Using species-environmental amplitudes to predict pH values from vegetation

Peppler-Lisbach, Cord

University of Oldenburg, IBE, Landscape Ecology Group, P.O. Box 2503, 26111 Oldenburg, Germany;
Fax +49 7985659; E-mail cord.peppler.lisbach@uni-oldenburg.de

Abstract

Question: Species optima or indicator values are frequently used to predict environmental variables from species composition. The present study focuses on the question whether predictions can be improved by using species environmental amplitudes instead of single values representing species optima.

Location: Semi-natural, deciduous hardwood forests of northwestern Germany.

Methods: Based on a data set of 558 relevés, species responses (presence/absence) to pH were modelled with Huisman-Olff-Fresco (HOF) regression models. Species amplitudes were derived from response curves using three different methods. To predict the pH from vegetation, a maximum amplitude overlap method was applied. For comparison, predictions resulting from several established methods, i. e. maximum likelihood/present and absent species, maximum likelihood/present species only, mean weighted averages and mean Ellenberg indicator values were calculated. The predictive success (squared Pearson's r and root mean square error of prediction) was evaluated using an independent data set of 151 relevés.

Results: Predictions based upon amplitudes defined by maximum Cohen's $\varkappa$ probability threshold yield the best results of all amplitude definitions ($R^2 = 0.75$, RMSEP = 0.52). Provided there is an even distribution of the environmental variable, amplitudes defined by predicted probability exceeding prevalence are also suitable ($R^2 = 0.76$, RMSEP = 0.55). The prediction success is comparable to maximum likelihood (present species only) and – after rescaling – to mean weighted averages. Predicted values show a good linearity to observed pH values as opposed to a curvilinear relationship of mean Ellenberg indicator values. Transformation or rescaling of the predicted values is not required.

Conclusions: Species amplitudes given by a minimum and maximum boundary for each species can be used to efficiently predict environmental variables from species composition. The predictive success is superior to mean Ellenberg indicator values and comparable to mean indicator values based on species weighted averages.

Keywords: Calibration; HOF model; Realised niche; Response curve; Woodland .

Nomenclature: Wisskirchen & Haeupler (1998); Koperski et al. (2000).

Introduction

Using species composition to predict environmental variables, 'bio-indication', is common in community ecology (Zonneveld 1983; Diekmann 2003). In a numerical sense, this is calibration *sensu* Jongman et al. (1995). Particularly widespread are predictions based on single indicator values (IV) of plant species, especially by calculating averages of Ellenberg's IVs. While indicator values have proven to work reasonably well in many cases, it has been argued that IVs suffer from the lack of information about species amplitudes (Ertsen et al. 1998; Gégout et al. 2003; Wamelink et al. 2005). The term 'amplitude' is used here in the sense of 'environmental amplitude', i.e. that part of a gradient where the species is most likely to occur. In other words, the amplitude describes the realised niche referring to the gradient under consideration.

Describing species behaviour along ecological gradients, response curves

It is widely assumed that species show a unimodal, Gaussian response to environmental gradients (Jongman et al. 1995). However, there is evidence from several studies, that response curves are often skewed or monotonous (e.g. Minchin 1989; Lawesson & Oksanen 2002; Rydgren et al. 2003). Therefore, various methods for modelling different types of species response curves are applied (Guisan & Zimmermann 2000; Austin 2007). For presence-absence data, logistic regression models (= generalized linear models, GLM, with a logit link, McCullagh & Nelder 1989) are particularly adopted. Most commonly, monotone (sigmoidal) and symmetrical bell-shaped curves are modelled by using second order polynomials (e.g. Oksanen & Minchin 2002a; Gégout et al. 2003; Coudun & Gégout 2005). Generalized additive models (GAM, Yee & Mitchell 1991) and other non-parametrical approaches (e.g. Kernel estimation: Gégout & Pierrat 1998) allow for various response shapes and are more descriptive of the data. Due to their ability to represent complex behaviours along gradients, they can also be used to find suitable parametric models (Guisan et

al. 2002). Huisman-Olff-Fresco (HOF) models (Huisman et al. 1993; Oksanen & Minchin 2002a) can be seen as an extension of parametric logistic regression modelling, allowing for a limited set of response shapes, including monotonic, plateau, symmetric and skewed response curves. HOF-models have been increasingly used in recent years (Lawesson & Oksanen 2002; Oksanen & Minchin 2002a; Rydgren et al. 2003). HOF models are applied in the present study as they provide parsimonious models, yet able to meet relevant theoretical considerations about the nature of species responses along the pH gradient. Some of the sites investigated show extreme pH values as low as pH (CaCl$_2$) 2.5. Thus, skewed responses representing a stronger limitation towards the extreme part of a gradient might occur as hypothesized by Austin (1990). Furthermore, ecophysiological evidence and the observed behaviour of some species suggest the existence of pH thresholds (triggered e.g. by Al-toxicity, Tyler & Falkengren-Grerup 1998), resulting in a plateau-like response shape. As opposed to non-parametrical models, HOF models provide the possibility to test statistically for these basic response types (e.g. symmetric vs. skewed, sigmoidal vs. plateau, Oksanen & Minchin 2002a).

Species optima and amplitudes

Apart from modelling continuous response curves along environmental gradients, species behaviour can be characterised more simply by the optimum and the environmental amplitude. Optima are derived either from unimodal response curves or, as a simpler approximation, by weighted averaging (ter Braak & Barendregt 1986; Jongman et al. 1995; Schaffers & Sýkora 2000). A similar approach to describe species behaviour along gradients is the well-known indicator value approach introduced by Ellenberg (see review by Diekmann 2003).

Established parameters measuring species amplitudes are, in case of Gaussian response curves, tolerance (t) and range ($2t$) (ter Braak & Looman 1986; Jongman et al. 1995; Heegard 2002), where tolerance equals the standard deviation of the optimum. For other unimodal response models such as Gaussian logit curves, skewed or irregular shaped curves (modeled e.g. with GAM), t is not an appropriate measure (Heegard 2002). Therefore, Heegard (2002) introduced novel border parameters, but these are still applicable only to unimodal response types with a known optimum. Amplitudes used as properties in predicting specific environmental variables require a more general definition in order to include species with monotonous responses. Gégout & Pierrat (1998) propose a definition of species amplitudes which can also be applied to monotonous response shapes. The amplitude is defined by the part of the gradient containing 80% of the distribution of probability of presence. Austin

et al. (1990) use an arbitrary threshold of probability of occurrence at $P = 0.1$ to define a "feasible qualitative environmental realized niche" (QUERN). For an 'optimal' QUERN a threshold of $= 0.7$ is adopted. In the present study, two additional amplitude definitions based on different probability thresholds are tested. All these amplitude definitions provide an estimation of the preferred interval of a gradient independently of specific assumptions about species responses.

Predicting environmental variables from vegetation

The indicator value methods use single values for each species, representing species optima. Environmental variables are predicted by calculating a weighted average of present species indicator values across samples (Gégout el al. 2003; Wamelink et al. 2005). As opposed to IV methods, maximum likelihood methods use the complete information of species response curves to predict environmental variables (Jongman et al. 1995; Gégout et al. 2003; Wamelink et al. 2005).

Here, a novel approach of predicting environmental variables from species environmental amplitudes is introduced. It is based on the assumption that the predicted value of an environmental variable must be an element of the intersection of species niches (cf. Zonneveld 1983). The niche referring to the variable under consideration is quantified by the lower and upper limits of the species amplitude. The part of the gradient where the intersection (or overlap) of amplitudes of all present species is maximised is taken as the predicted value (Fig. 1).

In particular, this study deals with the following questions: 1. Can species amplitudes improve prediction of pH in forest soils compared to established indicator value and maximum likelihood methods? 2. Which definitions of species amplitudes are appropriate? 3. Is prediction success influenced by an uneven distribution of the predictor variable along the gradient? The questions are approached by comparing prediction success based on three different amplitude definitions with those of established methods by using explained variance (R^2) and root mean square errors of prediction (RMSEP).
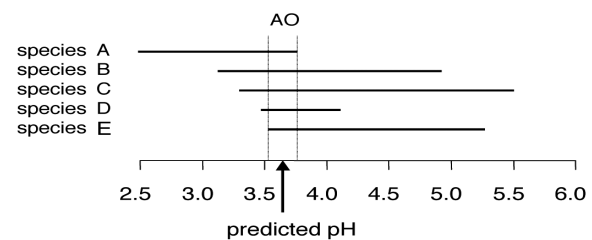
Fig. 1. pH prediction with the amplitude overlap (AO) method.

Material and Methods

Study area

The study area is located in NW-Germany within a radius of 50 km around the city of Oldenburg (53° 09' N, 8° 12' E). Sample plots were placed in stands of semi-natural deciduous forests, predominantly ancient woodlands. Dominant tree species are *Quercus robur*, *Carpinus betulus*, *Fagus sylvatica*, *Fraxinus excelsior* and *Alnus glutinosa*. Soil substrate is mainly ground moraine boulder clay (Saale period), more or less covered by sand layers, or clay ('Lauenburger Ton'). Soil types are podzols, pseudogley and gley soils, including intermediate types.

Data sets

The training data set consists of 558 phytosociological relevés, mostly unpublished data (collected 1996-1998), except 93 taken from Huntke (2002). The evaluation data set comprises 151 unpublished relevés from the same region (but different locations), collected between 2003 and 2005. Thus, the evaluation data set is completely independent of the training data set and is not a random subset. Although both data sets cover nearly the same range of pH gradient (Table 1), 14 modelled species are completely absent in the evaluation data set, due to local distribution patterns. Likewise, other species occurring in the evaluation set have not been modelled due to absence or low prevalence in the training data set.

The sampling design was preferential and aiming to cover the major gradients (dry - moist, base-poor - base-rich). Plot sizes were mostly between 100 and 200 m^2 (median 150 m^2, interquartile range 100 - 200 m^2). Species presence/absence data were used as dependent variables in the analyses.

To test for effects of uneven distribution of pH values across the training data set, a second reduced data was created (see Table 1). For this purpose, the pH gradient was partitioned in seven classes with a class width of 0.5 pH, except for class seven including all pH > 5.5. In the full data set, there is a considerable accumulation of samples at pH values < pH 4.0. In the stratified, reduced data set, for each of the pH classes < pH 4.0, the sample size was limited to 55 randomly selected plots, resulting in an overall sample size of 312 plots. Thus, the uneven distribution of pH values could be widely reduced.

Soil samples were mixed from three to five subsamples within the plot from the upper 10 cm of the mineral soil. pH values were measured in the laboratory in CaCl$_2$ (according to Schlichting et al. 1995). The range of pH-values covers the whole extent of the pH gradient in woodland sites of the region.

Response curve modelling

Response curves for 127 species of the understorey vegetation (vascular plants and epigeic bryophytes with at least ten occurrences in the training data set) were modelled with Huisman-Olff-Fresco (HOF) models (Huisman et al. 1993). Calculations were performed with R 2.2.1 (Anon. 2005), using the 'gravy' package, v. 0.0-21 by J. Oksanen (Oksanen & Minchin 2002a, b). HOF models consist of a set of five hierarchical models (types I to V) of increasing complexity, involving one to four model parameters. Constant (type I), sigmoidal (type II), sigmoidal with plateau below $P = 1$ (type III), unimodal symmetric (type IV) and unimodal skewed (type V) response curves can be modelled and tested for significance. The selection of the best model was done according to the likelihood ratio test of residual deviance (χ^2-method) with $P = 0.05$ (see Oksanen & Minchin 2002a). The function of the gravy package selecting the best model was slightly modified. Fitted type V models with identical signs of parameters b and c, resulting in a monotonic curve, were skipped in favour of the next less complex significant model. The same was done with type V and type III models showing extremely high coefficients (> 100), resulting in response shapes with very abrupt increases or decreases (absolute values of slopes converging to infinity) and peak-like maxima (cf. Zelený, unpubl. URL: http://sci.muni.cz/botany/zeleny/selfish-hof.php). Model goodness of fit is expressed by Nagelkerke's R^2 (R^2N) as an analogous measure to the R^2 of least squares estimated regression models (Nagelkerke 1991).

Table 1. Distribution of plots in relation to pH.

Data set	≤ 3	≤ 3.5	≤ 4	pH-class ≤ 4.5	≤ 5	≤ 5.5	≤ 6.1	Min.	Max.
Training data sets:									
Full data set	149	193	69	56	47	31	13	2.5	6.1
Reduced data set	55	55	55	56	47	31	13	2.5	6.1
Evaluation data set	36	30	14	18	15	26	12	2.5	5.9

Defining amplitudes

Different methods of defining species amplitudes were tested. The first two are based on probability thresholds used in predictive modelling (see Liu et al. 2005). A.prev is the pH-amplitude where the predicted probability of occurrence exceeds prevalence, i.e. *a priori* probability (= proportion of observed occurrences in the data set). A.kappa is defined by the threshold for predicted probability resulting in a maximum of Cohen's κ (Cohen 1960). For this, κ-values for an array of thresholds with 0.01 pH resolution within the observed pH range were calculated and the threshold yielding the maximum κ-value was taken as the classification cut value. The derived amplitude can be interpreted as the part of the gradient with a maximum prediction success compared to a random distribution for the given prevalence of the species. A.area defines the pH-amplitude comprising 80 % of the area under the response curve (Gégout & Pierrat 1998; Coudun & Gégout 2005). All amplitude calculations have an accuracy of 0.01 pH units.

Prediction methods

Predictions derived from species amplitudes are based on the maximum amplitude overlap method. For a given species composition, amplitudes of present species are used to sum up all species predicted to be present for a certain pH value. The pH with a maximum of predicted species is the predicted pH (Fig. 1). The calculations were performed for a pH scale with a step width of 0.05 pH units. In cases with more than one pH-value yielding the maximum species number, the mean of all values reaching the maximum was taken as the predicted pH. The predictions were calculated for the three different amplitude definitions (AO.prev, AO.kappa, AO.area method). For comparison, various established calibration procedures were performed. Maximum likelihood predictions (ML method; Jongman et al. 1995; Gégout et al. 2003; Wamelink et al. 2005) based upon present and absent species require fitted response curves for all the species. A probability (likelihood) of observing a certain pH-value for a given combination of present and absent species can be calculated by:

$$L_j(pH) = \prod_{i=1}^{k} p_i(pH)^{y_{ij}} \cdot \prod_{i=1}^{k} [1 - p_i(pH)]^{1-y_{ij}} \quad (1)$$

The pH that maximises the likelihood is the predicted pH-value (ML-p/a method, 'full' method *sensu* Wamelink et al. 2005). The likelihood based on present species only is calculated accordingly (ML-p method, 'present' method *sensu* Wamelink et al. 2005):

$$L_j(pH) = \prod_{i=1}^{k} p_i(pH)^{y_{ij}} \quad (1)$$

The maximum likelihood estimation was performed with the nlm function in R 2.2.1.

Prediction by weighted averaging was carried out by first calculating indicator values of each species, using presence/absence data according to Jongman et al. (1995). The predicted pH value is the mean indicator value calculated from the indicator values of the species present (IV-wa method). Due to double averaging, IV-wa predictions lead to a compression of the predicted pH values. To account for this, pH values were rescaled using linear regression as described by ter Braak & van Dam (1989) and Wierda et al. (1997).

Additionally, mean indicator values were also calculated using the R (soil reaction) indicator value of Ellenberg et al. (1992) (IV-e method). Observed pH values and mean soil reaction values often show a curvilinear relationship (Ertsen et al. 1998; Diekmann 2003). To transform mean *R* value (*mR*) to pH scale, non-linear regression was applied according to Ertsen et al. (1998).

Predictions are evaluated using two quality measures, the squared linear correlation coefficient (R^2), and the root mean square error of prediction (RMSEP). The RMSEP is the square root of the mean of all squared differences between observed and predicted pH (see Gégout at al. 2003; Wamelink et al. 2005). The R^2 expresses a linear relationship between predicted and observed values independently of the scale of predicted values. For the sake of comparability, RMSEP values require identical scales. Thus, prediction results based on Ellenberg's ordinal indicator values cannot be compared to the others using RMSEP, unless transformed to the pH scale by regression (see above).

Results

The prediction success of the models derived from the stratified data set did not generally improve compared to the full data set with uneven sampling (see below, Table 2). Thus, the results of the reduced data set are not shown in detail. If not stated otherwise, the results refer to the full data set.

Response curves and species amplitudes

The 127 modelled species show all response shapes HOF models allow for (Fig. 2, App. 1). Eight species are indifferent (type I, Fig. 2a); all other species are significantly related to pH. 79 models (62%) have sigmoidal (type II, cf. Fig. 2b,c and type III, cf. 2d) curves, 59 with an increasing and 20 with a decreasing response to increasing pH. All Type III models (48) have plateaus at high pH values. 40 species (31 %) have unimodal response curves, with 31 species (24%) displaying symmetric (type IV, cf. Fig. 2e) and nine species (7%) showing skewed (type V, cf. Fig. 2f) responses. All species

with a skewed response show a steeper decrease towards more acid soil conditions (cf. Fig 2f).

Model goodness of fit (R^2N) of type II to type V models range from below 0.1 to 0.72, with model R^2N of 100 species exceeding 0.1 and 45 species exceeding 0.3. Type III models with mean R^2N of 0.36 yield a significantly (Behrens-Fisher-Test, $P < 0.01$) higher proportion of explained variance than model types II, IV and V with mean R^2N of 0.19, 0.18 and 0.18, respectively. Model parameters, variance explained (R^2N) and species amplitudes are given in App. 1.

The width of species amplitudes depends on amplitude definition, prevalence and position on the gradient (App. 2). A.kappa yields amplitude widths with the greatest variance and results in very small amplitude widths for a number of species (mean = 1.48, SD = 0.84, models type I omitted, cf. App. 2). As opposed to that, A.prev shows a high mean amplitude width (mean = 1.85, SD = 0.71). A.area results in intermediate mean amplitude widths with a relatively small variance (mean = 1.67, SD = 0.58). Contrary to this general pattern, type III models yield narrower A.area widths than A.kappa and A.prev (cf. Fig. 2d), due to steep slopes occurring in the plateau response curves. There is a clear tendency for amplitudes to increase with prevalence, especially for A.area (Spearman's $\varrho = 0.71$, $P < 0.001$). This effect is weaker concerning A.kappa (0.56, $P < 0.001$) and A.prev (0.29, $P = 0.001$). The deviances between amplitude widths of different definitions are more pronounced in species with high prevalences. Position on the gradient influences amplitude widths in so far as species with positions near the lower end of the gradient have much narrower amplitudes than those with intermediate or high gradient positions. This applies to all amplitude definitions, but especially to A.prev and, to a lesser extent, to A.kappa. Species with high positions on the gradient do not generally yield narrower amplitudes (with the exception of A.area), mainly due to the high proportion of type III models showing wide A.kappa and A.prev amplitudes.

Evaluation of pH predictions

Prediction methods differ in performance (Table 2), ranging from 0.53 to 0.76 in explained variance and from 0.51 to 1.05 in mean prediction error. Referring to the R^2-values, ML-p, IV-wa and AO.kappa (full data set) and AO.prev, ML-p, IV-wa and AO.kappa (reduced data set) yield the best results for the evaluation data set. ML-p/a works well with the training data set , but performs considerably weaker when applied to the evaluation data set. IV-e, AO.area predictions have intermediate performances. Referring to RMSEP values, the situation is partly different. Predictions with the smallest mean error are yielded by ML-p and AO.kappa with RMSEP of about 0.5 pH (full data set). Smallest mean errors derived from the reduced data set show ML-p, AO.kappa and AO.prev. All other methods result in predictions with mean errors exceeding 0.6 or even 0.7. Some insights into the reasons for different performances can be derived considering Appendix 3. As indicated by the R^2-values, ML-p, IV-wa, AO.kappa and AO.prev (reduced data set only) display the highest linearity between predicted and observed pH. Indicator value (IV) predictions show a distinct compression with predicted values ranging only from ca. 3.0 to 4.5, which is reflected in the weaker RMSEP value. RMSEP of IV-wa improves considerably after rescaling (0.57 vs. 0.75). A certain compression of the y axis can also be recognised regarding AO.prev. Methods showing an evident deviation from linearity are especially IV-e with a curvilinear relationship, but also ML-p/a, AO.area and AO.prev (full data set only).

Performance of predictions derived from the reduced data set is weaker than that of predictions derived from the full data set, except for AO.prev. AO.prev predictions are significantly enhanced by a more balanced sampling structure, but a more even distribution of pH values does not generally lead to better predictions. Considering both quality measures, ML-p, IV-wa with rescaling, AO.kappa and AO.prev (reduced data set) show the best results.
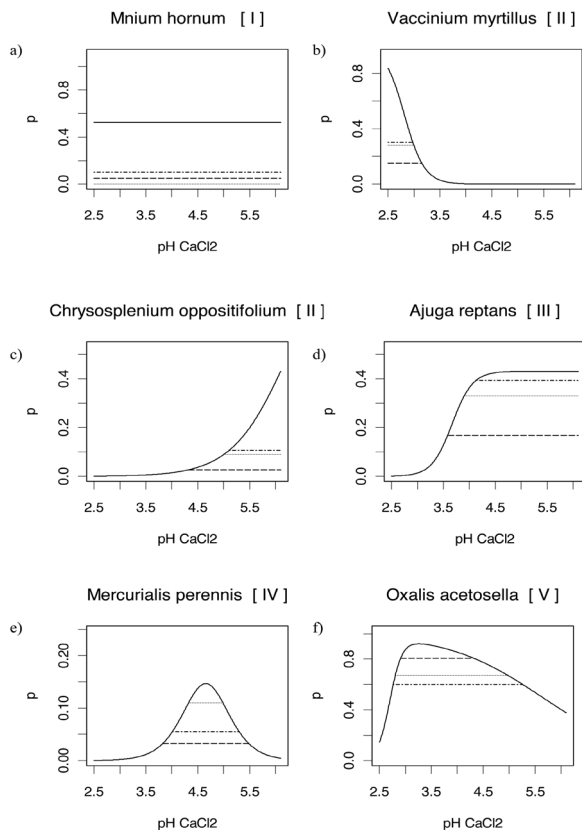
Fig. 2. Response curves with amplitudes of selected species. HOF model type in brackets.

––– : A.prev   ·········· : A.kappa   ·–·– : A.area.

Table 2. Evaluation of pH predictions for different prediction methods. All correlations are significant at the $P = 0.001$ level.

| Model | Models reduced data set | | | | Models full data set | | | |
| Data set | training | | evaluation | | training | | evaluation | |
Prediction method	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
AO.prev	0.70	0.52	0.76	0.55	0.70	0.50	0.70	0.67
AO.kappa	0.72	0.46	0.72	0.54	0.73	0.43	0.75	0.52
AO.area	0.59	0.63	0.64	0.65	0.57	0.67	0.66	0.63
ML-p/a	0.71	0.50	0.58	0.73	0.76	0.42	0.64	0.70
ML-p	0.75	0.47	0.73	0.54	0.72	0.45	0.75	0.51
IV-wa	0.70	0.66	0.73	0.75	0.74	0.57	0.76	0.75
IV-wa (rescaled)	0.73	0.71	0.70	0.57	0.74	0.48	0.76	0.57
IV-e	0.64	-	0.67	-	0.60	-	0.68	-
IV-e (transformed)	0.69	0.49	0.66	0.60	0.70	0.44	0.57	0.69

Discussion

Response curves and gradient properties

The results confirm the variability in response shapes stated by previous studies (e.g. Oksanen & Minchin 2002a; Rydgren et al. 2003). The relatively low proportion of unimodal responses (31%) can be connected to gradient length (Lawesson & Oksanen 2002; Rydgren et al. 2003). While the lower end of the pH gradient in the present study (pH 2.5) approximately coincides with the lowest possible pH in woodland sites of central Europe, the upper end is truncated by the almost complete absence of calcareous soils. Thus, sites with pH values exceeding 6.0 hardly occur. Presumably for the same reason, skewed responses are rare in the data set, as opposed to other observations (Minchin 1989; Austin et al. 1994; Lawesson & Oksanen 2002). The fact that all skewed response curves are right-skewed supports the hypothesis of Austin (1990) that species display stronger limitations towards the extremes than towards the central portion of a gradient. Plateau responses are frequent (48) with a significantly better goodness of fit than all other model types. Therefore, plateau response models can be regarded as particularly suitable for many species to describe their occurrence along the pH gradient in the study area.

Prediction success

The different prediction methods explain between 0.53 and 0.76 percent variance soil pH with mean errors mainly between 0.51 and 1. Comparable studies from woodlands with data-based prediction methods report R^2-values ranging from 0.15 to 0.47 and RMSEP from 0.47 to 0.9 (Gégout et al. 2003; Gégout & Krizova 2003). Wamelink et al. (2005) present results from different prediction methods derived at a broader scale. They report RMSEP values ranging from 0.6 to 1.3, depending on prediction method and region of evaluation data. Numerous studies contain correlations between mean Ellenberg's R-value and pH measurements (see

Diekmann 2003). In the present study, predictions from mean R-values (IV-e method) perform rather weakly. The results from the evaluation data set show that even using a non-linear regression to meet the weakness of a curvilinear relation between IV-e predictions and observed values (see Schaffers & Sýkora 2003; Diekmann 2003) is not a successful approach. Another prediction method showing a considerable decline in prediction success when applied to the evaluation data set is the ML-p/a method, apparently due to some differences in species composition between training and evaluation data set. Lack or low frequencies of modelled species in the evaluation data set leads to a shift towards lower predicted pH values, as plots with lower species richness tend to have lower pH values. The ML-p/a method will only yield good predictions, if the absence of species in the evaluation data can be really linked to pH. If species are missing for any other reason (dispersal, range, other vegetation types), the results can be misleading. Based on these considerations, the good prediction success of ML-p/a reported by Gégout et al. (2003) can be linked to the circumstance that the evaluation data set used was a random subsample of the whole data set. Wamelink et al. (2005) report a much weaker general performance of the ML-p/a method. Comparing the prediction success reported by Wamelink et al. (2005) to those yielded by the present study or by Gégout et al. (2003) and Gégout & Krizova (2003), respectively, reveals the tendency of better results from small study areas than from large ones. Likewise, training data sets comprising a wide range of vegetation types may result in weaker prediction success than those confined to certain types, e.g. woodlands.

Amplitude overlap methods are among the most effective prediction methods in the present study with the exception of AO.area (discussion see below). Nonetheless, a general weakness (also applying to IV methods) can be seen in that fixed univariate amplitudes do not account for possible interactions with other environmental factors (Pakeman et al. 2008). Otherwise, e.g. species amplitudes for pH should vary with water regime. Yet, one has to take into account that amplitudes and IVs are only simple approximations of species behaviour along a gradient and

their advantage lies in this simplicity (straightforward, easy-to-use). More reliable predictions could possibly be achieved by maximum likelihood methods based on more complex multiple regression models accounting for different factors and their interactions (Jongman et al. 1995). This would also improve the general validity of the models beyond a regional scale. Predictions based on univariate models perform reasonably well within a certain region because they implicitly reflect interactions with other factors as characteristic landscape properties (cf. Austin 2005). Their performance, though, strongly depends on the existence of similar interaction patterns.

Species amplitudes

In the present study, amplitude definitions are tested on their ability to predict the pH value of a site with a given species combination. Thus, possible amplitude definitions were selected that maximise the prediction success based on modelled response curves. This is admittedly a rather heuristic approach, as theoretical considerations about species responses take a back seat. The suitability of an amplitude definition for calibration purposes requires a predictive performance as high as possible. Defined too broad, amplitudes might include parts of the gradient where the absence of a species is in fact more likely than its presence. This problem particularly applies to A.area and A.prev (full data set). A.area definitions mostly lead to wider amplitudes compared to A.kappa, especially concerning species with high prevalence and unimodal response curves. As opposed to that, species with type III models almost completely show narrower A.area amplitudes. Both circumstances lead to a shift in amplitude overlap towards higher pH values between pH 2.7 and 4.5, resulting in a curvilinear relationship between predicted and observed values and weaker predictions (App. 3g). AO.prev (full data set) suffers from a similar distorting effect caused by the prevalence threshold applied. Because of oversampling at the lower part of the gradient, prevalence of species preferring high pH values is considerably reduced, leading to wide A.prev amplitudes. This means that the amplitudes of species with pronounced positions at the upper part of the gradient are extended towards lower pH values. Thus, A.prev amplitude overlap predictions shift towards lower pH values particularly in the middle and upper part of the gradient, resulting in a poor performance (App. 3e). This negative effect is more or less removed in the reduced data set, leading to better AO.prev predictions. In general, however, a more balanced sampling cannot fully compensate for the loss of information by omitting 45 % of the plots.

While A.area mainly depends on the shape of the response curve, the probabilistic approaches of A.kappa and A.prev additionally account for species *a priori* probability and error structure respectively (A.kappa), leading to a wide variation in amplitude widths. Thus, A.kappa and A.prev (even sampling provided) seem to be more effective in delineating that part of the gradient where the predictive significance of a species occurrence is maximised.

Application

With given species amplitudes derived from response curve modelling (App. 1), it takes nothing but a simple spreadsheet software to calculate pH predictions for given species assemblages. In this, the method is nearly as easy-to-use as indicator value methods. An Excel-spreadsheet for pH prediction based on species amplitudes modelled in this study can be obtained from the author. Yet, the predictions are restricted to forest sites in comparable landscapes with a comparable species pool.

References

Anon. (R Development Core Team) 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, AT. URL http://www.R-project.org.

Austin, M.P. 1990. Community theory and competition in vegetation. In: Grace, J.G. & Tilman, D. (eds.) *Perspectives on plant competition*, pp. 215-238. Academic Press, San Diego, CA, US.

Austin, M.P. 2005. Vegetation and environment: discontinuities and continuities. In: van der Maarel, E. (ed.) *Vegetation ecology*, pp. 52-84. Blackwell Publishing, Malden, MA, US.

Austin, M.P. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200: 1-19.

Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60: 161-178.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a β-function. *Journal of Vegetation Science* 5: 215-228.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 30: 37-46.

Coudun, C. & Gégout, J.-C. 2005. Ecological behaviour of herbaceous forest species along a pH gradient: a comparison

between oceanic and semicontinental regions in northern France. *Global Ecology and Biogeography* 14: 263-270.

Diekmann, M. 2003. Species indicator values as an important tool in applied plant ecology – a review. *Basic and Applied Ecology* 4: 493-506.

Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulißen, D. 1992. Zeigerwerte von Pflanzen in Mitteleuropa 2nd. ed. *Scripta Geobotanica* 18: 1-258.

Ertsen, A.C.D., Alkemade, J.R.M. & Wassen, M.J. 1998. Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability and salinity in the Netherlands. *Plant Ecology* 135: 113-124.

Gégout, J.-C. & Krizova, E. 2003. Comparison of indicator values of forest understory plant species in Western Carpathians (Slovakia) and Vosges Mountains (France). *Forest Ecology and Management* 182: 1-11.

Gégout, J.-C. & Pierrat, J.-C. 1998. L'autécologie des espèces végétales: une approche par régression non paramétrique. *Écologie* 29: 473-482.

Gégout, J.-C., Hervé, J.-C., Houllier, F. & Pierrat, J.-C. 2003. Prediction of forest soil nutrient status using vegetation. *Journal of Vegetation Science* 14: 55-62.

Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Guisan, A., Edwards, T.C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157: 107-122.

Heegaard, E. 2002. The outer border and central border for species-environmental relationships estimated by non-parametric generalised additive models. *Ecological Modelling* 157: 131-139.

Huisman, J., Olff, H. & Fresco, L.F.M. 1993. A hierarchical set of models for species response analysis. *Journal of Vegetation Science* 4: 37-46.

Huntke, T. 2002. Rezente Laubwaldgesellschaften und potentielle natürliche Vegetation des Mansholter Holzes, Landkreis Ammerland. *Drosera* 2002: 1-31.

Jongman, R.H.G., ter Braak, C.J.F. & van Tongeren, O.F.R. 1995. *Data analysis in community and landscape rcology.* Cambridge University Press, Cambridge, UK.

Koperski, M., Sauer, M., Braun, W. & Gradstein, S.R. 2000. Referenzliste der Moose Deutschlands. *Schriftenreihe für Vegetationskunde* 34: 1-519.

Lawesson, J.E. & Oksanen, J. 2002. Niche characteristics of Danish woody species as derived from coenoclines. *Journal of Vegetation Science* 13: 279-290.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.

McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models.* 2nd. ed. Chapman & Hall, London, UK.

Minchin, P.R. 1989. Montane vegetation of the Mt. Field massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio* 83: 97-110.

Nagelkerke, N.J.D. 1991. A note on general definition of the coefficient of determination. *Biometrika* 78: 691-692.

Oksanen, J. & Minchin, P.R. 2002 a. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157: 119-129.

Oksanen, J. & Minchin, P.R. 2002 b. *Non-linear maximum likelihood estimation of Beta and HOF response models.* URL: http://cc.oulu.fi/~jarioksa/softhelp/hof3.pdf

Pakeman, R.J., Reid, C.L., Lennon, J.J. & Kent, M. 2008. Possible interactions between environmental factors in determining species optima. *Journal of Vegetation Science* 19: 201-208.

Rydgren, K., Økland, R.H. & Økland, T. 2003. Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. *Journal of Vegetation Science* 14: 869-880.

Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. *Journal of Vegetation Science* 11: 225-244.

Schlichting, E., Blume, H.-P. & Stahr, K. 1995. *Bodenkundliches Praktikum.* 2nd. ed. Blackwell Wissenschaft, Berlin, DE.

ter Braak, C.J.F. & Barendregt, L.G. 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences* 78: 57-72.

ter Braak, C.J.F. & Looman, C.W.N. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65: 3-11.

ter Braak, C.J.F. & van Dam, H. 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* 178: 209-223.

Tyler, G. & Falkengren-Grerup, U. 1998. Soil chemistry and plant performance – ecological considerations. *Progress in Botany* 59: 634-658.

Wamelink, G.W.W., Goedhart, P.W, van Dobben, H.F & Berendse, F. 2005. Plant species as predictors of soil pH: Preplacing expert judgement with measurements. *Journal of Vegetation Science* 16: 461-470.

Wierda, A., Fresco, L.F.M., Grootjans, A.P. & van Diggelen, R. 1997. Numerical assessment of plant species as indicators of the groundwater regime. *Journal of Vegetation Science* 8: 707-716.

Wisskirchen, R. & Haeupler, H. 1998. *Standardliste der Farn-und Blütenpflanzen Deutschlands.* Ulmer, Stuttgart, DE.

Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587-602.

Zonneveld, I.S. 1983. Principles of bio-indication. *Environmental Monitoring and Assessment* 3: 207-217.

*For App. 1-3, see below (online version)
also available at JVS/AVS Electronic Archives;
www.opuluspress.se/*

App. 1. HOF model parameters of species' response curves and boundaries of amplitudes along the pH gradient.

Species	HOF model type	HOF model parameters*					A.kappa		A.prev		A.area	
		a	b	c	d	R²	lower	upper	lower	upper	lower	upper
Acer platanoides	I	2.7669	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Acer pseudoplatanus	II	0.5772	-1.3292	-	-	0.03	3.42	6.10	3.64	6.10	3.43	6.10
Adoxa moschatellina	III	5.9726	-19.1640	-1.0579	-	0.55	3.61	6.10	3.56	6.10	4.10	6.10
Aegopodium podagraria	IV	-6.4899	9.4283	6.3059	-	0.28	3.95	5.93	3.92	5.97	4.18	5.71
Agrostis capillaris	II	1.9992	8.5956	-	-	0.11	2.50	2.74	2.50	3.22	2.50	3.21
Agrostis stolonifera	II	4.3568	-2.7484	-	-	0.06	4.35	6.10	3.90	6.10	4.20	6.10
Ajuga reptans	III	5.9576	-18.3289	0.2752	-	0.36	3.91	6.10	3.59	6.10	4.14	6.10
Anemone nemorosa	III	3.7195	-29.5875	-3.3527	-	0.48	3.01	6.10	3.13	6.10	3.57	6.10
Angelica sylvestris	V	-2.4278	8.5135	26.4243	61.2577	0.31	4.07	4.36	3.91	5.14	3.96	4.82
Athyrium filix-femina	V	-3.2118	3.5806	3.0349	13.4516	0.29	3.45	5.48	3.32	5.88	3.38	5.67
Atrichum undulatum	III	4.0825	-19.7064	0.8918	-	0.14	3.43	6.10	3.32	6.10	3.81	6.10
Betula pubescens	II	1.9707	10.0449	-	-	0.13	2.50	2.84	2.50	3.19	2.50	3.10
Brachythecium rutabulum	IV	-3.2046	6.6291	2.6350	-	0.06	3.34	4.83	3.27	4.91	3.05	5.12
Brachypodium sylvaticum	III	6.2083	-16.7916	0.0264	-	0.42	3.91	6.10	3.70	6.10	4.25	6.10
Calamagrostis canescens	I	2.9407	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Cardamine amara	II	7.0005	-6.3603	-	-	0.27	5.16	6.10	4.31	6.10	5.09	6.10
Cardamine flexuosa	III	6.2341	-21.3336	0.5361	-	0.31	4.18	6.10	3.51	6.10	4.05	6.10
Cardamine pratensis	III	8.2030	-19.2312	0.3931	-	0.43	4.32	6.10	3.86	6.10	4.42	6.10
Carex elongata	II	6.2443	-4.8789	-	-	0.17	5.64	6.10	4.16	6.10	4.86	6.10
Carex pilulifera	IV	-2.9867	23.1638	3.1600	-	0.24	2.81	3.15	2.67	3.29	2.66	3.29
Carex remota	IV	-5.7521	5.4399	2.9699	-	0.24	4.03	6.10	3.73	6.10	4.14	6.10
Carex sylvatica	III	6.8571	-18.9926	-0.7004	-	0.55	3.92	6.10	3.68	6.10	4.24	6.10
Carpinus betulus	V	-0.1264	1.3538	5.3860	60.0566	0.07	2.92	3.80	2.90	4.09	2.84	5.29
Ceratocapnos claviculata	IV	-2.7804	19.0802	2.5186	-	0.25	2.70	3.29	2.65	3.35	2.63	3.37
Chaerophyllum temulum	III	6.0492	-15.5235	2.7347	-	0.15	4.26	6.10	3.74	6.10	4.28	6.10
Chrysosplenium alternifolium	III	10.7341	-28.9432	0.5429	-	0.44	4.21	6.10	3.75	6.10	4.28	6.10
Chrysosplenium oppositifolium	II	6.9673	-6.6935	-	-	0.30	5.01	6.10	4.32	6.10	5.11	6.10
Circaea lutetiana	III	3.8116	-13.0311	-1.4218	-	0.46	3.41	6.10	3.51	6.10	4.02	6.10
Cirsium palustre	III	7.3596	-17.3202	1.6297	-	0.26	4.02	6.10	3.84	6.10	4.41	6.10
Convallaria majalis	IV	-2.3213	6.6612	2.6174	-	0.04	3.38	4.29	3.12	4.55	2.84	4.83
Corylus avellana	IV	-2.3321	3.7342	1.2643	-	0.02	3.44	5.03	3.27	5.19	2.93	5.54
Crataegus laevigata	III	4.1597	-17.0404	0.3579	-	0.22	3.85	6.10	3.40	6.10	3.91	6.10
Crepis paludosa	III	9.3068	-21.0239	0.2364	-	0.48	4.11	6.10	3.92	6.10	4.47	6.10
Dactylis glomerata	II	5.6190	-4.7998	-	-	0.18	4.98	6.10	4.12	6.10	4.81	6.10
Deschampsia cespitosa	III	1.7975	-12.8512	-2.4960	-	0.30	3.21	6.10	3.28	6.10	3.63	6.10
Deschampsia flexuosa	II	-0.5462	18.7721	-	-	0.33	2.50	2.89	2.50	3.10	2.50	2.89
Dicranella heteromalla	IV	-1.6743	10.3201	1.8609	-	0.13	2.92	3.31	2.62	3.61	2.53	3.70
Dicranum scoparium	II	0.5962	7.6359	-	-	0.17	2.50	3.07	2.50	3.28	2.50	3.34
Dryopteris carthusiana	V	-0.4502	4.1752	6.2160	74.0045	0.16	2.84	3.49	2.82	3.65	2.75	4.41
Dryopteris dilatata	II	-0.8175	5.1050	-	-	0.22	2.50	3.18	2.50	3.48	2.50	3.97
Epilobium angustifolium	IV	-3.9632	32.4952	5.8049	-	0.17	2.91	3.17	2.79	3.29	2.79	3.29
Epilobium montanum	IV	-5.3629	10.9898	7.5162	-	0.14	4.08	5.13	3.71	5.51	3.86	5.36
Epipactis helleborine	IV	-8.8536	12.5757	10.0531	-	0.30	5.08	5.33	4.25	6.10	4.61	5.80
Equisetum arvense	IV	-9.1064	15.1279	9.6684	-	0.31	4.60	4.87	4.00	5.47	4.21	5.26
Equisetum hyemale	IV	-3.6919	8.4608	5.7761	-	0.08	3.98	5.05	3.55	5.48	3.59	5.44
Euonymus europaea	III	5.1367	-11.8659	-0.5286	-	0.44	4.11	6.10	3.80	6.10	4.38	6.10
Eurhynchium hians	IV	-8.3645	14.2414	9.4876	-	0.27	4.70	4.81	3.98	5.53	4.19	5.32
Eurhynchium praelongum	III	5.5332	-16.2195	0.5888	-	0.30	4.02	6.10	3.62	6.10	4.18	6.10
Eurhynchium striatum	II	4.8177	-5.5469	-	-	0.29	4.92	6.10	4.08	6.10	4.75	6.10
Fagus sylvatica	II	-1.5746	2.5360	-	-	0.10	2.50	3.59	2.50	3.69	2.50	4.98
Festuca gigantea	III	6.4383	-16.0192	0.1934	-	0.41	4.23	6.10	3.77	6.10	4.34	6.10
Filipendula ulmaria	III	9.4698	-21.6232	-0.1687	-	0.54	4.21	6.10	3.91	6.10	4.46	6.10
Frangula alnus	II	0.4874	9.7663	-	-	0.19	2.50	2.78	2.50	3.22	2.50	3.17
Fraxinus excelsior	II	1.7695	-3.7741	-	-	0.19	3.81	6.10	3.68	6.10	3.89	6.10
Gagea spathacea	III	10.1809	-37.1570	1.3869	-	0.25	3.63	6.10	3.46	6.10	4.00	6.10
Galeopsis tetrahit ag.	V	4.1304	-35.2231	-1.7729	-3.1849	0.18	2.99	4.45	3.00	4.35	2.90	5.13
Galium aparine	III	1.7672	-11.6106	-0.3147	-	0.11	3.33	6.10	3.32	6.10	3.65	6.10
Galium odoratum	III	4.5949	-15.5321	-0.5649	-	0.39	3.62	6.10	3.52	6.10	4.05	6.10
Galium palustre	III	7.9502	-20.4569	1.3915	-	0.29	4.15	6.10	3.77	6.10	4.32	6.10
Geranium robertianum	III	4.1294	-12.3882	-0.5312	-	0.37	3.62	6.10	3.59	6.10	4.12	6.10
Geum urbanum	III	5.4141	-14.4525	-1.2209	-	0.55	3.81	6.10	3.69	6.10	4.25	6.10
Glechoma hederacea	III	11.3725	-39.9444	0.9990	-	0.32	3.51	6.10	3.50	6.10	4.03	6.10
Hedera helix	V	-2.3736	3.1130	0.6515	30.8932	0.11	2.75	4.09	2.78	3.94	2.61	5.13
Holcus lanatus	I	3.6599	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Holcus mollis	IV	-2.0094	6.3580	3.0085	-	0.02	3.32	4.52	3.16	4.68	2.87	4.97

App. 1, cont.

| Species	HOF model type	HOF model parameters*					A.kappa		A.prev		A.area	
		a	b	c	d	R^2	lower	upper	lower	upper	lower	upper
Hypnum cupressiforme	II	-0.0686	7.0342	-	-	0.21	2.50	3.18	2.50	3.33	2.50	3.48
Ilex aquifolium	II	-1.4293	4.8812	-	-	0.25	2.50	3.37	2.50	3.57	2.50	4.19
Impatiens noli-tangere	III	2.9917	-17.2687	0.2026	-	0.13	3.32	6.10	3.28	6.10	3.71	6.10
Impatiens parviflora	V	6.5838	-32.8827	0.0807	-5.3737	0.10	3.32	3.57	3.11	4.17	3.04	4.45
Iris pseudacorus	IV	-6.8487	9.9881	8.2630	-	0.24	5.00	5.45	4.11	6.10	4.52	5.92
Juncus effusus	I	2.0614	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Lamium galeobdolon ag.	III	3.6704	-18.6167	-1.5012	-	0.39	3.31	6.10	3.31	6.10	3.78	6.10
Leucobryum glaucum	II	-0.3034	17.0372	-	-	0.30	2.50	3.08	2.50	3.11	2.50	2.92
Listera ovata	IV	-6.5839	11.6275	7.6936	-	0.23	3.95	5.47	3.85	5.57	4.03	5.39
Lonicera periclymenum	II	-0.8084	2.4147	-	-	0.08	2.50	3.88	2.50	3.61	2.50	4.85
Lophocolea bidentata	I	3.1931	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Luzula pilosa	II	0.4030	2.7921	-	-	0.07	2.50	3.19	2.50	3.50	2.50	4.48
Lysimachia nemorum	III	6.1136	-15.3636	0.9235	-	0.30	4.33	6.10	3.76	6.10	4.32	6.10
Lysimachia vulgaris	IV	-4.6391	9.2001	6.7967	-	0.11	4.11	5.36	3.71	5.76	3.88	5.60
Maianthemum bifolium	V	-0.8625	6.1255	1.5680	33.8196	0.21	2.72	3.29	2.65	3.51	2.55	3.92
Melica uniflora	III	3.3717	-16.8341	0.6115	-	0.13	3.31	6.10	3.32	6.10	3.78	6.10
Mercurialis perennis	IV	-6.8525	12.2414	7.8031	-	0.23	4.33	4.98	3.83	5.48	4.01	5.30
Milium effusum	III	2.0191	-29.4467	-2.4500	-	0.20	2.81	6.10	3.05	6.10	3.41	6.10
Mnium hornum	I	-0.0933	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Moehringia trinervia	IV	-3.0242	7.6392	3.1003	-	0.06	3.12	4.77	3.20	4.68	3.01	4.88
Molinia caerulea	II	-1.0752	22.1427	-	-	0.38	2.50	2.89	2.50	3.07	2.50	2.86
Oxalis acetosella	V	-3.3962	3.8954	1.7241	30.3263	0.18	2.81	4.98	2.91	4.31	2.77	5.27
Paris quadrifolia	III	8.0125	-15.7151	1.2818	-	0.33	4.52	6.10	4.04	6.10	4.63	6.10
Phalaris arundinacea	IV	-8.5797	13.1451	8.8417	-	0.32	4.41	5.36	4.05	5.72	4.29	5.48
Phyteuma nigrum	III	10.8281	-28.0498	1.2123	-	0.35	4.03	6.10	3.79	6.10	4.32	6.10
Picea abies	II	0.3677	12.8057	-	-	0.23	2.50	2.88	2.50	3.16	2.50	3.02
Platanthera chlorantha	III	7.9484	-20.3278	0.2129	-	0.45	4.22	6.10	3.77	6.10	4.33	6.10
Plagiomnium affine	III	15.6661	-54.7609	1.7296	-	0.27	3.89	6.10	3.51	6.10	3.86	6.10
Plagiomnium undulatum	III	7.8876	-22.0898	-0.3068	-	0.52	3.82	6.10	3.69	6.10	4.24	6.10
Poa nemoralis	I	4.0037	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Poa trivialis	III	4.8959	-13.3064	-0.6158	-	0.44	3.81	6.10	3.67	6.10	4.23	6.10
Polygonatum multiflorum	III	3.1717	-30.7037	0.5824	-	0.05	3.13	6.10	3.09	6.10	3.48	6.10
Polytrichum formosum	II	-0.5455	4.6123	-	-	0.18	2.50	3.37	2.50	3.47	2.50	4.05
Primula elatior	III	10.2346	-25.9685	-0.7808	-	0.64	3.81	6.10	3.81	6.10	4.35	6.10
Prunus avium	I	3.5227	-	-	-	0.00	2.50	6.10	2.50	6.10	2.50	6.10
Pteridium aquilinum	II	0.1370	15.4817	-	-	0.26	2.50	2.97	2.50	3.12	2.50	2.94
Pulmonaria obscura	IV	-7.4834	12.7941	6.9426	-	0.30	4.08	4.98	3.75	5.31	3.91	5.15
Quercus robur	IV	-3.2318	5.2028	-0.3764	-	0.10	2.61	4.37	2.75	4.22	2.50	4.76
Ranunculus auricomus ag.	III	11.8737	-33.0122	-0.0041	-	0.53	3.81	6.10	3.73	6.10	4.24	6.10
Ranunculus ficaria	III	6.0274	-17.7137	-2.6981	-	0.72	3.61	6.10	3.62	6.10	4.18	6.10
Ranunculus repens	III	7.2820	-16.6325	0.1928	-	0.44	4.32	6.10	3.87	6.10	4.44	6.10
Rubus fruticosus ag.	II	-0.6811	3.2470	-	-	0.13	2.50	3.79	2.50	3.56	2.50	4.52
Rubus idaeus	IV	-1.9514	4.6897	-0.2721	-	0.10	2.52	3.77	2.50	3.91	2.50	4.55
Rumex sanguineus	IV	-7.2572	10.1754	7.0459	-	0.31	4.69	5.37	4.01	6.05	4.32	5.74
Sambucus nigra	IV	-3.3936	16.2347	5.3396	-	0.31	3.07	3.87	3.01	3.92	2.96	3.97
Sanicula europaea	III	7.5529	-20.2471	0.9911	-	0.32	3.91	6.10	3.72	6.10	4.27	6.10
Scrophularia nodosa	II	3.8272	-2.9134	-	-	0.08	4.89	6.10	3.89	6.10	4.21	6.10
Scutellaria galericulata	IV	-5.8559	11.7813	7.5137	-	0.15	4.19	4.89	3.70	5.38	3.85	5.23
Senecio sylvaticus	IV	-2.9618	14.8229	5.4964	-	0.07	3.28	3.78	3.02	4.03	2.95	4.10
Sorbus aucuparia	II	-1.3654	4.6480	-	-	0.23	2.50	3.37	2.50	3.57	2.50	4.24
Stachys sylvatica	III	7.9529	-26.6316	-0.5644	-	0.54	3.51	6.10	3.53	6.10	4.07	6.10
Stellaria holostea	III	0.8167	-11.1627	-1.5473	-	0.12	3.43	6.10	3.26	6.10	3.49	6.10
Taraxacum sect. *Ruderalia*	II	3.9795	-2.7121	-	-	0.07	5.38	6.10	3.88	6.10	4.17	6.10
Thuidium tamariscinum	IV	-5.1955	10.4081	6.1025	-	0.14	4.05	4.86	3.60	5.31	3.70	5.21
Trientalis europaea	II	0.2139	14.5573	-	-	0.25	2.50	2.89	2.50	3.13	2.50	2.96
Urtica dioica	III	2.5959	-8.3311	-0.7902	-	0.26	3.82	6.10	3.56	6.10	3.99	6.10
Vaccinium myrtillus	II	-1.6232	18.5497	-	-	0.41	2.50	2.99	2.50	3.15	2.50	2.98
Valeriana procurrens	IV	-8.2887	12.8624	8.7856	-	0.31	4.53	5.25	4.05	5.73	4.29	5.49
Veronica chamaedrys	II	5.7815	-4.2692	-	-	0.13	5.20	6.10	4.09	6.10	4.71	6.10
Veronica montana	III	7.0210	-16.4292	0.5477	-	0.38	4.53	6.10	3.84	6.10	4.41	6.10
Viburnum opulus	IV	-6.2135	11.1471	7.6553	-	0.19	4.45	5.02	3.83	5.65	4.02	5.46
Viola reichenbachiana	III	5.2744	-14.9224	-0.2122	-	0.41	3.91	6.10	3.65	6.10	4.20	6.10

* pH values are rescaled by pH'=(pH-2.5)/3.6.

App. 2. Amplitude width vs. species gradient position and prevalence (full data set).
(**a**) A.prev, (**b**) A.kappa, (**c**) A.area Gradient position derived from species weighted averages.
Gradient position classes [pH]: A: ≤ 3.5 ; B : ≤ 4.5 ; C : > 4.5; Prevalence classes: 1: ≤ 0.2 ; 2 : ≤ 0.4 ; 3 : > 0.4

a)

b)

c)

OPULUSPRESS
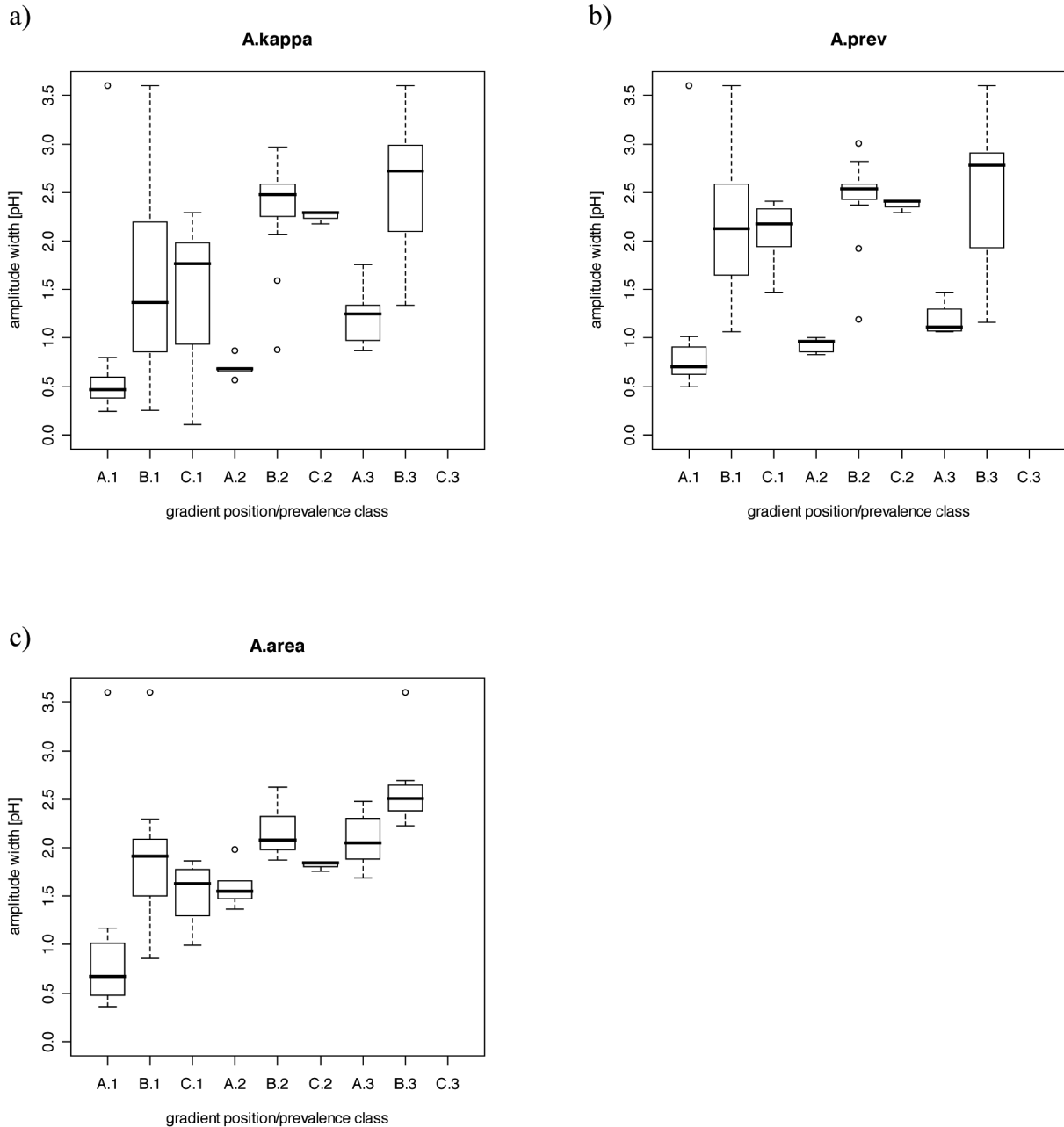
App. 3. Predicted vs. observed pH values for different prediction methods (full data set)
(**a**) ML-p/a, (**b**) ML-p, (**c**) IV-wa, (**d**) IV-e, (**e**) AO.prev, (**f**) AO.kappa, (**g**) AO.area