Vegetation data access and evaluation of taxon names Version 0.1

Florian Jansen

November 10, 2009

Abstract

An example session to show the usage of R library *vegdata*.
After installation of package vegdata you can run this script with

```
> vignette("vegdata")
```

The package can be downloaded from `http://geobot.botanik.uni-greifswald.de/download/`. It is not yet available through official R-cran Sites.

1 Provided functionality

1.1 Taxonomic evaluation

One of the most important step in using vegetation data (from different sources) is to take care about the taxonomic content of the used taxon names. That is, to make sure, that exactly one (correct and valid) name defines one biological entity. Most researchers remember to convert synonyms to valid names, but in many cases the care about e.g. monotypic subspecies or ambiguous taxonomic levels, is lacking. The package offers the function `tv.taxval()` with options for the adjustment of formas, synonyms, monotypic taxa, subspecies, members of aggregates and undetermined genera.

1.2 Cover standardization

Turboveg provides different abundance codes and all kinds of user defined cover codes can easily be added. For vegetation analysis an unique platform is needed which will mostly be the percentage cover of the observed area, so for every abundance code class the mean cover percentage is defined in Turboveg. Since different scales can occur in a database and the storage format of the code table in Turboveg is somewhat strange, the function `tv.coverperc()` provides automatic conversion.

1.3 Layer aggregation

The main sample unit in vegetation science is a so called relevé. A Braun-Blanquet relevé is a sample of names and coverage (abundance) of species in a specified area (usually between 1 an 1000 sqm). It contains (is intended to contain) a complete list of phytoautotrophic plants (or a special guild) in that area. The information can be stored in a three-column list of relevé ID, Taxon ID and performance measure (e.g. cover code).

Often additional information about the kind of occurrence of the specific plants is wanted. In Turboveg one additional column for the most widespread attribute is default: growth height classes. E.g. in a forest it is of interest, if a woody species reaches full height (tree layer) or occurs only as a small individual (herb layer). Other attributes like micro location (hummock or depression, rock or dead wood), development stage (juvenile or not, age, flowering status etc.) or the month of survey in a multi-seasonal survey could be of interest and can be added in Turboveg. For analysis you may want to differentiate species growing in different layers. Function `tv.veg()` provides options for species-plot attribute handling.

1.4 Vegetation matrix

Turboveg stores relevés as a list of occurrences (s. above) but almost all functions and programs for vegetation analyses are using plot / species tables with a 0 value for non-occurrence (measured absence). Function `tv.veg()` inflates the Turboveg list to matrix format.

2 Examples

Maybe the best way to introduce into the functionalities of the package is a session with example code.

2.1 Preparations

You can download the package from `http://geobot.botanik.uni-greifswald.de/download`.
Please ensure the right path to the R installation directory on your local computer. For a default Windows installation the procedure can look like this:

We load the library as usual into our R environment.

```
> library(vegdata)
```

If you want to use the taxonomic standard list for Germany, please look at: `'http://geobot.botanik.uni-greifswald.de/portal/reflist'`

```
> download.file("http://geobot.botanik.uni-greifswald.de/download/GermanSL%201.1/species.zip",
+     file.path(tempdir(), "vegdata.zip"), method = "internal",
+     mode = "wb")
> zip.unpack(file.path(tempdir(), "vegdata.zip"), file.path(tv.home(),
+     "Species"))
```

The package includes some example datasets and the taxonomic reference list Germansl 1.1, which can be loaded with option sysPath=TRUE. To get a better idea about the intended functionality of the code you can copy the datasets into your local Turboveg directory structure (see Turboveg help) and omit option "sysPath=TRUE".

```
> download.file("http://geobot.botanik.uni-greifswald.de/download/GermanSL%201.1/elbaue.zip",
+     file.path(tempdir(), "elbaue.zip"), method = "internal",
+     mode = "wb")
> zip.unpack(file.path(tempdir(), "elbaue.zip"), file.path(tv.home(),
+     "Species"))
```

Define your Turboveg Database (for general information about database structure see Turboveg Help). The function tries to guess the path of Turboveg installation (for Windows and Linux with Wine), please specify explicit in case of trouble.

Main functions The package contains two main functions:

tv.veg() is a wrapper for other functions to support the database load from Turboveg, taxonomic amendment, cover unification, layer combination and matrix generation.

tv.site() will load the site (header) data and does some basic corrections.

But first look at the metainfo which should always be created for every Turboveg database. We recommend a simple text format stored in the (Turboveg) database directory which can than be loaded by:

```
> metainfo("database_name")
```

2.2 Site data

```
> require(vegdata)
> db <- "taxatest"
> site <- tv.site(db, sysPath = TRUE)
```

```
 The following columns contain no data and are omitted
 [1] REFERENCE  TABLE_NR   NR_IN_TAB  PROJECT    AUTHOR     SYNTAXON
 [7] UTM        ALTITUDE   EXPOSITION MOSS_IDENT LICH_IDENT

 The following numeric columns contain only 0 values and are omitted
 [1] COV_TOTAL  COV_TREES  COV_SHRUBS COV_HERBS  COV_MOSSES COV_LICHEN
 [7] COV_ALGAE  COV_LITTER COV_WATER  COV_ROCK   TREE_HIGH  TREE_LOW
[13] SHRUB_HIGH SHRUB_LOW  HERB_HIGH  HERB_LOW   HERB_MAX   CRYPT_HIGH

 The following numeric fields contain 0 values:
[1] INCLINATIO

Please check if these are really measured as 0 values
 or if they are not measured and wrongly assigned
 because of Dbase restrictions.
If so, use something like:
 site$Column_name[site$Column_name==0] <- NA
 summary(site[,c('INCLINATIO')])
```

The function **tv.site()** is quite straightforward. After loading the file tvhabita.dbf of the specified database, warnings are given for plots without specified relevé area or date and the fields are checked if they are empty (a lot of predefined header fields in Turboveg are often unused) or contain probably mistakable 0 values in numerical fields, due to DBase deficiencies (DBase can not handle NA = not available values reliably). It is stated in the output, if you have to check 0 values.

2.3 Vegetation data

Now we care about the species occurrencies data.
Simple loading of species observation data from Turboveg is done by function **tv.obs()**

```
> obs <- tv.obs(db, sysPath = TRUE)

reading observations ...

> head(obs)

  RELEVE_NR SPECIES_NR COVER_CODE LAYER DET_CERT SEASON MICROREL FLOWER
1         2         27         2b     0        0      0 Schlenke      0
2         2        100          1     0        0      0 Schlenke      0
3         2       4685          4     1        0      0 Schlenke      0
4         2       4685          1     2        1      0 Schlenke      0
5         2       4685          1     6        0      0     <NA>     10
6         2       4721          4     0        0      1 Schlenke      0
```

Data is stored in Turboveg as a flat table of occurrency values, that is one species-plot occurrence per row. Field **RELEVENR** contains the plot number, **SPECIESNR** the taxon codes, **COVERCODE** the performance code and all other columns show species-plot attributes like growth hight classes.

2.3.1 Names and entities

If you want to know the species name for a species number or shortletter or vice versa you can use:

```
> tax("ACERNEG", sysPath = TRUE)

  SPECIES_NR LETTERCODE    SHORTNAME    ABBREVIAT   NATIVENAME AUTHOR SYNONYM
6          8    ACERNEG Acer negundo Acer negundo Eschen-Ahorn     21   FALSE
  VALID_NR
6        8

> tax(27, tax = TRUE, sysPath = TRUE)

     SPECIES_NR                 ABBREVIAT SYNONYM VALID_NR
2780         27 Achillea millefolium agg.   FALSE       27
                    VALID_NAME GRUPPE RANG   AGG       AGG_NAME
2780 Achillea millefolium agg.      S  AGG 60728 Achillea spec.
                          NACHWEIS                          SECUNDUM HYBRID
2780 BfN (Wisskirchen & Haeupler 1998) BfN (Wisskirchen & Haeupler 1998)      0
     BEGRUENDUN IN_QUELLE_ AUTONYM ELTER_1 ELTER_2 ELTER_3 EDITSTATUS
2780       <NA>       <NA>    <NA>    <NA>    <NA>    <NA>        BfN
```

As stated in the beginning the care about the taxonomic entegrity of your database should stay at the beginning of every vegetation analyses. For Turboveg databases with taxnomic reference list GermanSL (versions 0.9, 1.0 or 1.1) this can be done semi-automatic.

To run the taxonomic adjustments of the example dataset use **tv.taxval()**

```
> obs <- tv.taxval(db, obs, sysPath = TRUE)


Original number of taxa: 35

 6 Synonyms found in dataset, adapted
      SPECIES_NR                              ABBREVIAT VALID_NR
4338       20098       Achillea millefolium var. lanata       34
5230       27309                Armeria bottendorfensis    20585
7609       20096      Achillea millefolium ssp. collina       29
7763       25203                        Abies alpestris     4269
9853       20583  Armeria maritima ssp. bottendorfensis    20585
20090      80425 Brachythecium salebrosum var. palustre    80400
                         VALID_NAME
4338             Achillea pannonica
5230  Armeria maritima ssp. halleri
7609               Achillea collina
7763                    Picea abies
9853  Armeria maritima ssp. halleri
20090       Brachythecium mildeanum

 6 Variants, forms, subspecies etc. found also at species level in dataset, combined at species level
      SPECIES_NR                          ABBREVIAT   AGG             AGG_NAME
1865          33 Achillea millefolium ssp. sudetica    31 Achillea millefolium
5313        4722 Ranunculus ficaria ssp. bulbilifer  4721   Ranunculus ficaria
13135       7161       Poa trivialis ssp. trivialis 26611        Poa trivialis
14576        100 Agrimonia eupatoria ssp. eupatoria    99  Agrimonia eupatoria
26287      90072     Adonis aestivalis * aestivalis    76    Adonis aestivalis
26288      90073   Adonis aestivalis * provincialis    76    Adonis aestivalis

 4 members of occurring aggregates in dataset, aggregated:
      SPECIES_NR            ABBREVIAT   AGG                  AGG_NAME
1863          31 Achillea millefolium    27 Achillea millefolium agg.
12032       2923  Hieracium pilosella 12273 Hieracium subg. Pilosella
12188         29     Achillea collina    27 Achillea millefolium agg.
14194         34   Achillea pannonica    27 Achillea millefolium agg.

 5 Monotypic taxa found in dataset, species converted to lower rank.
     AGG_NR               AGG_NAME AGG_RANG MEMBER_NR
4        99    Agrimonia eupatoria      SPE       100
198    4721     Ranunculus ficaria      SPE      4722
319   26611          Poa trivialis      SPE      7161
966   66142        Acoraceae spec.      FAM     61329
1404  94619 Anthocerotopsida spec.      KLA     94653
                              MEMB_NAME MEMB_RANG
4    Agrimonia eupatoria ssp. eupatoria       SSP
198  Ranunculus ficaria ssp. bulbilifer       SSP
319        Poa trivialis ssp. trivialis       SSP
966                        Acorus spec.       GAT
1404             Anthocerotales spec.       ORD

 1 Monotypic taxa found in dataset, species converted to lower rank.
    AGG_NR     AGG_NAME AGG_RANG MEMBER_NR      MEMB_NAME MEMB_RANG
851  61329 Acorus spec.      GAT        69 Acorus calamus       SPE

 Warning: Critical Pseudonym(s) in dataset, please check
          to_check checknr        check against SPECIES_NR   VALID_NAME
4876 Galium mollugo    2555 Galium mollugo auct.      27395 Galium album
     VALID_NR                          SECUNDUM
4876     2549 BfN (Wisskirchen & Haeupler 1998)

 Warning: Critical species in dataset, please check
                               to_check checknr
2112             Arenaria serpyllifolia   13703
6438  Dactylis glomerata ssp. glomerata   13464
12052                    Galium mollugo    2555
                                check against SPECIES_NR
2112           Arenaria serpyllifolia s. str.        567
```

Have a look at **?tv.taxval** or **args(tv.taxval)** for options.

Taxonomic evaluation of vegetation data sets can only be performed with checklists containing appropriate taxonomic information (see tax.dbf and monotypic-D.dbf for GermanSL 1.1 Jansen & Dengler (2008).
If your database is not referenced with GermanSL you can not use **tv.taxval** and you have to execute **tv.veg()** with option tax=FALSE or convert your database to GermanSL (Export to XML in Turboveg and re-import choosing the new GermanSL) assuming you have a central european database. You can download the reference list in different formats from:
`http://geobot.botanik.uni-greifswald.de/portal/reflist`.

German SL is based upon the **taxon views** (Berendsohn (1995) of available checklists but contains more than 16,000 synonyms which can be used to switch between different taxon views.

To deal with a different taxonomic concept than the one used in GermanSL, you can use the option **concept**. For this a file is necessary indicating the new synonymy status, valid names and new aggregation. Within the package only a small example list (**korneck1996.dbf**) for the taxonomic view of *Armeria maritima* from (Korneck *et al.*, 1996) is implemented. Please compare the following examples.

```
> tv.taxval("taxatest", quiet = TRUE, sysPath = TRUE)
> tv.taxval("taxatest", concept = "korneck1996", quiet = TRUE,
+     sysPath = TRUE)
```

2.3.2 Cover values

Cover is coded in Turboveg as an alphanumeric code. Different codes can be combined by using the mean cover percentage per cover class. Function **tv.coverperc** will do this job according to the definitions in *Turboveg/Popup/tvscale.dbf*.

```
> head(obs)
```

```
  RELEVE_NR SPECIES_NR COVER_CODE LAYER DET_CERT SEASON MICROREL FLOWER
1         2         27         2b     0        0      0 Schlenke      0
2         2        100          1     0        0      0 Schlenke      0
3         2       4685          4     1        0      0 Schlenke      0
4         2       4685          1     2        1      0 Schlenke      0
5         2       4685          1     6        0      0     <NA>     10
6         2       4722          4     0        0      1 Schlenke      0
```

```
> obs <- tv.coverperc(db, obs, sysPath = TRUE)
```

```
[1] Cover code used: Braun/Blanquet (old)
code        r        +        1        2        3        4        5
perc         1         2         3        13        38        68        88
[1] Cover code used: Braun/Blanquet (new)
code        r        +        1        2m        2a        2b        3        4        5
perc         1         2         3         4         8        18        38        68        88
```

```
> head(obs)
```

```
  RELEVE_NR SPECIES_NR COVER_CODE LAYER DET_CERT SEASON MICROREL FLOWER
1         2         27         2b     0        0      0 Schlenke      0
2         2        100          1     0        0      0 Schlenke      0
3         2       4685          4     1        0      0 Schlenke      0
4         2       4685          1     2        1      0 Schlenke      0
5         2       4685          1     6        0      0     <NA>     10
6         2       4722          4     0        0      1 Schlenke      0
  COVERSCALE COVER_PERC
1         02         18
2         02          3
3         02         68
4         02          3
5         02          3
6         02         68
```

2.3.3 Pseudo-species, layer combinations and vegetation matrix

tv.veg() is a wrapper for the above mentioned functions and produces a vegetation matrix with releves as rows and species as columns. Additionally care about species-plot attribute differentiation and combination, the inflation of a vegetation matrix and the handling of species names is provided.

If we have more than one occurrence of the same species in a plot, e.g. because tree species growing as young stands and adult specimens were differentiated according to growth hight classes we have to create either pseudo-species which differntiate the occurrences in the resulting vegetation matrix or to combine species occurrences from different layers. For the latter we can use different calculations i.e. mean, max, min or first value. If we assume an independent occurrence of a species in different layers, a tree with a cover of 50% in tree layer and 50% in herb layer can be accounted with an overall cover of 75%. This is done with option lc = ’layer , the default.

If you want to differentiate species according to layer or other species-plot attributes you can specify which attributes should be used for differentiation, and how pseudo-species should be labelled (e.g. speciesname.layercode). Two example data frames for layer differentiation are included in the package. lc.0 uses all Turboveg layers (0 to 9) for pseudo-species differentiation. lc.1 combines tree layers and shrub layers to a maximum of three pseudo-species per taxon.

```
> lc.1
```

```
   LAYER  COMB
1      0     0
2      1  Tree
3      2  Tree
4      3  Tree
5      4 Shrub
6      5 Shrub
7      6     0
8      7     0
9      8     0
10     9     0
```

```
> veg <- tv.veg(db, lc = "sum", comb = list(lc.1, c("LAYER")),
+     dec = 1, quiet = TRUE, sysPath = TRUE)
```

```
> veg[, 1:7]
```

```
  ABIEALB.Tree ABIECON.Tree ACERPSE ACERPSE.Shrub ACHI#MI ACOUCAL ADONAES
1            3            3       3            13      44       0       3
2            0            0       0             0      18       0       6
3            0            0       0             0       0       3       0
```

If you want to use only presence/absence information in your analyses you can do:

```
> veg[veg > 0] <- 1
```

2.4 Additional functions

freqtab produces a relative or absolute frequency table of a vegetation table classification with the possibility to filter according to threshold values.

```
> db <- "elbaue"
> veg <- tv.veg(db, quiet = TRUE, sysPath = TRUE)


reading observations ...
[1] "Germansl1.1"
Original number of taxa: 53

 6 Synonyms found in dataset, adapted

 No conflicting aggregates found.

 1 Monotypic taxa found in dataset, species converted to lower rank.

 Warning: Critical species in dataset, please check

 No undetermined genera or above found.

 Number of taxa after validation: 53


 converting cover code ...

 creating pseudo-species ...

 combining occurrences using type LAYER and creating vegetation matrix ...

 replacing species numbers with shortletters ...


> site <- tv.site(db, sysPath = TRUE)


 The following columns contain no data and are omitted
 [1] REFERENCE  TABLE_NR   PROJECT    AUTHOR     SYNTAXON   UTM
 [7] ALTITUDE   EXPOSITION INCLINATIO MOSS_IDENT LICH_IDENT REMARKS

 The following numeric columns contain only 0 values and are omitted
 [1] COV_TOTAL  COV_TREES  COV_SHRUBS COV_HERBS  COV_MOSSES COV_LICHEN
 [7] COV_ALGAE  COV_LITTER COV_WATER  COV_ROCK   TREE_HIGH  TREE_LOW
[13] SHRUB_HIGH SHRUB_LOW  HERB_HIGH  HERB_LOW   HERB_MAX   CRYPT_HIGH

 The following numeric fields contain 0 values:
[1] INUND    INUND_50

Please check if these are really measured as 0 values
 or if they are not measured and wrongly assigned
 because of Dbase restrictions.
If so, use something like:
 site$Column_name[site$Column_name==0] <- NA
 summary(site[,c('INUND','INUND_50')])


> cluster <- site$MGL < -100
> freqtab(veg, cluster, limit = 20, relfr = TRUE, sysPath = TRUE)
```

```
[1] "Number of clusters:  2"
                                           FALSE. TRUE.
Achillea millefolium                            0    38
Agrostis capillaris                            10    54
Agrostis stolonifera                           30    15
Alopecurus geniculatus                         30     0
Alopecurus pratensis                           45    77
Anthoxanthum odoratum                          20    23
Cardamine pratensis                            45     8
Carex acuta                                    60     0
Carex praecox                                   5    69
Carex vesicaria                                30     0
Carex vulpina                                  25    15
Elymus repens                                  15    77
Erophila verna                                  0    31
Euphorbia esula                                 0    23
Galium palustre                                65    15
Galium verum ag.                                5    46
Glyceria maxima                                45     0
Holcus lanatus                                 30     8
Juncus effusus                                 35     0
Phalaris arundinacea                           70    15
Poa palustris                                  45    38
Poa pratensis ag.                              30    62
Poa trivialis s. trivialis                     40    15
Potentilla anserina                            25     0
Ranunculus flammula                            30     0
Ranunculus repens                              70    31
Rorippa sylvestris                             25    15
Rumex thyrsiflorus                              0    69
Silene flos-cuculi                             25     0
Sium latifolium                                35     0
Stellaria palustris                            45     8
Taraxacum Sec. Alpina, Hamata et Ruderalia     20    62
Trifolium repens                               30     0
Vicia cracca                                   15    23
Vicia tetrasperma                               5    31
```

Use help(package='vegdata') for a complete list of available functions and data sets in vegdata.

At `http://geobot.botanik.uni-greifswald.de/download/` a development version of package vegdata is available with additionla functionalities (but less stability).

The package vegdata serves only as a helper for further analysis of vegetation data which can already be done by powerful R packages like vegan. But with the functions shown above we are now ready to execute all kinds of analyses in the wide area of vegetation analyses.

2.5 NMDS

For instance we can do a “Nonmetric Multidimensional Scaling with Stable Solution from Random Starts Axis Scaling and Species Scores” which is a wrapper for Kruskal’s Non-metric Multidimensional Scaling (Cox & Cox, 1994, 2001) from Jari Oksanen (Oksanen *et al.*, 2008).

```
> library(vegan)
> veg.nmds <- metaMDS(veg, distance = "bray", trymax = 5, autotransform = FALSE,
+     noshare = 1, expand = TRUE, trace = 2)
> plot(veg.nmds)
```

To show the result in comparison with environmental measurements we do some magic.

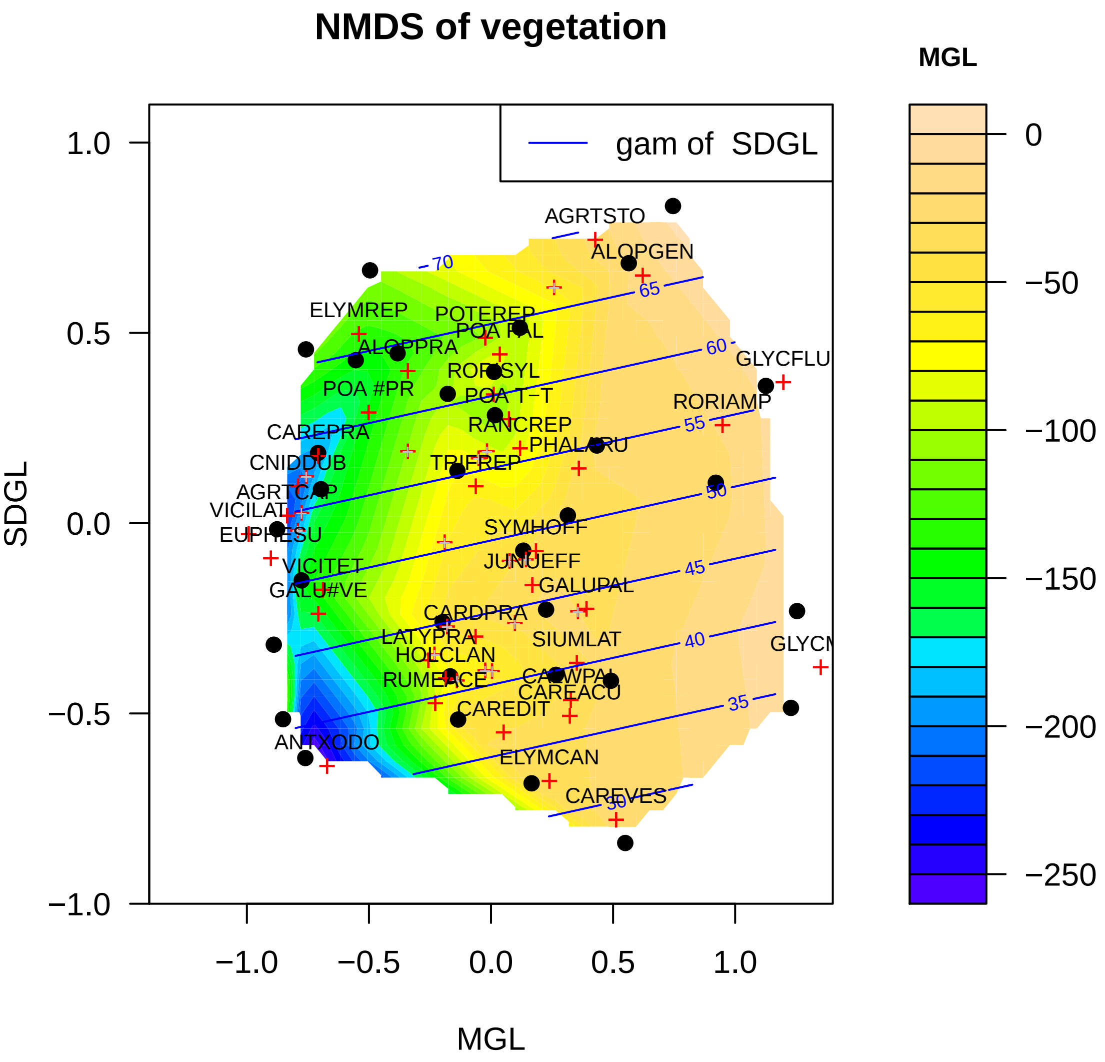
```
> library(labdsv)
```

```
This is mgcv  1.5-6 . For overview type ‘help("mgcv-package")’.
```

```
> library(akima)
> nmds.plot <- function(nmds, site, disp, var1, var2) {
+     var.1 <- site[, var1]
+     var.2 <- site[, var2]
+     lplot <- nrow(nmds$points)
+     lspc <- nrow(nmds$species)
+     samp <- sample(1:lspc, lspc/5)
+     stems <- colSums(veg)
+     filled.contour(interp(nmds$points[, 1], nmds$points[, 2],
+         var.1), ylim = c(-1, 1.1), xlim = c(-1.4, 1.4), color.palette = topo.colors,
+         xlab = var1, ylab = var2, main = paste("NMDS of vegetation"),
+         key.title = title(main = var1, cex.main = 0.8, line = 1,
+             xpd = NA), plot.axes = {
+             axis(1)
+             axis(2)
+             points(veg.nmds$points[, 1], veg.nmds$points[, 2],
+                 xlab = "", ylab = "", pch = 19)
+             points(veg.nmds$species[, 1], veg.nmds$species[,
+                 2], xlab = "", ylab = "", col = 2, pch = "+")
+             ordisurf(veg.nmds, var.2, col = 4, choices = c(1,
+                 2), add = TRUE)
+             orditorp(veg.nmds, display = disp, pcol = "gray",
+                 pch = "+", pos = 3)
+             legend("topright", paste("gam of ", var2), col = 4,
+                 lty = 1)
+         })
+ }
```

```
> print(nmds.plot(veg.nmds, site, disp = "species", var1 = "MGL",
+     var2 = "SDGL"))
```

```
NULL
```



The first axis of our NMDS plot show the influence of mean groundwater level on the patterns of the dataset. *Glyceria maxima* is marking the wet side of the gradient, whereas *Cnidium dubium Agrostis capillaris* or *Galium verum agg,* occur only at low mean groundwater level. The second axis can be assigned to the fluctuation of water levels measured as standard deviation of mean groundwater level. Species indicating high water fluctuation are *Agrostis stolonifera* or *Alopecurus geniculatus* whereas *Carex vesicaria* occurrs only at more balanced situations.

References

Berendsohn, W.G. (1995). The concept of "potential taxa" in databases. *Taxon*, 44, 207–212.

Cox, T.F. & Cox, M.A.A. (1994, 2001). *Multidimensional Scaling.* Chapman & Hall.

Jansen, F. & Dengler, J. (2008). Germansl - eine universelle taxonomische referenzliste für vegetationsdatenbanken. *Tuexenia*, 28, 239–253.

Korneck, D., Schnittler, M. & Vollmer, I. (1996). Rote Liste der Farn- und Blütenpflanzen (Pteridophyta et Spermatophyta) Deutschlands. *Schriftenreihe für Vegetationskunde*, 28, 21–187.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L. & Stevens, M.H.H. (2008). *vegan: Community Ecology Package*.