Black symbols correspond to negative (*z*−) indicator taxa, whereas red corresponds to positive (*z*+) indicator taxa. Symbols are sized in proportion to *z* scores. Horizontal lines overlapping each symbol represent 5th and 95th percentiles among 500 bootstrap replicates. Vertical lines indicate the simulated true values for negative (black vertical line) and positive (red vertical line) underlying thresholds. (c) TITAN sum(*z*−) and sum(*z*+) values corresponding to all candidate change points (*x_i*) along the environmental gradient. Black and red vertical lines represent the cumulative frequency distribution of change points (*x_cp*, or thresholds) among 500 bootstrap replicates for sum(*z*−) and sum(*z*+), respectively. (d) Deviance reduction in Bray-Curtis distance values for each candidate change point (*x_i*) along the environmental gradient. The dashed blue line represents the cumulative frequency distribution of change points (thresholds) among 500 bootstrap replicates.

in community composition (inferred from nMDS axis 1 scores) between 20% and 30% watershed developed land (5th–95th bootstrap percentiles).

DATA ANALYSES

We performed TITAN analysis on all four data sets in R (R Development Core Team 2009, version 2·9·2) using a custom package TITAN (see Appendix S3) written by MEB and RSK. nCPA was performed using a custom function within TITAN based on the db-MRT method of De'Ath (2002) in the package mvpart. We log$_{10}(x + 1)$ transformed taxa abundances to reduce the influence of highly variable taxa on indicator score calculations in each data set, which was particularly important for taxa with low occurrence frequencies. Taxa with <5 occurrences were deleted (following previous analyses of these data) and we used Bray-Curtis distance as the dissimilarity metric for all nCPA assessments (King *et al.* 2005).

We compared TITAN and nCPA by plotting sum(*z*) and deviance reduction values as a function of increasing values of *x* and identified community change points as the *x* value that resulted in the maximum sum(*z*) or deviance reduction, respectively. We computed cumulative

change-point distributions by finding the maximum IndVal (individual taxa), TITAN sum(*z*) and nCPA deviance reduction across 500 bootstrap replicates for the simulated and real data sets. Indicator purity and two levels of reliability (proportion of bootstrap replicates with $P \leq 0.05$ and $P \leq 0.01$) were also computed from the bootstrap replicates. We used output from the 500 unique iterations of each simulated data set to compute: (i) the frequency at which the 0·05 and 0·95 bootstrap quantile intervals captured the true thresholds, (ii) median change-point values corresponding to IndVal *z*-score maxima along the environmental gradient, and (iii) mean reliability and purity of each taxon.

Results

SIMULATION 1

TITAN accurately interpreted both negative and positive indicator taxa regardless of their relative position along the environmental gradient, while diagnostic indices helped distinguish the relative information content in taxon-specific
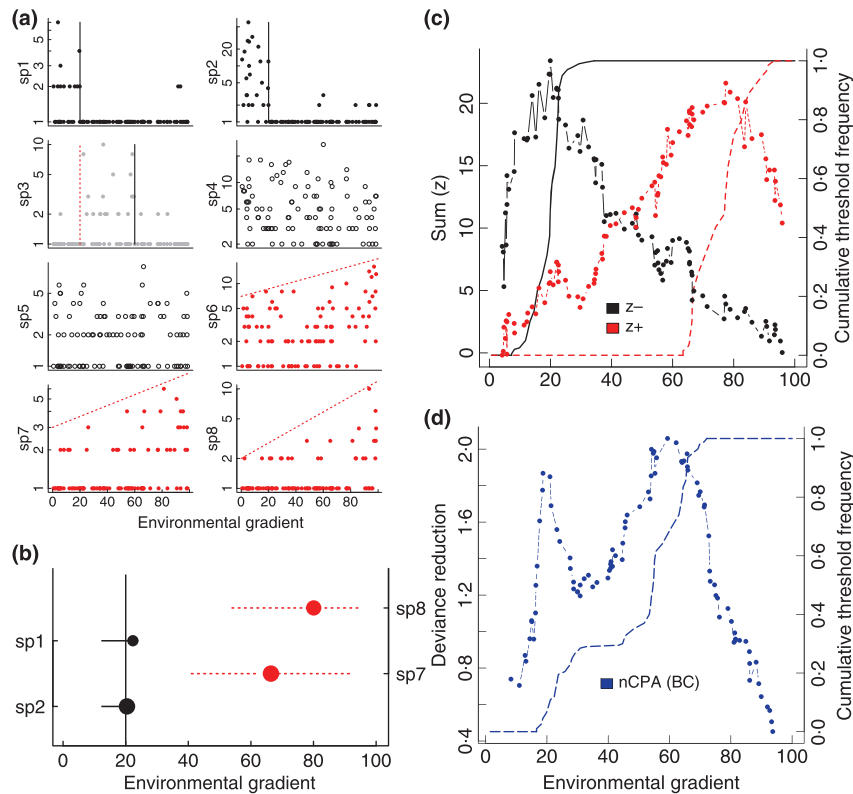
Fig. 4. Threshold Indicator Taxa ANalysis (TITAN) and change-point analysis (nCPA, Bray-Curtis distance) of a negative threshold and positive wedge-shaped community response to a simulated environmental gradient (Simulation 2). (a) Sp1–2 should decline at 20 (black vertical line), Sp3 should increase at 20 (red vertical line) and decrease at 60 (black vertical line), Sp4–5 should not vary predictably with the environment, and Sp6–8 should increase with a wedge-shaped distribution (red dotted lines) instead of a threshold response. See Fig. 3 for additional details.

distributions. All 8 taxa produced indicator assignments and observed change points consistent with the simulated parameters, and all had significant ($P \le 0.05$) raw IndVal scores in at least 46% of the 500 data sets. All six taxa simulating a threshold decline or increase had median change-point distributions that overlapped their true threshold value of x (Table 1; see Fig. 3b for example), whereas the distribution for Sp3 (simulated unimodal distribution) included both thresholds. Change-point distributions were much broader for the two random taxa (Sp7–8). All five monotonic taxa were both reliable (mean reliability over 500 data set iterations ≥0.99 for $P \le 0.05$ and $P \le 0.01$) and pure indicators (mean purity over 500 iterations ≥0.99), whereas Sp3 was impure and only moderately reliable (reliability ≥0.69 from $P \le 0.05$ to $P \le 0.01$; mean purity = 0.74). Both taxa (Sp4–5) simulating independence from the environment gradient often produced significant raw IndVal scores at $P \le 0.05$ (64% and 61%, respectively), but were much less reliable at $P \le 0.01$ (38%, 35%) and remained impure indicators (mean purity 0.74, 0.72). The degraded reliability was expected for random taxa, but in the case of Sp3, limited reliability suggests its distribution often resulted in IndVals indistinguishable from random scores.

As an assessment of community-level thresholds, TITAN sum($z-$) peaked at a median of 39.97 (mean = 40.21) across

500 data sets, corresponding closely with the true threshold value of 40 for negative indicator taxa (Table 2; see examples in Fig. 3a and c). The sum($z+$) median peaked at 60.04 (mean = 60.22) close to the simulated true threshold of 60. Cumulative change-point frequency distributions for bootstrap replicates were relatively narrow at the 90% level and straddled the true thresholds for negative and positive taxa in 100% of simulation data sets. nCPA deviance reduction reached its maximum at a median of 59.58 (mean = 53.58; Fig. 3d). The inner 90% of the change-point distribution was broader than either sum(z) and overlapped the positive threshold in 91% of the iterations, but was slightly above the negative threshold more often, capturing the true value in only 62% of the data sets. This comparison illustrated that despite near-synchronous thresholds among negative and positive responders, TITAN distinguished the two precisely, whereas nCPA produced broader confidence limits and a biased observed change-point estimate.

SIMULATION 2

All eight taxa produced indicator assignments and observed change points consistent with the simulated parameters, and all had significant ($P \le 0.05$) raw IndVal scores in more than 43% of the 500 data sets. Taxa with higher overall frequencies

Table 1. Threshold Indicator Taxa ANalysis (TITAN) individual taxa results from Simulations 1 and 2

	+/−	IndVal	Freq.	z	Change point Obs.	5%	95%	Capture	Purity	Reliability ≤0·05	≤0·01
Simulation 1											
Sp1	z−	82·3	47·8	15·45	40·04	30·94	48·71	1·00	1·00	1·00	1·00
Sp2	z−	54·59	29·9	10·88	39·61	25·21	48·45	0·99	1·00	1·00	1·00
Sp3	z−	28·81	16·9	6·04	57·5	38·21	61·71	0·66, 0·62*	0·74	0·74	0·72
Sp4	NA	54·82	100	1·91	48·84	8·45	92·22	NA	0·74	0·64	0·38
Sp5	NA	49·02	69·8	1·91	51·22	8·84	92·16	NA	0·72	0·61	0·35
Sp6	z+	60·61	40·3	10·47	60·33	51·16	73·03	1·00	1·00	1·00	1·00
Sp7	z+	82·19	48·3	15·5	59·94	51	68·61	1·00	1·00	1·00	1·00
Sp8	z+	59·59	40	10·37	60·33	50·9	72·78	1·00	1·00	1·00	1·00
Simulation 2											
Sp1	z−	50·41	11	12·49	19·63	9·51	27·13	0·99	1·00	1·00	0·99
Sp2	z−	83·61	26	15·51	20·08	13·04	28·5	1·00	1·00	1·00	1·00
Sp3	z−	52·42	32·6	9·63	60·18	49·49	67·63	0·99, 0·14*	0·96	0·96	0·96
Sp4	NA	55·28	100	1·91	49·2	8·74	92·16	NA	0·73	0·63	0·38
Sp5	NA	45·04	59·6	1·91	53·22	8·96	92·1	NA	0·73	0·62	0·37
Sp6	z+	49·94	51·8	4·36	71·75	30	90·37	NA	0·93	0·90	0·77
Sp7	z+	51·2	28·8	8·69	70·5	52·12	87·14	NA	1·00	1·00	0·98
Sp8	z+	45·42	24·3	8·12	71·68	51·79	88·72	NA	1·00	0·99	0·97

TITAN observed change points (obs.) and bootstrap confidence intervals (median among 500 simulation iterations) correspond to the value of x resulting in the largest indicator value (IndVal) z scores for each taxon (see Figs 3a and 4a). Capture rates show the proportion of 500 simulation iterations in which the 90% bootstrap confidence interval included the true threshold. Purity is the mean proportion of correct response direction ($z−$ or $z+$) assignments, reliability is the mean proportion of p-values ≤0·05 or ≤0·01 among 500 simulation iterations, and Freq. is the number of non-zero observations.
*Values correspond to increasing and decreasing change points, respectively for unimodal taxa.
NA, not applicable to random or wedge-shaped responses.

tended to have higher raw IndVals, but not necessarily higher z scores. Rather, z scores corresponded more closely to the relative difference between the left and right side of the threshold

Table 2. Threshold indicator taxa analysis (TITAN) community-level results from Simulations 1 and 2

Method	Obs.	0·05	0·95	Capture (−)	Capture (+)
Simulation 1					
TITAN sum(z −)	39·97	30·76	49·84	1·00	≤0·01
TITAN sum(z +)	60·04	51·41	68·39	≤0·01	1·00
nCPA (Bray-Curtis)	59·58	38·10	64·82	0·62	0·91
Simulation 2					
TITAN sum(z −)	20·17	12·53	28·22	1·00	NA
TITAN sum(z +)	70·21	54·27	85·08	≤0·01	NA
nCPA (Bray-Curtis)	20·03	15·35	61·29	0·93	NA

TITAN observed change points (Obs.) and bootstrap 5th and 95th quantiles of change points (median among 500 simulation iterations) correspond to the value of the x resulting in the largest sum of indicator value (IndVal) z scores among all negative ($z−$) and positive ($z+$) taxa, respectively (see Figs 3c and 4c), whereas nCPA thresholds correspond to the maximum deviance reduction among sample units (Bray-Curtis distance; Figs 3d and 4d). Capture rates show the proportion of 500 simulation iterations in which the 90% bootstrap confidence interval included the true threshold.
NA, not applicable.

in the assigned probabilities of the negative binomial distribution function used to simulate abundances (Table 1; see Fig. 4b for example). Both taxa simulating threshold declines had median change-point distributions that overlapped their true change-point value of x, whereas the distribution for Sp3 included only the declining change point. All five monotonic taxa and unimodal Sp3 were both reliable and pure indicators (reliability ≥0·96 from $P ≤ 0.05$ to $P ≤ 0.01$; mean purity over 500 data sets ≥0·92) with the exception of Sp6 (wedge-shaped increaser; inconsistent reliability; mean purity = 0·93). Both random taxa (Sp4–5) produced significant raw IndVal scores at $P ≤ 0.05$ in 63% and 62% of the iterations, respectively, but were much less reliable at $P ≤ 0.01$ (38%, 37%) and were impure indicators (mean purity 0·73, 0·73).

During assessment of community-level response, TITAN sum($z−$) peaked at a median of 20·17 (mean = 20·02) across 500 iterations, corresponding closely with the true threshold value of 20 for negative indicator taxa (Table 2; see examples in Fig. 4a and c). Cumulative change-point frequency distributions for bootstrap replicates were quite narrow at the 90% level (Fig. 4c), yet straddled the true threshold in 100% of simulation data sets. The sum($z+$) median peaked at 70·21 (mean = 69·78) and appropriately exhibited a broader cumulative frequency distribution of change points in accordance with the gradual, wedge-shaped increasing response for Sp6–8 (Fig. 4b and d). The nCPA deviance reduction reached its maximum at a median of 20·03 (mean = 25·34), close to

the negative threshold decline, and increased again between 55 and 75. The inner 90% of the change-point distribution was broader than either sum(z), stair-stepped, and overlapped the negative threshold in 93% of simulation data sets. This comparison illustrated that when negative and positive responders have asynchronous change points, the aggregate dissimilarity metric used by nCPA can be biased by different change-point locations.

EVERGLADES DATA

The majority of the negative ($z-$) indicator taxa declined sharply between 8 and 20 μg/L TP, resulting in a sum($z-$) change point of 14·5 μg/L (Fig. 5a and b; Table 3; Appendix S1). Positive ($z+$) indicators increased sharply between 20 and 40 μg/L resulting in a distinct sum($z+$) peak at 30·3 μg/L. Most individual taxa change points overlapped considerably in the 15–40 μg/L range, providing evidence in support of an ecological community threshold.

The observed peak in nCPA deviance reduction occurred at 18·6 μg/L, with a secondary peak at ~35 μg/L (Fig. 5d). The cumulative distribution of nCPA change points spanned the range of both sum($z-$) and sum($z+$) distributions. TITAN sum($z-$) and the lower end of the nCPA change-point distribu-

tion supported TP community changes reported in King & Richardson (2003), although sum($z-$) results implied a much sharper, synchronous change with TP. This comparison demonstrates the increased precision obtained through taxon-specific analysis in TITAN.

MARYLAND STREAM DATA

TITAN identified > 20 taxa with synchronous declines in response to developed land between 0·5% and 2·5%, resulting in a distinct peak in sum($z-$) at 1·89% (Fig. 6a and b; Table 3; Appendix S2). The cumulative distribution of sum($z-$) change points among bootstrap replicates was quite narrow (Fig. 6b; Table 3). A few additional taxa fell out in an approximately linear sequence from 5% to 20% developed land. The strong synchrony of change in many taxa at low levels of development was consistent with an ecological community threshold.

Relatively few taxa exhibited positive associations with increasing amounts of watershed development (Fig. 6a and b; Appendix S2). Those that did were widely distributed along the developed-land gradient, spanning most of the range of values and approximating a linear distribution of observed taxon change points with increasing urbanization (Fig. 6a). The asynchronous
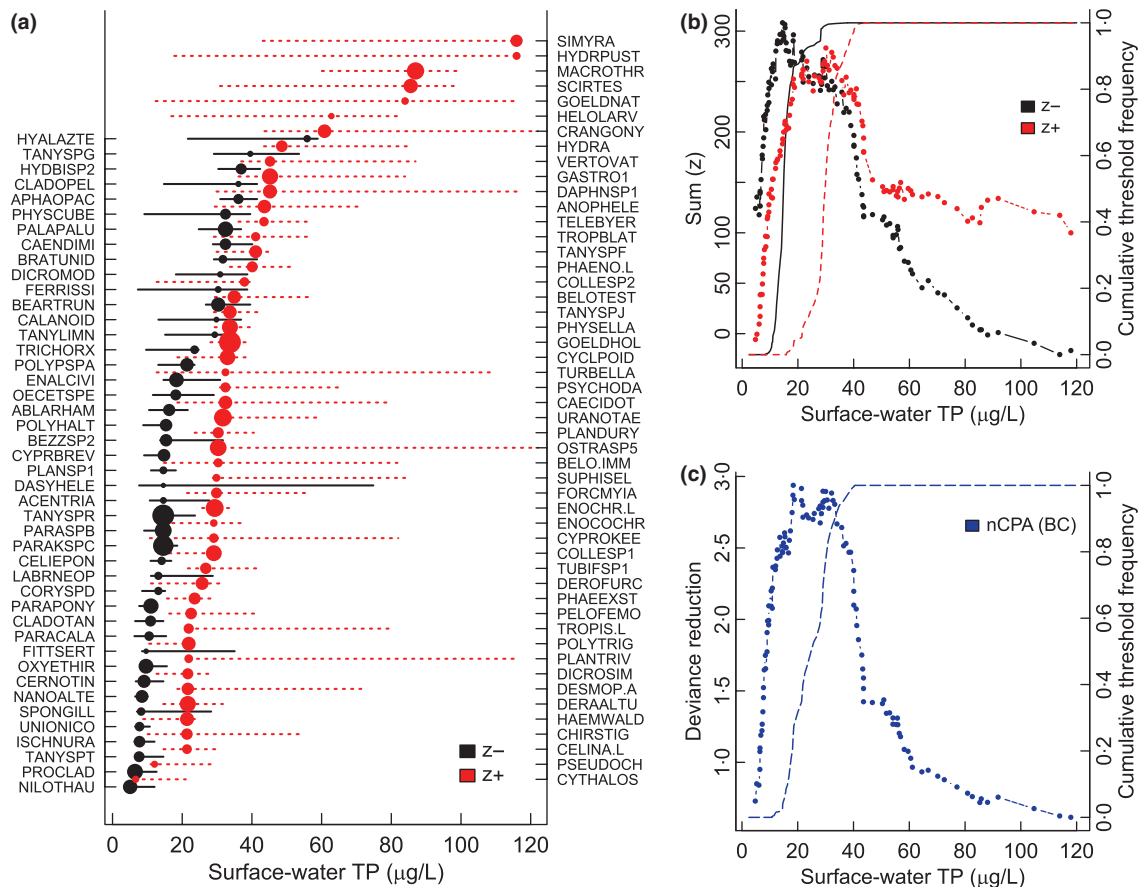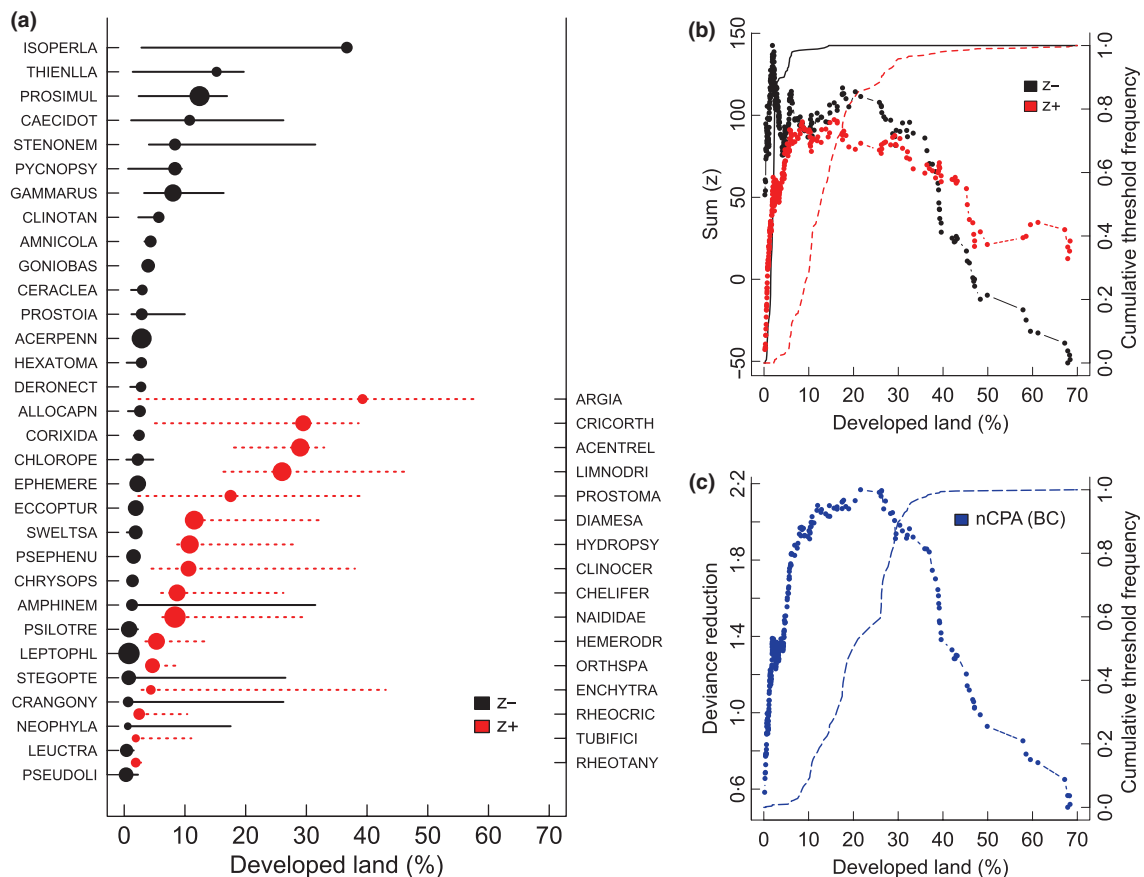

Fig. 5. Threshold Indicator Taxa ANalysis (TITAN) and change-point analysis (nCPA, Bray-Curtis distance) of macroinvertebrate community response to a surface-water total phosphorus gradient in the Everglades (n = 126). For clarity, only taxa with purity = 1·0 are shown in (a). Taxa codes in (a) are explained in Appendix S1. See Fig. 3 for additional details.

Table 3. TITAN community-level thresholds estimated from macroinvertebrate taxa responses to total phosphorus (TP) in the Everglades and percent developed land in Maryland watersheds

Method	Everglades (TP, μg/L)				Maryland (developed land)			
	Obs.	5%	50%	95%	Obs.	5%	50%	95%
TITAN sum($z-$)	14·5	11	14·7	21·6	1·89	0·81	1·96	3·38
TITAN sum($z+$)	30·3	21·3	29·8	39·5	16·3	6·15	13·3	26·2
nCPA (Bray-Curtis)	18·6	16·8	25·7	36·1	23·7	10·5	21	30·6

TITAN observed change points (Obs.) correspond to the value of the x resulting in the largest sum of indicator value (IndVal) z scores among all negative ($z-$) and positive ($z+$) taxa, respectively, whereas nCPA thresholds correspond to the maximum deviance reduction among sites (Bray-Curtis distance). Quantiles (5%, 50%, 95%) correspond to change points from 500 bootstrap replicates.

Fig. 6. Threshold Indicator Taxa ANalysis (TITAN) and change-point analysis (nCPA, Bray-Curtis distance) of macroinvertebrate community response to a watershed developed land gradient in Maryland ($n = 295$). Taxa codes in (a) are explained in Appendix S2. See Fig. 3 for additional details.

distribution of positive taxa change points meant that the corresponding maximum of their sum($z+$) showed a relatively weak (poorly defined) peak at 16·3% developed land (Fig. 5b). Further, tolerant taxa exhibited relatively wide bootstrap frequency distributions representing substantial uncertainty about the existence of a threshold because of gradual increases in frequency and abundance. The lower distribution of most tolerant taxon change points only marginally overlapped with those from the majority of declining, sensitive taxa.

nCPA deviance reduction trends closely resembled TITAN sum($z+$) (Fig. 6b and c). nCPA peaked at 23·7% developed

land and yielded bootstrap frequency distributions similar to sum($z+$), although nCPA distributions were skewed slightly higher. nCPA bootstrap frequency distributions did not overlap with TITAN sum($z-$) bootstrap distributions. The nCPA and TITAN sum($z+$) change points corresponded closely to the threshold range of 20–30% developed land reported in King *et al.* (2005), but these estimates were far higher than the TITAN sum($z-$) threshold. This comparison demonstrates how the subtle bias exhibited by nCPA analysis of aggregate metrics in Simulation 2 can be exacerbated to distort estimates of community change.

Discussion

Estimation of taxon-specific change points is arguably the key output from TITAN because this information is precisely what many aggregate community metrics obscure. Response to any novel environment likely differs when considering different biophysical gradients or taxon-specific life history, so discernment of negative and positive response patterns makes sense from an evolutionary perspective (Huggett 2005). There is little reason to expect that all taxa will respond to environmental perturbation at the same level or in the same manner, and responses may vary considerably between infrequent and more widespread taxa. Further, taxa positively associated with anthropogenic environmental gradients do not represent well-organized communities, but rather are a comprised of historically native taxa that either directly (resource subsidy) or indirectly (e.g., realized niche expansion, reduced competition or predation) benefit from it, or an invasive taxon that was not historically present but is able to cross ecosystem boundaries because of a variety of factors related to the novel environment. Therefore, unbiased measures of association along environmental gradients are critical for distinguishing response type, magnitude, and interpreting their relevance.

TITAN deconstructs community-level dissimilarity to assess synchrony of taxon-specific change points. Output can be used alone in novel biodiversity conservation applications, analogous to a species sensitivity distribution used in ecological risk assessment (Newman, Ownby, *et al.* 2000). As in other binary partitioning techniques, TITAN will find taxon-specific or sum(z) maxima in most data sets, so interpretation of change-point values makes little sense without measures of uncertainty and information content. While this approach will not prove the existence of ecological community thresholds, graphical and tabular display of change-point distributions provides simple and intuitive supporting evidence for interpreting biotic responses, as do diagnostic measures of rescaled IndVals, uncertainty, reliability and purity.

Our simulations demonstrate TITAN's precision, efficacy and flexibility. In the first simulation, the negative (z−) and positive (z+) indicator taxa thresholds were indistinguishable by nCPA, yet they remained distinct for TITAN. In the second simulation, we illustrated how, through wide confidence bands and reduced sum(z) scores, TITAN communicates uncertainty when underlying abundance distributions do not show a clear threshold response. In both cases, IndVal scores emphasized sharp changes in frequency and abundance, but many (>40%) randomly generated distributions were nonetheless deemed to contain significant change points following permutation. This pattern illustrated how frequent or abundant taxa with only modest differences in IndVals between groups are often statistically significant ($P \leq 0.05$) in large data sets despite dubious ecological significance. However, such patterns are readily distinguished from more meaningful responses through the diagnostic use of reliability and purity. We note that those taxa deemed significant by permutation do not always

achieve reliability or purity ≥ 0.95, but taxa with reliability or purity ≥ 0.95 are by definition significant at $P \leq 0.05$, and usually much lower.

The Everglades data set represented a synchronous, balanced shift in taxonomic structure during the transition from an oligotrophic community to one more typically associated with southeastern USA wetlands where nutrient limitation is less severe (King & Richardson 2003; Qian, Pan, & King 2004). In contrast to the two closely related thresholds in Simulation 1, this example illustrates how, where a relatively synchronous shift from negative to positive indicator taxa exists, nCPA and TITAN produce similar results. However, the sharp decline of several negative indicator taxa prior to any observed increase in positive indicator taxa with increasing TP resulted in a lower sum(z−) threshold and narrower quantile intervals than nCPA and sum(z+). Thus, through TITAN, individual taxon or sum(z) graphics and quantitative indices may be used in a diagnostic fashion to better understand the general nature of community response.

In the Maryland data, we illustrate new insight gained by distinguishing between taxa subsets with distinct responses to an environmental gradient. Here, most observed taxa are native and evolved without the hydrological and chemical alterations linked to urbanization (e.g., Walsh *et al.* 2005), thus watershed development likely represents a novel selective pressure. Taxa sensitive to the novel environment (or a strong covariate) responded synchronously at strikingly and remarkably low levels of development and exhibit a classic pattern consistent with decline or even potential extirpation. King *et al.* (2005) missed this pattern because we relied upon a whole-community metric similar to the nCPA results in this article. The multivariate nCPA approach did detect a change, but it did not distinguish which taxa were changing and how substantial this change was for each taxon. In this example, so-called 'tolerant' taxa occurred all along the gradient and simply became generally more prevalent in streams draining more heavily developed watersheds. In contrast with negative-responding taxa, positive or tolerant responses did not support characterization as a threshold because of wide confidence bands and gradual cumulative change-point probabilities, much like Sp6–8 in Simulation 2. Because loss of sensitive taxa or other environmental changes may not benefit all community remnants in the same way, the causes for increases in tolerant taxa along the watershed development gradient are likely taxon-specific and there is no reason to expect a synchronous tolerant community response.

In both simulations and the Maryland data set, nCPA was necessarily less sensitive to disparate change points. However, there was clearly a tendency in each data set for nCPA to average away the acute response of some taxa with that of widely distributed taxa. In the Everglades, where negative and positive threshold taxa were relatively synchronous, nCPA fell between the negative (z−) and positive (z+) responses. In both Simulation 1 and Maryland, nCPA cumulative probabilities tracked the positive (z+) response more closely and were

more likely to capture a greater number of positive threshold values. In Simulation 2, the mean nCPA value shifted to the right of its median, whereas the mean and median sum($z-$) were nearly identical. Our contrast between nCPA and TITAN should not be construed as criticism, but rather an illustration of how TITAN can be used to help interpret results from explanatory or predictive models developed using nCPA or MRT, and how under conditions where taxa turnover is synchronous on both sides of a change point, nCPA and TITAN will yield comparable results.

Sample size, minimum split size and taxon frequency are important considerations in TITAN. Small numbers of sample units will necessarily result in uncertain change-point estimates, but an appropriate minimum sample size will be data set specific. At least three and preferably five sample units should be used as the minimum split size to compute z scores. Similarly, we tentatively recommend that taxa with < 3–5 occurrences be excluded because these taxa are too infrequent to estimate interpretable z scores. In general, the recommendations of De'Ath & Fabricius (2000) for RT, De'Ath (2002) for MRT and Dufrêne & Legendre (1997) for IndVal will also apply to TITAN for these considerations.

As with any classification model, long gradients with high beta-diversity or multiple change points on the same gradient (e.g., multiple interacting limiting factors, Gaussian or otherwise modal distributions) will require recursive partitioning of sample units to isolate thresholds. In Simulation 1, TITAN was able to distinguish unimodal Sp3 as reliably impure, but in Simulation 2, the positive response was close enough to the edge of the gradient that resampling only emphasized the negative threshold. TITAN is identical in form to tree-based modelling, thus could be used in a similar way to identify covarying secondary and tertiary limiting factors corresponding to sharp community changes. Our focus here was on single-predictor thresholds, so evaluation of additional variables fell beyond the scope of this paper. However, multiple predictors may be useful in identifying a hierarchical set of environmental limiting factors of species distributions, or to assess variable importance.

Emerging analytical techniques such as Random Forests (Breiman 2001) can be powerful methods for exploring community responses along multiple environmental gradients, and TITAN would fit naturally within a similar framework. At each level of the tree, sum(z) could be contrasted among several covarying predictors and the variable that consistently had the highest sum(z) among bootstrap replicates could be selected as the best predictor, as done in current tree-based predictive models (e.g., De'Ath 2007). We also recommend examination of the continuous trend in sum(z) along with individual taxa change points, where secondary peaks may reveal additional transition points in community structure. TITAN and extensions of this method should prove useful for detecting taxon-specific and community-level thresholds and for addressing a variety of basic and applied ecological questions.

References

Andersen, T., Carstensen, J., Hernandez-Garcia, E. & Duarte, C.M. (2009) Ecological regime shifts: approaches to identification. *Trends in Ecology and Evolution*, **24**, 49–57.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth Int, Belmont.

Brenden, T.O., Wang, L. & Su, Z. (2008) Quantitative identification of disturbance thresholds in support of aquatic resource management. *Environmental Management*, **42**, 821–832.

Cade, B.S. & Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, **1**, 412–420.

De'Ath, G. (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology*, **83**, 1105–1117.

De'Ath, G. (2007) Boosted trees for ecological modelling and prediction. *Ecology*, **88**, 243–251.

De'Ath, G. & Fabricius, K.E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**, 3178–3192.

Dufrêne, M. & Legendre, P. (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, **67**, 345–366.

Groffman, P.M., Baron, J.S., Blett, T., Gold, A. J., Goodman, I., Gunderson, L. H. *et al.* (2006) Ecological thresholds: they key to successful environmental management or an important concept with no practical application? *Ecosystems*, **9**, 1–13.

Huggett, A.J. (2005) The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation*, **124**, 301–310.

King, R.S. & Richardson, C.J. (2003) Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management*, **31**, 795–809.

King, R.S., Baker, M.E., Whigham, D.F., Weller, D.E., Kazyak, P.F. & Hurd, M.K. (2005) Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications*, **15**, 137–153.

Konrad, C.P., Brasher, A.M.D. & May, J.T. (2008) Assessing stream flow characteristics as limiting factors on benthic invertebrate assemblages in streams across the western United States. *Freshwater Biology*, **53**, 1983–1998.

Manly, B.F.J. (1997) *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, 2nd edn. Chapman & Hall, London, UK.

Martin, K.L. & Kirkman, L.K. (2009) Management of ecological thresholds to re-establish disturbance-maintained herbaceous wetlands of the south-eastern USA. *Journal of Applied Ecology*, **46**, 906–914.

May, R.M. (1977) Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, **269**, 471–477.

McCune, B. & Grace, J.B. (2002) *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, OR, USA.

Newman, M.C., Ownby, D.R., Mezin, L.C.A., Powell, D.C., Christensen, T.R.L., Lerberg, S.B. & Anderson, B. (2000) Applying species-sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient numbers of species. *Environmental Toxicology and Chemistry*, **19**, 508–515.

Okland, B., Skarpaas, O. & Kausrud, K. (2009) Threshold facilitations of interacting species. *Population Ecology*, **51**, 513–523.

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

Qian, S.S., Pan, Y. & King, R.S. (2004) Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. *Ecological Indicators*, **4**, 29–37.

Sasaki, T., Okayasu, T., Jamsran, U. & Takeuchi, K. (2008) Threshold changes in vegetation along a grazing gradient in Mongolian rangelands. *Journal of Ecology*, **96**, 145–154.

Shelford, V.E. (1913) *Animal Communities in Temperate North America*. University of Chicago Press, Chicago, IL, USA.

Sonderegger, D.L., Wang, H., Clements, W.H. & Noon, B.R. (2009) Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment*, **7**, 190–195.

Suding, K.N. & Hobbs, R.J. (2009) Threshold models in restoration and conservation: a developing framework. *Trends in Ecology & Evolution*, **24**, 271–279.

Toms, J. & Lesperance, M.L. (2003) Piecewise regression: a tool for identifying ecological thresholds. *Ecology*, **84**, 2034–2041.

Townsend, C.R., Uhlmann, S.S. & Matthaei, C.D. (2008) Individual and combined responses of stream ecosystems to multiple stressors. *Journal of Applied Ecology*, **45**, 1810–1819.

Utz, R.M., Hildebrand, R.H. & Boward, D.M. (2009) Indentifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. *Ecological Indicators*, **9**, 556–567.

Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M. & Morgan, R.P. (2005) The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society*, **24**, 706–723.

Wang, L., Robertson, D.M. & Garrison, P.J. (2007) Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implications for nutrient criteria development. *Environmental Management*, **39**, 194–212.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Taxon-specific results from Threshold Indicator Taxa Analysis (TITAN) of macroinvertebrate community composition in response to surface-water total phosphorus (TP, ug/L) in the Everglades.

Appendix S2. Taxon-specific results from Threshold Indicator Taxa Analysis (TITAN) of macroinvertebrate community composition in response to watershed developed land cover (%) in Maryland.

Appendix S3. Program TITAN for R 2·9·2 for performing all analyses described in this paper.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.