Environmental Pollution 200 (2015) 120–125

Contents lists available at ScienceDirect

Environmental Pollution

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

How to assess species richness along single environmental gradients? Implications of potential versus realized species distributions

Thomas M.W.J. van Goethem [a, *], Mark A.J. Huijbregts [a], G.W. Wieger Wamelink [b], Aafke M. Schipper [a]

[a] Department of Environmental Science, Institute for Water and Wetland Research, Radboud University Nijmegen, P.O. Box 9010, 6500 GL, Nijmegen, The Netherlands

[b] Alterra, Wageningen University and Research Centre, P.O Box 47, 6700 AA, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form
19 January 2015
Accepted 15 February 2015
Available online 23 February 2015

Keywords:
Survey data
pH
Stressor–response relationships
Species sensitivity distributions (SSDs)
Environmental quality standards (EQS)

ABSTRACT

Quantifying relationships between species richness and single environmental factors is challenging as species richness typically depends on multiple environmental factors. Recently, various methods have been proposed to tackle this challenge. Using a dataset comprising field observations of grassland vegetation and measured pH values, we compared three methods for deriving species richness response curves. One of the methods estimates species richness close to the maximum species richness observed at the sites, whereas the other two provide estimates of the potential species richness along the environmental gradient. Our response curves suggest that potential species richness of grasslands is slightly more sensitive to acidification than realized plant species richness. However, differences in corresponding environmental quality standards (EQS) for acidification were small compared to intrinsic spatial differences in natural soil pH, indicating that natural background values are more important to consider in the derivation of EQS for pH than methodological differences between the three approaches.

1. Introduction

Environmental factors that determine species distribution patterns and species richness are of primary interest to nature conservation (Pausas and Austin, 2001). Quantifying the influence of individual factors on species communities in a systematic way can help to improve our understanding and predictive ability of biodiversity patterns, derive environmental quality standards, and underpin abatement priorities (Latour and Reiling, 1993; Van Goethem et al., 2013; Wamelink et al., 2013). However, species distributions are typically dependent on multiple environmental factors, including both abiotic and biotic drivers (Pulliam, 2000; Schipper et al., 2014; Soberón, 2007). As confounding environmental factors generally result in considerable scatter among species richness observations, it is not straightforward to extract relationships between species richness and single factors from field data (Cade and Noon, 2003; Van den Brink et al., 2002).

Recently, various methods have been proposed to tackle this challenge (Leung et al., 2005; Struijs et al., 2011; Kefford et al., 2011; Iwasaki and Ormerod, 2012; Azevedo et al., 2013; Cormier & Suter II, 2013). Most of these methods are based on occurrence data (e.g. presence–absence data), which are generally more readily available than abundance data (Pearce and Boyce, 2006; Potts and Elith, 2006). One method is to relate site-specific observations of the number of species present to a particular environmental variable with quantile regression (Iwasaki and Ormerod, 2012). Most regression techniques relate changes in the mean of a response variable to one or more explanatory variables. With quantile regression, any part of the distribution of a variable can be used as response (Cade and Noon, 2003). Quantile regression based on one of the upper boundaries of the response variable distribution (e.g. the 0.95 or 0.99 quantile) is expected to show the constraints imposed by the explanatory environmental variable of concern (Iwasaki and Ormerod, 2012; Lancaster and Belyea, 2006). A second method is to assess the number of species present within regular intervals along a particular environmental gradient by pooling multiple samples per interval ('pooled samples method'). The number of species per interval is then assessed either by simply counting the number of unique species across all samples within the interval (Struijs et al., 2011) or by establishing a species

* Corresponding author.
E-mail address: tgoethem@science.ru.nl (T.M.W.J. van Goethem).

http://dx.doi.org/10.1016/j.envpol.2015.02.017
0269-7491/

accumulation curve (SAC) per interval, thus correcting for potential differences in the number of samples between the intervals (Kefford et al., 2011). With a third method, observations of multiple species across multiple samples are used to first establish species-specific occurrence ranges, represented by the minimum and maximum values of the environmental variable of concern where the species has been observed. These occurrence ranges are then stacked across the species to arrive at an estimate of species richness ('occurrence range method'; Verbrugge et al., 2012; Azevedo et al., 2013; Cormier et al., 2013).

Given the differences in approach, these three methods are expected to yield different species richness estimates, reflecting differences in potential and realized species richness. Potential species richness refers to the species that could occur at a specific site, while realized species richness refers to the species that actually occur there (Jiménez-Valverde et al., 2008). By modelling an upper quantile of the distribution of species richness actually observed at the sampling sites, the quantile regression method yields an estimate of the maximum species richness that may be realized at a particular location with a given pH. In contrast, the other two methods yield species richness estimates representing the pool of plant species corresponding with a given pH, i.e., the potential species richness. Species richness typically increases with an increasing number of samples (Kefford et al., 2011). Hence, aggregating observations from multiple sampling sites at each given interval along a particular environmental gradient, as is done in the pooled samples method, is expected to yield considerably higher values of species richness than can be observed at specific sampling sites (Kefford et al., 2011). The occurrence range method, finally, is expected to yield the highest estimates of species richness, by aggregating the species occurrences over the full environmental gradient rather than for each given interval separately.

The goal of this paper was to compare the three methods by applying them to the same species-environment dataset and quantifying the differences in the resulting species richness response curves. The dataset comprises presence-absence observations of terrestrial plant species along a gradient of soil pH measurements (pH 3–10) collected from 4412 sampling sites of grassland vegetation across the Netherlands (Wamelink et al., 2012). The methods were compared by quantifying the shapes of the response curves (magnitude, width) along the pH gradient. Furthermore, we compared the methods in terms of environmental quality standards, i.e. the pH levels corresponding with a predefined relative reduction in species richness (Van Straalen and Denneman, 1989; Posthuma et al., 2002). To achieve this we converted the species richness estimates to relative values with a maximum of 100%, thus obtaining field-based species sensitivity distributions (f-SSDs), i.e., empirical distributions describing interspecies variation in sensitivity to a particular environmental variable.

2. Methods

2.1. Species richness response curves

2.1.1. Quantile regression

The quantile regression method to estimate species richness along the pH gradient was based on Cade and Noon (2003) In our study, three models were constructed at the 95% quantile (Visser and Sasser, 2009): a linear model ($y = \beta_0 + \beta_1 \cdot x$), a Gaussian model ($y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2$) and a baseline model where species richness is estimated by a constant (i.e., an intercept-only model). The most parsimonious model was selected based on the Bayesian Information Criterion (Lee et al., 2013). The different models were also constructed for the 97.5% and 99% quantiles to assess the influence of the quantile selection on the species richness estimates. The quantile regression was performed with the quantreg package in R (Koenker et al., 2013).

2.1.2. Pooled samples method

With the pooled samples method (Kefford et al., 2011), we derived species accumulation curves (SACs) for each interval i along the pH gradient. The SACs were derived using a resampling rarefaction method (100 times) that calculates the mean number of species observed (SR_{est}) in 1 to n samples, where n is the total number of samples pooled. The SR_{est} in k samples, SR_{est} (k), is the mean number of species estimated in k samples. The SR_{est} (inf) is the mean number of species where one added sample leads to a maximum increase of less than one species (Verberk et al., 2006). For each interval i we considered the SR_{est} (inf) as an estimate for $SR_{i,j}$ (Kefford et al., 2011). The intervals i were set at 0.1 pH unit, so that there were enough observations in each interval to derive SR_{est} (inf) (Table S2). The SACs were extrapolated up to a maximum of 5 times to ensure that SR_{est} (inf) could be estimated for all intervals (Colwell et al., 2012). As a sensitivity check the response curves were also derived based on 50, 20 and 1 samples. The SACs were determined using the computer software Estimates 7.5.1 (Colwell et al., 2004).

2.1.3. Occurrence range method

Following (Azevedo et al., 2013), we defined the occurrence range for each species as the range between minimum and maximum pH values corresponding to the occurrence of that species as observed in the field. A species was considered to be absent at pH values outside this range, and potentially present at values inside its occurrence range. Species richness (SR_i) was computed as the number of species potentially present at each pH interval i as

$$SR_i = \sum_s O_{s,i} \quad (1)$$

where $O_{s,I}$ is the occurrence of each species s at pH interval i, with $O = 0$ when the pH value is outside a species' occurrence range and $O = 1$ if the pH value is within its occurrence range. The intervals i for pH were set at 0.1. To assess the sensitivity of SR to changes in occurrence ranges, the species occurrences were also derived based on the 5th and 95th and 2.5th and 97.5th percentiles of the pH values corresponding to the field occurrence of that species.

2.2. Dataset

The ecological conditions (EC) database compiled by Wamelink et al. (2012) was used in this study. This database comprised vegetation relevés from the Netherlands, each accompanied by a measured value of at least one abiotic soil parameter. The database contained 5243 grassland relevés with a measured pH value, covering the period from 1936 to 2011 (Table 1). pH values were measured in H_2O extract and ranged from 3.0 to 10.1. Several relevés were part of a time series: the dataset included 141 sites where a relevé was made at least twice. To remove potential confounding influences of temporal autocorrelation, we included only the most recently recorded relevés from each time series in the dataset. This

Table 1
Characteristics (Mean, SD, Median, Min, Max and various percentiles) of the measured pH values and species richness for 4412 relevés.

	Mean	SD	Median	Min	0.025	0.25	0.75	0.975	Max
pH values	5.7	1.3	5.6	3.0	3.8	4.7	6.4	7.9	10.1
Species richness	25	12	24	1	5	14	31	48	73