



DEPARTMENTS

Technological Tools

Note: Dr. David Inouye is the editor of the **Technological Tools** section. Anyone wishing to contribute articles or reviews to this section should contact him at the Department of Zoology, University of Maryland, College Park, MD 20742, E-mail: di5@umail.umd.edu.

DISTANCE 3.5

Research Unit for Wildlife Population Assessment, Mathematical Institute, University of St. Andrews, UK KY16 8PP. Freely available for download from the Internet at the *Distance* web site <http://www.ruwpa.st-and.ac.uk/distance/>

Distance sampling is a widely used group of related methods for estimating the density and/or abundance of biological populations. These methods include line transects, point transects (also called variable circular plots), trapping webs, and cue-counts. The methods have been used successfully in an incredibly diverse array of taxa, including shrubs and herbs, insects, amphibians, reptiles, birds, fish, and marine and land mammals. In all cases, the basic idea is the same. The ecologist performs a standardized survey along a series of lines or points, searching for “objects of interest” (i.e., individuals or clusters of indi-

viduals). For each object detected, they record the distance from the line or point to the object. Not all objects that the ecologist passes will be detected, but a fundamental assumption of the basic method is that all objects that are actually on the line or point are detected. Intuitively, one would expect that objects become harder to detect with increasing distance from the line or point, resulting in fewer detections with increasing distance. The key to distance sampling analyses is to fit a “detection function” to the observed distances, and use this fitted function to estimate the proportion of objects missed by the survey. From here it is a simple matter to calculate the number of objects present in the survey area. The methods are described in detail in the standard text, Buckland et al. 1993.

Distance sampling can be thought of as an extension of traditional “finite-population sampling” methods, (such as quadrat counts and strip counts) to the situation in which detection of all objects within the sampled area is not certain—a very common situation. For example, many songbird surveys rely on “point count” methods, in which observers stand at a point and count the number of birds within a given radius that are detected in a set time period. Often not all of the birds present are detected, so the method is usually taken to provide an index of abundance, rather than actual abundance. However, this is clearly not appropriate if

the goal of the study is to compare abundance between areas or time periods in which bird detectability differs (say due to habitat differences). It would then be impossible to tell whether observed differences in the index were due to differences in abundance or differences in detectability. Distance sampling allows the ecologist to estimate detectability in the different areas or time periods and so correct for the potential bias.

The standard software for analysis of distance sampling data is the program *Distance*. This software allows the user to fit detection functions interactively to distance sampling data and assess the fit of the functions using various diagnostic tools, including plots of the fitted functions superimposed on histograms of the data. Other useful features include the ability to deal with stratified surveys, to test for “size bias” in surveys in which the objects of interest are clusters of individuals, and to provide bootstrap estimates of variance.

Previous versions of *Distance* (1.0–3.0) were DOS-based programs, driven by a SAS-like command language. Those familiar with programming computers found the language relatively easy to master, but most ecologists struggled. Now, a new version of *Distance* is available (version 3.5) that is completely Windows-based, with an intuitive point-and-click graphical user interface.

Distance 3.5 is based around “projects,” which are database files

containing all of the data and analyses for one survey or series of surveys. Interactive “wizards” guide the user through the process of setting up a new project, and entering or importing the data. The main analysis interface is a live summary table called the “Analysis Browser,” which allows users to view and compare results from multiple analyses, and is the starting point for creating and running new analyses. Detailed information about analyses can be obtained via the “Analysis Details” windows, which can be launched from the Analysis Browser. Analysis Details windows are also the place where the analysis options (such as which detection function to use) are chosen, via two tabbed properties boxes.

In addition to being easier to learn and use, the new version of *Distance* offers a number of new features:

- Data import from text files in “flat file” format, enabling easy interface with other spreadsheet and database packages
- Export of data and analysis results via the Windows clipboard, including ability to copy high-resolution detection function plots for use in word-processing packages
- The ability to select different subsets of the data for analysis, for example, enabling large multiyear studies to be kept in one project file
- An extension of the stratification options to allow on-the-fly post stratification of data (for example, by sex or species)
- Multipliers, which offer a flexible way to scale the density estimate, e.g., to account for indirect counts (counts of nests or animal dung)
- The ability to run multiple analyses at once in the background; useful for large bootstrap analyses
- Faster analyses and fewer restrictions on data size, due to a new 32-bit analysis engine

To ease the process of upgrading from previous versions, *Distance 3.5* can import *Distance 3.0* and earlier command files. The new program also comes with an extensive on-line

manual and two sample projects that demonstrate various features of the software.

The only drawback to the new version is that it is relatively computer-resource hungry. Minimum requirements are for a 486DX processor with 16Mb RAM running Windows 95/98 or NT4. However, for satisfactory performance a Pentium processor with 32Mb RAM is recommended.

The new version is part of an ongoing project at the University of St. Andrews (jointly with the National Marine Mammal Laboratory in Seattle and the Colorado Cooperative Fish and Wildlife Unit in Fort Collins) to develop new methods and software for distance sampling surveys. The next version of *Distance* (version 4) is due for release in fall 2000 and will include several major advances in distance sampling theory, including geographic survey design (using a built-in GIS), spatial estimation of abundance, ability to deal with multiplatform surveys (for cases in which probability of detection on the line is <1), and ability to incorporate covariables that affect detectability into the analysis.

To stay up to date with the latest developments, there is a distance sampling e-mail listserver, details of which are given on the *Distance* web site.

Literature cited

Distance sampling book:

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. *Distance sampling: estimating abundance of biological populations*. Chapman and Hall, London, UK.

This book is currently out of print, but we hope to make it freely available over the Internet at the *Distance* web site in the first quarter of 1999.

Suggested citation for *Distance 3.5*:

Thomas, L., J. L. Laake, J. F. Derry, S. T. Buckland, D. L. Borchers, D. R. Anderson, K. P. Burnham, S.

Strindberg, S. L. Hedley, M. L. Burt, F. Marques, J. H. Pollard, and R. M. Fewster. 1998. *Distance 3.5*. Research Unit for Wildlife Population Assessment, University of St. Andrews, St. Andrews, UK.

Len Thomas
Research Unit for Wildlife
Population Assessment
Mathematical Institute
University of St Andrews
St Andrews, UK
len@mcs.st-and.ac.uk

CANOCO 4

Software for Canonical Community Ordination. Cajo ter Braak and Petr Šmilauer. Available from Microcomputer Power. Attn: Richard E. Furnas, 113 Clover Lane, Department W1, Ithaca, NY 14850-4930. Fax (607)-272-0782; Phone (607)-272-2188. Educational license \$249. <http://ww1.microcomputerpower.com/webpages/mcp/>

This package was tested in a Micron Pentium with a 300 Mhz CPU, 128 Mb RAM, and ample disk space. The package runs on Windows 95 or Windows NT. My machine has Windows NT. It will run in Windows 95 with 8 Mb available RAM and 32 Mb of hard drive, according to the manual.

This package is the latest in the series of CANOCO packages developed by ter Braak et al. As the title implies, this is a package with a choice of ordination methods, including principal components analysis, biplots (Gower and Hand 1996), correspondence analysis (Greenacre and Hastie 1987), and ter Braak’s canonical correspondence analysis (ter Braak 1986). The package is bundled with routines for drawing graphical displays (i.e., CANODRAW 3.1) and routines for editing the graphs to make them suitable for publication (CANOPOST 1.0). The prime audience for this package in-

cludes community and landscape ecologists interested in examining species distributions across habitats/sites and species distributions in relation to environmental variables. However, other scientists with multivariate data may also find this package useful.

Since many ecologists may not be familiar with the techniques provided in this package, two useful references were available from the distributor (Jongman et al. 1995 and ter Braak et al. 1996). The first is a book on landscape and community ecology emphasizing a number of methods in the package; the second is a collection of articles by ter Braak et al. As a statistician, I found the latter particularly interesting, since it provided motivation for many of the methods in CANOCO and insight into the assumptions being made. However, these articles would be stiff reading for the graduate students in my classes, i.e., first- and second-year graduate ecology students.

SAS provides procedures for correspondence analysis (CA) and principal components analysis (PCA), but it does not provide an algorithm for many of the extensions developed for use in community ecology such as detrended correspondence analysis (DCA). DCA avoids some of the weaknesses in correspondence analyses when displaying data consisting of counts of species within each of several habitats (Hill and Gauch 1980). CANOCO provides an extension to several ordination methods for compositional data, such as relative abundance measures that add to 100%. It is based on Aitchison's log-ratio model (Aitchison 1986). CANOCO output allows one to visualize the relationship of species distributions to environmental variables (e.g., Canonical Correspondence Analysis (ter Braak 1986, Palmer 1993).

In contrast to the previous version of CANOCO, the new interface is much more user friendly. Many features can be invoked by pointing and clicking with a mouse. Data can be imported from a spreadsheet program such as Excel or a database program

such as Access, with little effort. However, data can be input pretty much the same way as in the old version, with FORTRAN formats and data in aligned columns. An on-line tutorial is included to get users started.

Two interesting features include the ability to add a second data set to the plots even though they are not used to construct the axes, and permutation tests that offer nonparametric alternatives that provide tests of significance in the contexts of testing explanatory variables in methods such as canonical correspondence analysis (see Manly 1997).

The biggest problem I had was with the mechanism for generating hard copy. When I ran a principal components analysis, I had no problem getting a listing of the eigenvalues and the proportion of total variation explained. This was generated by the CANOCO package. I might also have liked to see the eigenvectors, but that is my personal bias. However, to get a plot of the principal component scores for each data point required the CANODRAW package. Again, this was not a problem since one can click on a button in CANOCO and get directly into the CANODRAW program. Using CANODRAW, I could view the plot on the screen of my PC and could then save the plot to a file. However, this package was designed with a limited number of printer drivers in mind. I have an Epson Stylus Color 800 inkjet printer and could not get it to print. By saving the graph to a file and copying it to an HP LaserJet 4SiMx in my building, I got some very nice hard copy. I then ran the CANOPOST program and was able to clip a picture into a Microsoft Word document and get hard copy at my desk. It would have been easier if the CANOPOST and CANODRAW programs were better integrated.

In short, if you believe that ordination methods will help you with the interpretation of your data, then you should give this package a serious look. If you own an older version of CANOCO, I believe the additional fea-

tures added in this version make consideration of an upgrade worthwhile. If you are interested in the utility of this package and other related ones, you may want to check out the following web sites for information: <http://www.okstate.edu/artsci/botany/ordinate> and <http://www.cpro.dlo.nl/cbw/canoco/>.

Literature cited

- Aitchison, J. 1986. The analysis of compositional data. Chapman and Hall, London, UK.
- Gower, J., and D. J. Hand. 1996. Biplots. Chapman and Hall, London, UK.
- Greenacre, M., and T. Hastie. 1987. The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association* **82**:437–447.
- Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis, an improved ordination technique. *Vegetatio* **42**:47–58.
- Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge, UK.
- Manly, B. 1990. Randomization tests and Monte Carlo methods in biology. Chapman and Hall, London, UK.
- Palmer, M. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* **74**:2215–2230.
- ter Braak, C. J. F. 1996. Unimodal models to relate species to environment. DLO-Agricultural Mathematics Group, Wageningen, The Netherlands.
- . 1986. Canonical correspondence analysis: a new eigenvector method for multivariate direct gradient analysis. *Ecology* **67**:1167–1179.

*Dr. Estelle Russek-Cohen
Department of Animal
and Avian Sciences
University of Maryland
College Park, MD 20742
E-mail: Ansc2@umdd.umd.edu*