The Bootstrapped Response Function

RESEARCH REPORT

THE BOOTSTRAPPED RESPONSE FUNCTION

J. GUIOT

Laboratoire de Botanique Historique & Palynologie, UA CNRS
1152, F-13397 Marseille cedex 13

ABSTRACT

The bootstrap procedure provides a way to test the significance of the regression coefficients and the stability of the estimates in response functions generated by regression on principal components. A subroutine RESBO, which calculates a bootstrapped response function, has been added to Fritts' program PRECON.

The principle of the response function is described in Fritts (1976) and discussed in Hughes, et al. (1982). To avoid problems with the great number of predictors and their intercorrelation, Fritts et al. (1971) introduced regression on principal components. As with all regression methods, the main problems with this procedure are testing the significance of the coefficients and the stability of the estimates. The response function obtained on a sample is considered satisfactory only if it explains the growth over independent years. The most straightforward way to assess the stability is to divide climatic and tree-ring data into a dependent calibration set and an independent verification set (Fritts 1976). If the set of tree-ring indices estimated from the verification-set climate data using the regression coefficients that were derived from the calibration data set is close to the observed values, the response function is judged as reliable.

Gordon et al. (1982) clearly set out the problem of verifying the predictive ability of a model calibrated on one data set when applied to another data set. Because regression coefficients are validated only to the dependent data, they result in overconfidence in the predictive power of the model. We can be convinced of that by simulating tree-ring indices by random numbers and by calculating response functions with real climatic data (Guiot 1981; Cropper 1985). These authors showed that simulated tree-ring series also can produce regression coefficients judged significant by standard Student's tests. This result is due mainly to an inadequate number of degrees of freedom. To test regression coefficients, Student's test involves $n-k-1$ degrees of freedom where n is the number of observations and k the number of regressors. If k is set to the number of principal components actually introduced into the regression on the basis of their correlation with the predictand (stepwise regression), the significance of the coefficients is overestimated; therefore, the number k must be chosen by *a priori* considerations independent of the predictand. A good practice is to select a relatively large number of principal components taking into account say 90 or 95% of the variance of the climatic data - or using the PVP criterion of Guiot (1981, 1985). The number k is then the number of principal components selected by such an *a priori* criterion.

The bootstrap procedure of Efron (1979) provides an interesting method to simultaneously test the regression coefficients and the stability of the response function. It has been applied to tree-ring data by Guiot (1990a) and Till and Guiot (1990). The idea is to replace the lack of information on the statistical properties of the data by a great number of estimates, each based on different subsamples of data. The comparison of these estimates shows the variability of the estimates. The subsampling is done by random extraction with replacement from the initial data set. The size of each subsample is the same as that of the initial data set (n) to avoid bias (Efron 1983). Each subsample forms a bootstrap test useful for cross-validation. Guiot (1990a) has shown that beyond 50 subsamples, the results do not change significantly, even if more than 1000 replications usually are recommended.

For each subsample, the regression coefficients and the multiple correlation are computed on the observations randomly selected (some observations of the initial data set are used repeatedly while others are omitted). An independent verification is done on the observations omitted from the subsample. Repeated 50 times, this procedure yields 50 sets of regression coefficients, 50 multiple correlations, and 50 independent verification correlations. A mean regression coefficients set with standard deviations is computed on these 50 estimates. Means and standard deviations are also computed for the multiple correlation and the independent correlation sets. The bootstrapped regression coefficients are judged significant at the 95% level if they are twice, in absolute value, their standard deviation (see Fritts and Dean 1991). A more precise method is to compute the interval between the 2.5th and the 97.5th percentile, which equals the 95% confidence level.

We also can predict a tree-ring index from climate using these 50 response functions accompanied with a confidence interval. This feature can be used to measure the reliability of the results and to improve simulation studies. For example, in the spirit of the paper of Fritts and Dean (1993), the amount of precipitation that must be added to or substracted from the real precipitation series to produce a significant growth shift can be easily quantified.

A subroutine RESBO, which computes a bootstrapped response function, has been added to the program PRECON of H. Fritts. In this interactive program, the user can monitor in real time the step by step evolution of the multiple correlations and stop the program when no significant change is detected. A complete package of statistical programs, including bootstrapped regression (but also time series analyses, multivariate analyses, and transfer functions) with a users' guide of 250 pages (Guiot 1990b) is available from the author.

REFERENCES

Cropper, J. P.
1985 *Tree-Ring Response Functions: An Evaluation by Means of Simulations.* Ph.D dissertation, The University of Arizona, Tucson. University Microfilms International, Ann Arbor.

Efron, B.
1979 Bootstrap methods; another look at the jackknife. *The Annals of Statistics* 7:1-26.

Fritts, H. C.
1976 *Tree Rings and Climate.* Academic Press, London.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach
1971 Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *Journal of Applied Meteorology* 10: 845-864.

Fritts, H. C., and J. S. Dean
1991 Dendrochronological modeling of the effects of climatic change on tree-ring width chronologies from Chaco Canyon and environs. *Tree-Ring Bulletin* (submitted).

Gordon, G. A., B. M. Gray, and J. R. Pilcher

1982 Verification of dendroclimatic reconstructions. In *Climate from Tree-Rings*, edited by M. K. Hughes, P. M. Kelly, J. R. Pilcher, and V. C. LaMarche, Jr., pp. 58-62. Cambridge University Press, Cambridge.

Guiot, J.

1981 *Analyse Mathématique de Données Géophysiques, Application à la Dendroclimatologie*. Ph.D dissertation, Institut d'Astronomie et Géophysique, Louvain-la-Neuve.

1985 The extrapolation of recent climatological series with spectral canonical regression. *Journal of Climatology* 5:325-335.

1990a Methods of calibration. In *Methods of Dendrochronology: Application to Environmental Sciences*, edited by E. Cook and L. Kairiukstis, pp. 165-178. Kluwer Academic Press and IIASA, Dordrecht.

1990b Methods and programs of statistics for paleoclimatology and paleoecology. In "Quantification des Changements Climatiques: Méthodes et programmes", edited by J. Guiot and Labeyrie, pp. *Monographie* 1, INSU, Paris.

Hughes, M. K., P. M. Kelly, J. R. Pilcher, and V. C. LaMarche, Jr., Editors

1982 *Climate from Tree-Rings*. Cambridge University Press, Cambridge.

Till, C., and J. Guiot

1990 Reconstruction of precipitation in Morocco since AD 1100 based on Cedrus atlantica tree-ring widths. *Quaternary Research* 33:337-351.