VNDM / NDM
Version 3.1

VNDM / NDM are two sister programs for analyzing areas of endemism, written by Pablo A. Goloboff (implementing the methods described in Szumik et al., 2002; Szumik & Goloboff, 2004; Aaagesen et al. 2013; Szumik & Goloboff 2015; and Casagranda & Goloboff in prep.).

VNDM reads the input data, which can be in the form of point (coordinate) data or in the form of a presence/absence grid, and allows viewing and handling the results. NDM is a a sort of "search engine", which is called from VNDM to do the computer-intensive parts of the analysis.

INPUT DATA

VNDM reads three types of input data: *.xyd, *.dat, or *.ndm.

XYD files contain point records (coordinates) for each of the species analyzed. It is the preferred form for inputting data!

DAT files contain a presence/absence matrix.

NDM files are binary, and may contain data (species distributions on a grid) and areas, or areas only. The only way in which areas (i.e. solutions) are stored is as NDM files. Note that the definition of the solutions in the NDM file must match the settings in the XYD or DAT file (i.e. the same number of rows, columns, species, and other settings).

XYD and DAT files can (must) be edited using a text-editor. NDM files can only be read/saved by VNDM and NDM.

XYD FILES

The format for *.xyd files is:

```
[ commands ] optional commands
spp n number of species, mandatory
xydata
sp 0 [ name of species 0 ] species name is optional, must
       4.56,20.25 be within square brackets; it
       5.61,21,22 can be followed by individual
       3.41,23.24 filling values
sp 1 [ name of species 1 ]
ETC.
```

The optional commands are:

nocommas	*lat-long data are not separated by commas; in this case, only two values can be read (if using commas, you can use coordinate data from other programs, which include more than two values separated by commas –only the first two are used, the rest are ignored).*
transpose	*read data as long-lat instead of the (default) lat-long.*
longlat	*included to make sure what you mean*
latlong	*ditto*
xnegative	*invert the sign of the latitude coordinate*
ynegative	*invert the sign of the longitude coordinate*
xnegativemap	*likewise, for map line definitions*
ynegativemap	*ditto*
autogrid	*determine grid size from point density*
automatrix	*after reading data, create binary matrix automatically*
mminusx n	*shift lat values by specified value*
mminusy n	*ditto, for long*
rows n	*use n rows for grid*
cols n	*ditto, columns*
gridx n c	*place grid origin at lat n, with a cell width of c*
gridy n h	*place grid origin at long n, with a cell height of h*
fill x y	*fill cells around a point, with x% of cell width and y% of height*
assume x y	*ditto, for filling cells with assumed records (see below, under dat).*
autosample N	*sample records and repeat searches, automatically, using N pseudoreplications. Reference areas are those read together with the data (i.e. as *.ndm file(s)). N indicates type: 1, all records; 2, selected species; 3, within (or outside) area (within, if* sampleareacut *is positive, outside, if* sampleareacut *is negative). This provides a rough estmate of the robustness of the results.*
autoconsense	*automatically produce consensus areas (to be used together with* autosample*)*
autosave	*autosave results, to different *.emf files; to be used in conjunction with* autosample *and* autoconsense
samplerec P	*probability of eliminating a record, (for all species, in* autosample 1, *or for selected species, in* autosample 2 *or* 3*).*
samplespp P	*probability of choosing a species for record elimination, in* autosample 2 or 3.
samplearea P	*probability of selecting a species for record elimination within or outside reference area (reference area is last area held in memory; this is used only as reference; recovery stats are not given for that area), when doing* autosample 3. *If P is positive, then it eliminates records within the reference area; if P is negative, then it eliminates records outside.*
sampleareacut S	*Determine the cutoff S for similarity in cell composition (for two areas to be considered "the same"). If S=0, no cutoff, uses average of all pseudoreplications. If S>0, then areas are matched (considered identical) only when cell composition is above threshold (i.e. count number of pseudoreplications above threshold). If S<0, then the*

samplesppcut P   *Determine the cutoff S for similarity in species composition (for two areas to be considered "the same"). S works as for* sampleareacut.

DAT FILES

The format for *.dat files is:

```
Spp N
Rows N
Cols N
data
        A0-0 0001
        A0-1 0101
        A0-2 1111
```

      If a cell is ommitted, then it is considered to be empty (i.e. containing no records) for all S species. If fewer than S species are indicated for a cell, the rest of the species (i.e. all the unspecified ones) are considered as absent from that cell. There may be several entries for a cell; subsequente entries continue defining presence/absence for the next species. This can be used to easily "fuse" or combine data sets specifying distributions of different sets of species. An example of this option is in the file liecoy.dat, in which cell 1-0 has two entries. The example also shows that the cells can be specified in any order you wish.

      The matrix may contain entries 0, 1 ("observed") or 2 ("assumed to be there but not observed"). Presence of a species outside an area decreases the score more than the *assumed* presence outside the area, and assumed presence in a cell inside the area does not decrease the score as much as a real absence (0).

The following commands may be optionally included before "data":

```
Xnegative
Ynegative
Gridx n c
Gridy n h
```

They work in the same way as in a XYD file.

NDM files

NDM files are binary files. As such, they cannot be edited or modified by the users, except through VNDM/NDM themselves.

READING SEVERAL DATA FILES

It is possible to specify, in the command line (e.g. when invoking the program from the DOS console) files of mixed types. In that case, VNDM establishes a precedence in the

following way:

1) *.dat files are always read first, and used to determine number of rows, cols, and species, as well as species distributions (absent, observed, assumed). If an *.xyd file is also specified, only the point records for the species are taken from that file (the filling of cells specified in the *.dat file is honored). If an *.ndm is also specified, it is expected to contain only areas (not a specification of the data themselves).

2) If no *.dat file is specified, then the *.xyd file is assumed to contain the data; if "automatrix" included in the file, then the matrix is created automatically from the point records after reading the file (using the values for filling specified in the file). If an *.ndm file is also included, it is expected to contain only areas (not a specification of the data themselves). Note that an an *.ndm file can be specified in conjunction with an *.xyd file only if "automatrix" is used; otherwise an error message is displayed.

3) If only an *.ndm file is specified, then that file is expected to begin with specification of data (numbers of species, columns and rows, and species distribtuions, optionally followed by specification of areas).

No more than a *.dat or *.xyd file can be specified; several *.ndm files can be specified if they contain only areas (e.g. in conjunction with a *.dat or *.xyd file).

The scores calculated for the areas are stored in the *.ndm file, so that this provides some (tenuous) safeguard against reading areas for a different data set; if a difference in score exists, a warning is displayed. Reading the areas anyway, may provide bogus definitions of included/excluded cells, if the *.ndm file corresponds to a different number of columns and rows.

LINE FILES
Another type of input file is *.lin files, which define lines to draw maps, contours, rivers, etc. Line files contain:
(R,G,B,t) *optional, color as red-green-blue, and*
thickness t
line *this string begins a new line*
    x,y
    x,y
    x,y
line
etc.

As in *.xyd files, more than two comma-separated values can be included for each point, but only the first two are considered (the others are ignored). All the points specified for each line are connected through a line of the specified color and thickness; to indicate the end of a line, you can either begin a new **line**, or give as coordinates the starting point (this is the method used by Global Mapper, so that files from that program can be easily imported into VNDM).

RUNNING NDM

If you select "Analyze with NDM", VNDM will create automatically all the options needed to run NDM with the parameters specified, create a temporay data file, and pass it onto NDM. The temporary data are always written into a file called *tmp.dat*, and the results saved by NDM to file *tmp.ndm* (containing only areas).

COMPILATION

The source code for both NDM and VNDM is included. Both programs can be compiled with the (free) Watcom compiler (http://www.openwatcom.org). You need to create a project for each of the two programs; for VNDM, before building, you may need to define "NEWEST_WATCOM" for the Resource Compiler Switches. Other than that, compilation should be straightforward.

REFERENCES

Aagesen, L., Szumik, C., & Goloboff, P. 2013. Consensus in the search for Areas of Endemism. **Journal of Biogeography** 40, 2011–2016.

Szumik, C., Cuezzo, F., Goloboff, P., & Chalup, A. 2002. An optimality criterion to determine areas of endemism. **Systematic Biology** 51:806-816.

Szumik, C., & Goloboff, P. 2004. Areas of endemism: improved optimality criteria. **Systematic Biology** 53:968-977.

Szumik, C., & Goloboff, P. 2015. Higher taxa and the identification of areas of endemism. **Cladistics** 31: 568–572