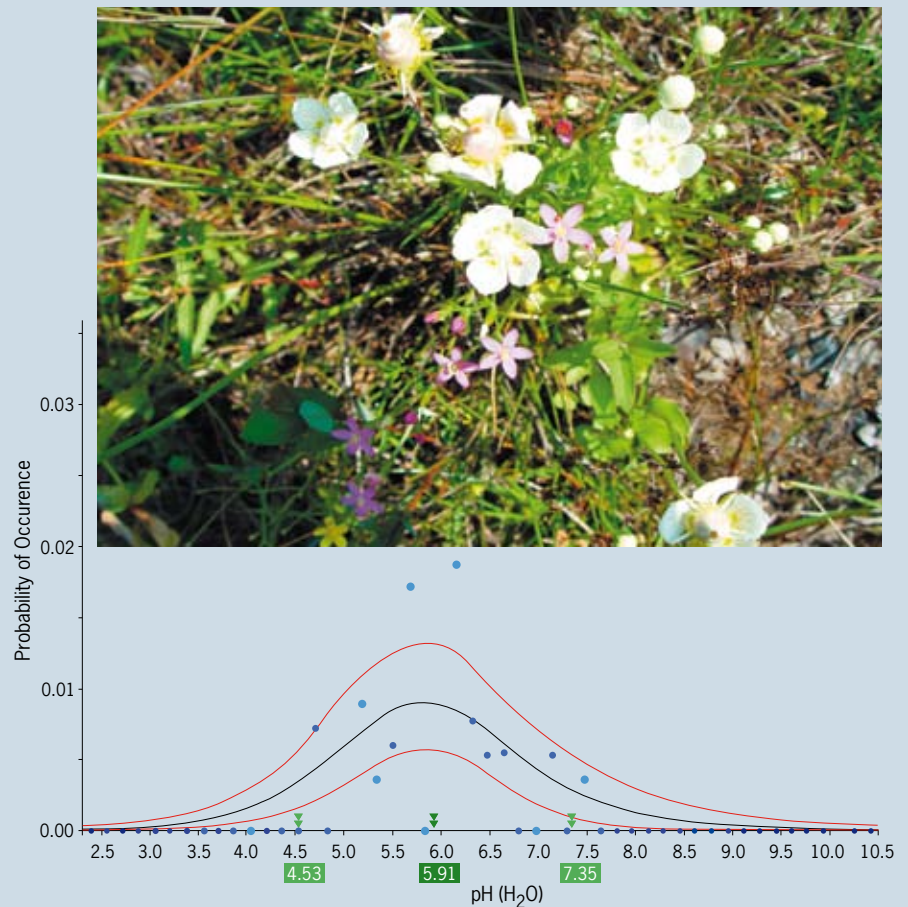




Response curves for plant species and vegetation types

G.W.W. Wamelink
P.W. Goedhart
J.Y. Frissel
R.M.A. Wegman
P.A. Slim
H.F. van Dobben

Alterra-rapport 1489, ISSN 1566-7197



Response curves for plant species and vegetation types

Response curves for plant species and vegetation types

G.W.W. Wamelink
P.W. Goedhart
J.Y. Frissel
R.M.A. Wegman
P.A. Slim
H.F. van Dobben

Alterra-rapport 1489

Alterra, Wageningen, 2007

ABSTRACT

G.W.W. Wamelink, P.W. Goedhart, J.Y. Frissel & R.M.A. Wegman, 2007. *Response curves for plant species and vegetation types*. Wageningen, Alterra, Alterra-rapport 1489. 20 blz.; 2 figs.; 2 tables.; 19 refs.

Responses of plant species and phytosociological classes were estimated based on field measured abiotic conditions and vegetation relevés. The responses for the phytosociological classes is limited to pH for now, however responses for plant species are estimated for 18 different soil conditions, for instance soil pH, spring groundwater table, potassium content of the soil, total nitrogen content of the soil. Most measured data were available for soil pH; therefore we were able to estimate most response curves for soil pH (547 species). Especially for nutrient contents of the soil the data availability is still limited, therefore responses for fewer species could be estimated. The optima for the pH curves were tested by predicting the soil pH for independent datasets throughout Europe. Compared with measured data the results were satisfying with an average difference of 0.5 pH unit. This report accompanies a CD (Plant species response) and a website (www.abiotic.wur.nl).

Keywords: planten soorten, abiotiek, pH, stikstof fosfaat, grondwaterstand, kalium, zout, zuurgraad, vocht, associatie, klasse, plant species abiotic, nitrogen, phosphorus, groundwater table, potassium, salt, acidity, moisture, association, phytosociological class.

ISSN 1566-7197

This report is available in digital format at www.alterra.wur.nl.

A printed version of the report, like all other Alterra publications, is available from Cereales Publishers in Wageningen (tel: +31 (0) 317 466666). For information about, conditions, prices and the quickest way of ordering see www.boomblad.nl/rapportenservice

© 2007 Alterra

P.O. Box 47; 6700 AA Wageningen; The Netherlands

Phone: + 31 317 474700; fax: +31 317 419000; e-mail: info.alterra@wur.nl

No part of this publication may be reproduced or published in any form or by any means, or stored in a database or retrieval system without the written permission of Alterra.

Alterra assumes no liability for any losses resulting from the use of the research results or recommendations in this report.

Contents

1	Introduction	7
2	Material & Methods	9
	2.1 Data set	9
	2.2 Response curves	9
	2.3 Associa	10
	2.4 Associations	10
	2.5 Reconstructed associations	11
	2.6 Quality control	11
	2.7 How to use the CD or website	11
3	Results	13
4	Discussion	15
	4.1 Regression to the mean	15
	4.1.1 Shortening of the response axis	15
	4.1.2 Shift of the response curve	15
5	Application range	17
	5.1 Species response curves	17
	5.2 Associations response curves	17
	5.3 Reconstructed association responses	17
	Literature	19

1 Introduction

The introduction is partly a summary of the introduction in Wamelink et al (2005), now extended to include soil parameters other than pH, and response curves for phytosociological classes.

This CD contains the attempt to characterize the response of a large set of plant species and phytosociological classes to single environmental factors (pH H₂O, pH KCl, spring groundwater table, highest groundwater table, lowest groundwater table, N-total, nitrate, ammonium, P-total, phosphate, potassium, chloride, calcium, sodium, C/N ratio and moisture content) on the basis of field measurements. The optima per response are given so they may be used as indicator values. Note that only the indicator values for species for soil pH are validated at this moment (2006).

Although many indicator systems work well in certain areas and vegetation types (Kruijne, De Vries & Mooi 1967, Zólomi et al. 1967, Landolt 1977, Ellenberg 1979, Ellenberg et al. 1991, Grime, Hodgson & Hunt 1988, Diekmann & Falkengren-Grerup 1998), there are several disadvantages:

1. Indicator systems do not provide information on ecological amplitudes. Species occur over a range of abiotic values, and the width of this range may vary per species. The indicator value per species is just one single value, which can be considered as the hypothetical optimum of the species.
2. Most systems are based on expert knowledge while only a minor part is based on field measurements. Wamelink et al. (2002) showed that expert systems can be biased, which restricts their application.
3. Often a transformation of the indicator values into physical values is necessary, for instance in the calculation of critical loads (Van Hinsberg & Kros 2001, Van Dobben & al 2006). The indicator values have an arbitrary scale, while the results of actual measurements are in physical units. The transformation of indicator values into variables with physical dimensions introduces a large amount of uncertainty (Ertsen, Alkemade & Wassen 1998, Schaffers & Sýkora 2000, Wamelink et al. 2002, Wamelink et al. 2003). It would be a significant improvement when this transformation could be circumvented by basing indicator values directly on actual measurements. The present indicator system is directly based on measurements in the field.

We developed a method to derive response curves for individual plant species to soil parameters. This method should be sufficiently general, so that it can be applied to all abiotic values. Our training set for the estimation of species responses consisted of a large data set (7509) of vegetation relevés and measured soil parameters from the Netherlands. The indicator values derived from this data set were applied to another data set containing vegetation relevés (app. 160000) with unknown soil pH that were syntaxonically identified. For each relevé, the average pH was estimated based on its species. The estimated pH was combined with the syntaxon to which it was assigned to construct a response curve for each syntaxon.

2 Material & Methods

2.1 Data set

The data set used for the estimation of the response functions for the species is an extension of the dataset described in [Wamelink et al. \(2002\)](#), now including 7509 records (date 14-12-2006). The ranges for the abiotic factors are given in [table 2.1.1](#) (date 14-12-2006).

Table 2.1.1. Measured ranges for abiotic parameters and the number of estimated species response functions

Abiotic parameter	Unit	n	Lower limit	Upper limit
pH H2O	-	547	2.3	10.5
pH KCl	-	280	2.5	9.1
Spring groundwater table	cm below surface	202	-48	212
Highest groundwater table	cm below surface	278	-61	501
Lowest groundwater table	cm below surface	255	-29	801
N total	g/Kg	122	0.1	31.0
nitrate	mg/Kg	39	150.0	0.1
ammonium	mg/Kg	47	0.09	365.1
P total	mg/Kg	183	0.0	1530
phosphate	mg/Kg	163	0.0	816
potassium	mg/Kg	164	0	1796
Chloride	mg/Kg	202	0.1	210000
Calcium	mg/Kg	58	0	22348
Sodium	mg/Kg	103	0	795
Magnesium	mg/Kg	58	0	806.4
C/N ratio	-	81	1.0	83.3
Moisture content	%	64	0	92

2.2 Response curves

For each combination of species and abiotic variable, a response curve was estimated when at least 25 records for that combination were available.

For relevés with a known abiotic variable such as pH, the presence-absence data of a species can be used to relate the probability of occurrence (p) of that species to pH. Penalized splines ([Eilers and Marx, 1996](#)) are used instead of smoothing splines, because penalized splines are easier to handle especially when using the bootstrap. However the fitted curves are very similar to the smoothing splines used in the paper by [Wamelink et al. \(2005\)](#). The concept of degrees of freedom of a spline is still used, to enable deviance testing for degrees of freedom. Note that a spline with one degree of freedom is equivalent to the linear logit model, while a spline with two degrees of freedom is already capable of fitting a bimodal response when the two modes are well separated. Penalized splines with 1, 2, ..., 10 degrees of freedom were fitted to all species responses. The "best" degrees of freedom of the spline was determined by backward deviance testing: the number of degrees of freedom was decreased one at a time, from 10 to 9, 8, etc., and was stopped when the resulting decrease in fit was

significant at the 1% level as judged on the basis of a deviance test. The solid black line in the resulting curves is the estimated species response curve. The solid red lines provide a 95% bootstrap interval for the response curve. This was obtained by bootstrapping the relevés and fitting the penalized spline to the bootstrapped data. A total number of 1000 bootstraps then yields 1000 response curves, and for every abiotic value the 2.5% and 97.5% percentage point of the response is saved.

2.3 Associa

Each relevé was assigned to a syntaxonomic unit (syntaxon) using the computer program ASSOCIA ([van Tongeren 2007](#), in press). ASSOCIA identifies vegetation relevés by comparing them with a training set that consists of relevés that have been classified beforehand. The training set used in our case was taken from Schaminée et al. ([1995](#), [1996](#) and [1998](#)) and [Stortelder et al. \(1999\)](#). ASSOCIA calculates the similarity of a given relevé to all relevés in the training set. The identification is based on both quantitative and qualitative data, i.e. presence-absence and abundance per plant species. A vegetation type is assigned to a relevé based on the calculated maximum likelihood using the dissimilarity between the relevé and the pre-classified relevés. The maximum likelihood combines the quantitative and the qualitative data into one index. ASSOCIA also calculates the ‘completeness’ and the ‘weirdness’ of the relevé. The completeness of the relevé gives information on the number of species that are expected to be present (according to the training set). The weirdness gives information on the number of ‘unexpected’ species in the relevé. The final assignment of a relevé to an association is based on the three above-mentioned characteristics, where the weirdness and incompleteness function as controllers. We used this information on all phytosociological levels. For each relevé, the association or sub-association that was closest to relevé was assigned to the relevé. This gives the advantage that relevés that are not so well developed (e.g., missing certain species) are classified as well. This gives a better estimation of the association response, since the limits of occurrence of the association can be estimated with greater accuracy. All relevés on the association and sub-association level are also part of the dataset for the higher hierarchical levels.

2.4 Associations

All relevés in the data set were assigned to a syntaxon by ASSOCIA (see [chapter 2.3](#)). Most of ASSOCIA's assignments appeared to be on hierarchical levels above the association. This resulted in relatively few assignments on the association and sub-association level, and therefore a response curve could be estimated for only 23 associations, a fraction of the over 300 associations. The uncertainty in the estimated responses is also relatively high. That is why we also used another approach to indirectly estimate the response of associations ([chapter 2.5](#)).

2.5 Reconstructed associations

In this novel approach we use an indirect method to estimate the response per syntaxon. This approach has now been used for soil pH but other responses will follow. As a starting point we used the 160,000 relevé data set described in chapter 1. For each relevé, the average pH value was inferred as the mean of the pH optima of its constituent species. A value was calculated when at least five species in a relevé were present with a known response for pH. A part of the relevés did not fulfil this criterion, however a set of over 130,000 relevés remained. For each syntaxonomic level a response curve was. This resulted in responses for 326 associations.

2.6 Quality control

The data set was subject to an intensive quality check, also as to gain the A-quality status for datasets. This is an internal Alterra quality standard, the requirements the dataset has to fulfil can be found on [quality reports](#) (unfortunately only in Dutch). It is expected that the A-quality status will be obtained in 2007. The pH responses were validated on independent data sets for Europe.

2.7 How to use the CD or website

Three data sets that can be selected:

1. Species response curves for various characteristics
2. syntaxon response curves for measured pH
3. syntaxon response curves for inferred pH

For the three sets, a pull down menu is available, where a species or a syntaxon can be selected. For all a new screen pops up with the following options: (for 'species' one may also read 'syntaxon')

The top of the screen displays four choices:

- "Home" takes you back to this screen so that you can select another dataset;
- "1 species" displays a single species response curve in a single frame at full size;
- "1 species 4 df" displays response curves of 1 species for 4 different degrees of freedom in 4 separate frames;
- "4 species" displays the response curves of 4 different species in 4 separate frames.

On the left you can choose one (or more) species for which response curves are displayed, as well as the number of degrees of freedom for the fitted penalized splines. This will display the fitted response curve in black and the 95% bootstrap interval in red for the fitted curve, see the Note below. The blue dots display the observed probabilities on an equidistant grid. These were obtained by first dividing the (transformed) abiotic values into 50 intervals. Then for every interval the observed probability was calculated by dividing the number of relevés at which the species was present by the total number of relevés in that interval. The size and colour of the blue dots depends on the number of relevés in the interval. Small dark blue dots represent few relevés, big light blue dots represent large numbers of

relevés; this is explained by means of the "Symbol Definition Rawdata" below the species information in the "1 species" menu.

You can also use the species translator on top of the screen to select a species for which the response curve must be displayed. In the "4 species" menu, this will only change the upper left species response curve.

The scale of the Y-axis of the graphs can be modified by means of the "Y" radio buttons.

In the "1 species" and "1 species 4df" menus species information is displayed on the left. This includes:

- "Npresent" the number of sites at which the species was present AND for which the abiotic value was available;
- "Mean Data" the mean of the abiotic values of all the relevés where that species is present;
- "Median Data" the median of the abiotic values of all the relevés where that species is present;
- "Sd Data" the standard deviation of the abiotic values of all the relevés where that species is present;
- "80% Data" the 80% interval, i.e. the interval between the 10 and 90-percentile, of the abiotic values of all the relevés where that species is present. Another percentage can be chosen at the top right of the page. Only intervals which are obtained by interpolation are given. So when a species is only present at say 39 sites, the 95% interval and higher are not given;
- "Df Curve" The degrees of freedom for which the Curve Mean/Median/Sd/80% information is given. The optimal number of degrees of freedom is given in parentheses. The best df was obtained by backward deviance testing: the number of degrees of freedom of the penalized spline was decreased one at a time, from 10 to 9, 8, etc., and was stopped when the resulting decrease in fit was significant at the 1% level as judged by a deviance test.
- "Mean Curve" the mean of the species response curve; this is calculated by viewing the response curve as a probability density;
- "Median Curve" the median of the species response curve, this is calculated by considering the response curve as a probability density function;
- "Sd Curve" the standard deviation of the species response curve; this is calculated by considering the response curve as a probability density function;
- "80% Curve" the interval of abiotic values which covers 80% of the area under the species response curve, another percentage can be chosen at the top right of the page;
- "Analysis of deviance" Detailed analysis of deviance with deviance contribution and corresponding p-value when degrees of freedom from 1 to 10 are subsequently added to the spline. This can be used to select degrees of freedom different from the optimal number

The Curve (or Data) interval and location are displayed in the graph by means of green arrows and accompanying numbers. The interval and location can be changed by using the picklists on the top right of the screen.

3 Results

The responses of the species and syntaxa are all available on the website/CD, including different degrees of freedom for the splines per species and syntaxon (www.abiotic.wur.nl). For each response the percentiles (1%, 5%, 10%, 20%, 50%, 80%, 90%, 95% and 99%) as well as the averages for the data and the response curve. All these values can also be found by using the links on the website/CD in Microsoft excel files. An example is given in [fig 3.1](#) with an explanation how to ‘read’ the results.

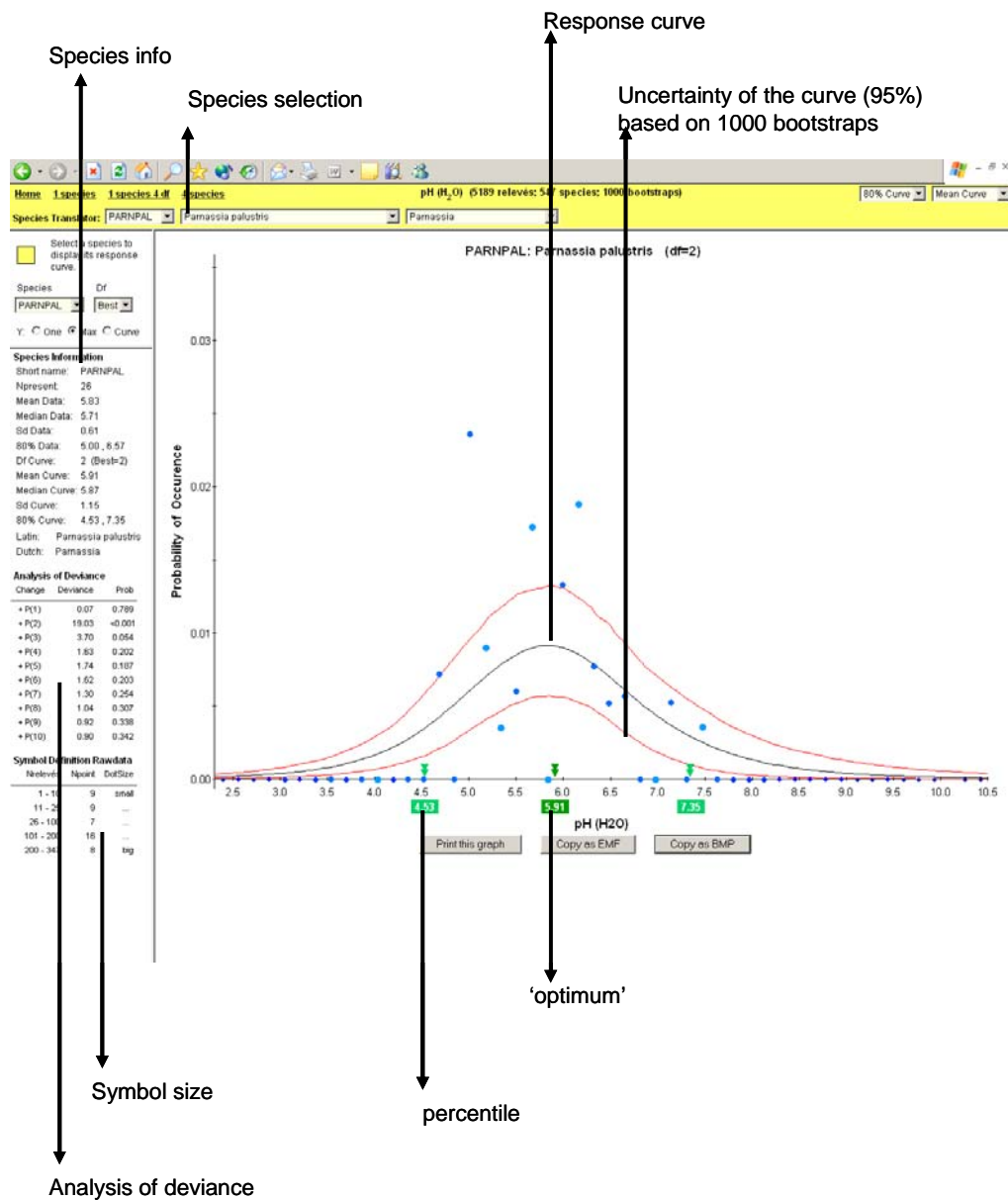


Fig. 3.1. Example of a species response curve on the CD/website with explanation of the features

4 Discussion

4.1 Regression to the mean

When calculating averages repeatedly, the results will always suffer from 'averaging to the mean'. In our case this phenomenon has two effects:

1. It shortens the responses.
2. It shifts optima and percentiles towards the overall average of the dataset.

Both phenomena are illustrated for pH.

4.1.1 Shortening of the response axis

The responses for soil pH-H₂O in the raw data, the minimum and maximum per species, the axis length for the reconstructed association responses and the minimum and maximum per reconstructed association are given in [Table 4.1.1.1](#). It is clear that every time an average is taken the pH axis is shortened; the minimum value rises from 2.3 in the raw dataset to 4.4 for the reconstructed associations, an increase of over two pH units. For the high pH value the decrease is even larger (three pH units). For a part this is natural since species and associations mostly do not have their optimum at the lowest or highest values, though for species this is theoretically possible. The latter is also a result of the way the optima for species are estimated. Even when the response curve reaches its maximum at the data limits, the optimum will be different from that, because we do not extrapolate.

Table 4.1.1.1. pH minimum and maximum values in the raw data, the averages of the species responses, the axis length of the reconstructed species response and the averages of the reconstructed associations.

	Minimum pH	Maximum pH
Raw data	2.3	10.5
Average species response	3.65	8.75
Axis for the reconstructed associations response	3.9	8.1
Average reconstructed associations response	4.39	7.46

4.1.2 Shift of the response curve

Using average species responses to estimate reconstructed association responses implies that the average is determined twice. This causes an extra shift of the optima, but also of the percentiles values, i.e. values lower than the average get higher and values higher than the average get lower. This effect is illustrated by regressing the averages of the directly estimated responses for associations on the averages of the reconstructed associations responses ([Fig. 4.1.2.1](#)). There is a highly significant relation between these averages, however the regression coefficient is only 0.54 instead of the expected 1 if there would be no shift. The regression line could be

used to recalculate the reconstructed pH values but is not done because the amount of available responses for directly estimated associations judged too low.

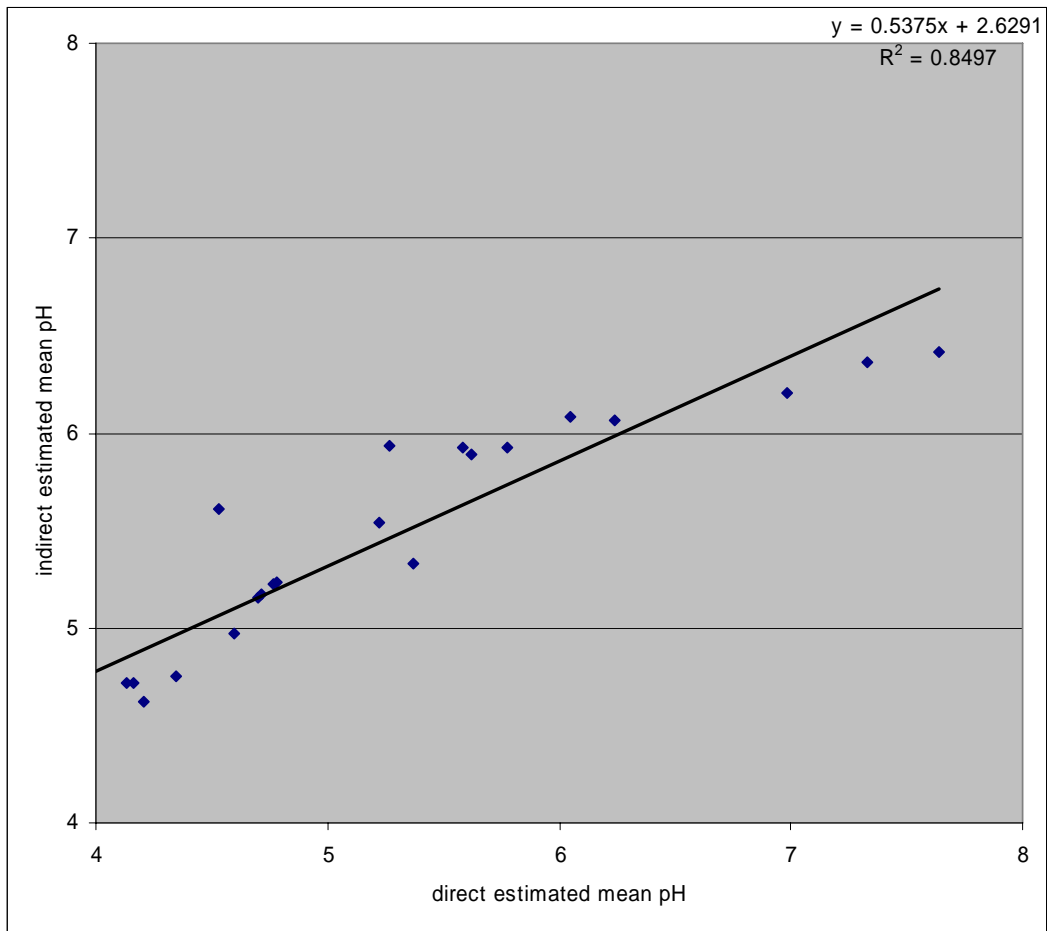


Fig. 4.1.2.1. Relation between the direct estimated average pH for associations and the indirect estimated average pH for the reconstructed associations.

5 Application range

5.1 Species response curves

The response curves of the species are solely based on field data collected in The Netherlands. Therefore, strictly speaking the responses are only applicable in The Netherlands and, to be more precise, only within the sites and vegetation types where the samples have been taken. However the application of the response curves to estimate pH values for relevés throughout Europe yielded satisfactory back predictions. The average uncertainty in forest was 0.5 pH units. The uncertainty for other vegetation types was higher, but in all cases lower than the uncertainty for Ellenberg indicator value for acidity. The uncertainty in Northern Europe was smaller than the uncertainty in Southern Europe. In our opinion, the values can be used for Europe, though especially for the south we warn to be cautious.

It may be clear that application of the indicator values outside the measured ranges (Table 2.1.1) for every parameter will increase the uncertainty and we recommend staying within the measured ranges.

5.2 Associations response curves

The directly estimated responses of associations are based on the same data set as the species responses and have therefore the same application ranges. However, due to lack of sufficient data the uncertainties are much larger than for the species responses. Therefore, we recommend to use them with great care. The responses are not yet tested on independent data.

5.3 Reconstructed association responses

The responses are based on the responses of the species in the relevés. Due to the effect of regression to the mean, the range of the response for pH is smaller than the range for the species. The lower pH value is 3.9 and the upper is 8.1. Especially the low values seem to be missing. Since the response is based on the averages of the responses of the species, the range is narrowed. The lowest and highest average of the response of the species is 3.65 for *Pseudotsuga menziesii* and 8.75 for *Zannichellia palustris ssp. pedicellata*. The directly estimated association responses have not tested yet been tested on an independent data set.

Literature

- Diekmann, M & Falkengren-Grerup, U. 1998. A new species index for forest vascular plants, development of functional indices based on mineralization rates of various forms of soil nitrogen. *Journal of Ecology* 86: 269-283.
- Eilers, H.C. & Marx, B.D. 1996. Flexible Smoothing with B-splines and Penalties. *Statistical Science* 11: 89-121.
- Ellenberg, H. 1979. *Zeigerwerte der Gefaesspflanzen Mitteleuropas - 2. Auflage*. *Scripta Geobotanica* 9: 9-97.
- Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulißen, D. 1991. *Zeigerwerte von Pflanzen in Mitteleuropa*. *Scripta Geobotanica* 18: 9-166.
- 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.
- Grime, J.P., Hodgson, J.G. & Hunt, R. 1988. *Comparative plant ecology, a functional approach to common British species*. Unwin Hyman, London.
- Hinsberg A. Van & Kros J. 2001. Dynamic modelling and the calculation of critical loads for biodiversity. (eds M. Posch, P.A.M. De Smet, J.P. Hettelingh & R.J. Downing), pp 73-80. Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven.
- Kruijne, A.A., Vries, D.D. de & Mooi, H. 1967. *Bijdrage tot de oecologie van de Nederlandse graslandplanten*. Pudoc, Wageningen.
- Landolt, E. 1977. *Oekologische Zeigerwerte zur Schweizer Flora, Heft 64*. Veröffentlichungen des Geobotanischen Institutes der ETH, Stiftung Rübel, Zürich.
- Schaminée, J.H.J., Stortelder, A.H.F. & Westhoff, V. 1995. *De vegetatie van Nederland*. Volume 2. Opulus press, Upsala.
- Schaminée, J.H.J., Stortelder, A.H.F. & Weeda, E.J. 1996. *De vegetatie van Nederland*. Volume 3. Opulus press, Upsala.
- Schaminée, J.H.J., Weeda, E.J. & Westhoff, V. 1998. *De vegetatie van Nederland*. Volume 4. Opulus press, Upsala.
- Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction, comparison with field measurements. *Journal of Vegetation Science* 11: 225-244.

Stortelder, A.H.F., Schaminée, J.H.J. & Hommel, P.W.F.M. 1999. De vegetatie van Nederland. Volume 5. Opulus press, Upsala.

Tongeren, O. van, Gremmen, N. & Hennekens, S. 2007. Assignment of relevés to pre-defined classes by supervised clustering of plant communities using a new composite index. *Journal of Vegetation Science*. Submitted.

Wamelink, G.W.W., Joosten, V., Dobben, H.F. van & Berendse, F. 2002. Validity of Ellenberg indicator values judged from physico-chemical field measurements. *Journal of vegetation science* 13: 269-278.

Wamelink, G.W.W. & Dobben H.F. van. 2003. Validity and uncertainty of Ellenberg indicator values. *Basic and applied ecology* 4: 515 - 523.

Wamelink, G.W.W, Goedhart, P.W., Dobben, H.F van & Berendse, F. 2005. Plant species as predictors of soil pH: replacing expert judgement by measurements. *Journal of vegetation science* 16:461-470.

Zólomi, B., Baráth, Z., Fekete, G., Jakucs, P. Kárpáti, I., Kovács, M. & Máthe, I. 1967. Einreihung von 1400 Arten der ungarischen Flora in ökologische Gruppen nach TWR-Zahlen. *Fragmenta Botanica Musei Historico-Naturalis Hungarici* 4: 101-142.