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.

STATISTICA NEURAL NETWORKS, VERSION 4.0

STATISTICA Neural Networks, Version 4.0; StatSoft, Incorporated, 2300 East 14th Street, Tulsa, OK 74104 USA; (918) 749-1119; Fax: (918) 749-2217; E-mail: info@statsoft.com; <http://www.statsoft.com>; Single-user license $1495.00; academic pricing available.

Neural network modeling is becoming a popular alternative to standard statistical analysis in the ecological sciences. Neural networks model the relationship between one or more input variables and a continuous, ordinal, or categorical output variable. The network is "trained" on a known data set, and the resulting model can be applied to novel data where the output is unknown. Neural network analysis is ideal for forecasting and classification problems where data sets are noisy and relationships are nonlinear, like many problems in ecology.

STATISTICA's Neural Networks (SNN) module is a "point-and-click," full-featured neural modeling package that can be used as an add-on to the STATISTICA analysis package, or as a stand-alone product. It runs on Windows 95/98/NT/2000/XP operating systems. The SNN program will run on almost any Windows-based PC, but an Intel Pentium system with a minimum of 16 MB of RAM is recommended for optimal performance. We tested the software on an Intel Celeron 465 Mhz processor with 192 MB of RAM, running Windows 98 SE.

Manuals include a concise introduction to neural networks and a series of tutorials that are helpful for users unfamiliar with the methodology. Background information is extensive, and ecologists will appreciate the references to primary literature.

Of the three neural network modeling packages we have used (including Neural Connections from SPSS and QNet 2000 from Vesta Services), SNN is by far the best, in terms of both ease of use and power. In fact, the only cumbersome aspect of the program is data management. SNN uses a proprietary file format, and input files stored in other formats (such as MS Excel) must be saved as comma-separated values (*.csv) files before they can be imported into SNN. This is an odd omission, because the main STATISTICA program contains conversion filters for many different formats, including MS Excel. However, this is not a problem unique to SNN; similar data-management problems exist in other neural network programs.

SNN's interface is more intuitive than other packages and allows users to view simultaneously data, network predictions/output, and performance statistics. The level of control that the user has over aspects of model development can be varied depending on his or her familiarity with neural network modeling. Beginners can make use of the *Intelligent Problem Solver*, which automates many aspects of model development. By asking simple questions, the *Solver* selects appropriate settings and parameter values for the training process. Seasoned users can take more control of model development and can customize each step of the process to suit their modeling needs. Experienced users can also make use of the *Intelligent Problem Solver* by selecting the *advanced* option in the dialogue window. We have used both the *Intelligent Problem Solver* and the more manual approach, and found little difference in performance.

Before analysis, input data must be apportioned to training, validation, and testing sets. This is relatively straightforward in SNN and can be handled automatically using the *Intelligent Problem Solver*. Training data are used to develop the neural network model, and the test and validation data are used to monitor the network's performance and goodness-of-fit, respectively.

The SNN program can generate a variety of neural network models, including multilayered perceptrons, radial basis functions, Kohonen networks, and generalized regression neural networks, among others. SNN allows the user to develop simultaneously models of different types, which is useful if the user is unsure of which model is most appropriate for his or her application. The program will train all possible networks within user-specified limits and will retain those models that have the best performance. Users can select the best model from the set or can compare results generated by different network types, including linear models. In addition to the wide variety of neural models available in SNN, there are also several available training algorithms. These include the common backpropagation method as well as the more efficient conjugate gradient descent.

SNN has a utility that makes use of another type of machine learning: genetic algorithms. SNN can test various combinations of variables against each other to evolve an optimal set for subsequent neural network development. The genetic algorithm also ranks variables in the optimal subset, indicating which variables are most important. This feature is useful for removing noisy, information-poor variables from large data sets prior to further analysis. Neither of the other two neural network packages mentioned above has this capability.

An additional module ($495) can convert the network output generated by SNN into C or C++ code that can be used in custom applications. This add-on is probably of more interest to engineers than to ecologists, but there are situations in which researchers may want to distribute a stand-alone program for use in certain management contexts.

Using a relatively simple data set (n = 100, with 10 independent variables), SNN trained a series of networks in under 5 minutes. Larger, more complex data sets will require longer training periods depending on how many network types the user has specified and the sample size. We developed models to predict continuous variables and for classification problems. The networks developed with SNN for predicting continuous variables produced qualitatively similar results to those generated by the other neural modeling applications with which we are familiar. Because the weights generated during training are not directly interpretable, SNN includes a simulation tool for investigating the relationships between independent and dependent variables. From a single dialogue window, users specify the variables to be investigated and the range over which they are to vary for the simulation analysis. The results are then displayed at the bottom of the window, and can be saved or exported to other applications. Because simulations are an important part of understanding the biology behind a neural network's predictions, this tool is an essential component. The simulation tool in SNN is much more intuitive and easy to use than a similar tool in SPSS's Neural Connections, and QNet 2000 does not provide a simulation tool at all.

SNN will generate classification networks if it detects a categorical dependent variable in the training data set. No additional input is required from the user, and model building is again very fast and efficient. SNN's default output activations cannot be directly interpreted as probabilities of class membership in problems involving more than two classes. The manual provides instruction on making the necessary changes to transform activations to probabilities, but it is not clear why this would not be the default option. Diagnostic output is very complete, and includes a sensitivity analysis to help determine the importance of different input variables to the final model.

SNN is an excellent package for both new and experienced users of neural networks. It is the most intuitive and easiest to use of the three packages mentioned in this review. Unfortunately, it is also the most expensive; Statsoft recently increased the price of a single-user license from $795 to $1495. For those who frequently use this type of analysis, the time and energy saved may well be worth the steep price; however, for occasional users, Statsoft's pricing policy has pushed the software out of reach.

We recommend SNN for those who are considering using neural networks in their research and are looking for a powerful, easy-to-use package, or for experienced users who are looking for an efficient and complete neural network modeling package.

Reviewed by Jeffrey J. Lusk
Department of Forestry
Oklahoma State University
Stillwater, OK 74078
E-mail: luskj@okstate.edu

Steven F. Wilson
EcoLogic Research
Gabriola Island, British Columbia
Canada V0R 1X1
E-mail: sfwilson@shaw.ca

THE CRITICAL VALUES PROGRAM FOR ASSESSING EDGE INFLUENCE

Many studies in fragmented and harvested landscapes have focused on the boundary between forested and nonforested ecosystems, and the influence of the edge environment on the adjacent forest (e.g., Chen et al. 1992, Laurance et al. 1998). Research has also begun on trends along edge-to-interior gradients at natural edges (e.g., lakeshore forest edges [Harper and Macdonald 2001]). To assess the effects of edges on forest structure, species composition, biodiversity, and other ecosystem attributes, there is a need to identify the area of the forested ecosystem that experiences a significant influence of the adjacent nonforested environment. Most researchers approach this by quantifying the distance of edge influence (DEI) or edge width. DEI quantifies the distance from the nonforested ecosystem into the forest, perpendicular to the edge, that is significantly different from interior forest. Such estimates of DEI are useful in predicting spatial patterns of edge influence in fragmented landscapes (e.g., Chen et al. 1996).

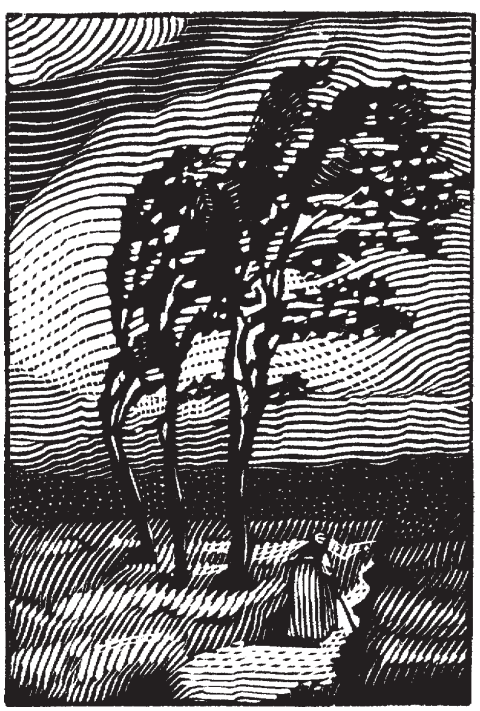
Although researchers continue to use a variety of methods for measuring DEI, there has been some convergence of methodologies (e.g., Chen et al. 1992, Laurance et al. 1998, Harper and Macdonald 2001). The critical values approach (Harper and Macdonald 2001) combines elements of several previously employed methods; it compares average values of response variables at different distances along an edge-to-interior gradient to critical values of reference (interior) forest conditions determined using randomization tests. As a response to a request for consistent methodology in order to compare edge influence from different studies (Murcia 1995), we have developed the Critical Values Program and are making it available for all researchers.

The Critical Values Program is the first program specifically designed to quantify DEI, and can be downloaded free from <http://www.rr.ualberta.ca/staff/emacdona/ellen.htm>. This program uses the critical values approach and can: (1) calculate DEI for one time period at one edge type (Harper and Macdonald 2001), (2) quantify changes in edge influence between two years at one edge type

(K. A. Harper and S. E. Macdonald, *unpublished manuscript*), and (3) assess differences between two different edge types (Harper et al. 2001). It is user friendly and it is written in Visual Basic in Excel (Microsoft Corporation, Seattle, Washington, 1997); instructions are provided in a Word document. The program will be updated as new methods are developed based on the critical values approach (e.g., for comparing more than two edge types).

Data that are appropriate for this analysis are generally from plots along transects established perpendicular to the edge, with at least three plots (or two plots with several subplots) designated as the "reference forest." The program can treat sampling designs with any number of transects, plots, or subplots. The program is also flexible with respect to the number of permutations in the randomization tests, and the significance level. Another advantage of this program is that numerous variables can be analyzed at the same time. Results are provided for each variable in the form of critical values of reference forest conditions, and averages or t values (when comparing two time periods or two edge types) for different distances from the edge; these averages or t values are highlighted if they are above or below the critical values, and therefore significantly different from the reference forest. Harper and Macdonald (2001) define the distance of edge influence as the set of two or more consecutive distances, with values that are significantly different from the reference forest.

The Critical Values Program could also be used for other applications in which a response to a given treatment, or at a specific location, is compared to the range of variation in reference sites where multiple samples are taken. For example, responses to gaps could be compared to reference forest conditions, or the responses to different types of gaps could be compared. This is a useful way to assess ecological significance by comparing the effect of the treatment to the range of variation in the reference condition. We hope that the Critical Values Program will be useful to researchers studying edge influence and other phenomena as a means of distinguishing the ecological significance from background heterogeneity.



Literature cited

Chen, J., J. F. Franklin, and J. S. Lowe. 1996. Comparison of abiotic and structurally defined patch patterns in a hypothetical forest landscape. Conservation Biology **10**:854–862.

Chen, J., J. F. Franklin, and T. A. Spies. 1992. Vegetation responses to edge environments in old-growth Douglas-fir forests. Ecological Applications **2**:387–396.

Harper, K. A., Y. Bergeron, P. Drapeau, and D. Lesieur. 2001. The structure of fire and clearcut edges in black spruce boreal forest. Ecological Society of America, 86th Annual Meeting Abstracts: 108.

Harper, K. A., and S. E. Macdonald. 2001. Structure and composition of riparian boreal forest: new methods for analyzing edge influence. Ecology **82**:649–659.

Laurance, W. F., L. V. Ferreira, J. M. Rankin-de Merona, and S. G. Laurance. 1998. Rain forest fragmentation and the dynamics of Amazonian tree communities. Ecology **79**:2032–2040.

Murcia, C. 1995. Edge effects in fragmented forests: implications for conservation. Trends in Ecology and Evolution **10**:58–62.

Karen A. Harper
Groupe de recherche en écologie forestière interuniversitaire
Université du Québec à Montréal
CP 8888 succ. A
Montréal, Québec,
Canada H3C 3P8
(514) 987-3000, Ext. 4723
E-mail: c1444@er.uqam.ca

S. Ellen Macdonald
Department of Renewable Resources
University of Alberta
Edmonton, Alberta,
Canada T6G 2E3
(780) 492-3070
E-mail: ellen.macdonald@ualberta.ca