TREE-RING RESEARCH, *Vol. 66(2), 2010, pp. 135–139*

RESEARCH REPORT

EVALUATION OF GOODNESS-OF-FIT STATISTICS FROM PRECON TO ESTIMATE THE STRENGTH OF MULTIVARIATE TREE GROWTH-CLIMATE ASSOCIATIONS

DAVID C. LEBLANC

Dept. of Biology, Ball State University, Muncie, IN 47306, USA

ABSTRACT

Although the primary purpose of response function analysis is to identify climate variables that have significant associations with tree radial growth, many researchers are also interested in assessing the strength of these associations. Existing response function programs use a liberal criterion to determine how many climate variables should be included in the analysis. The resulting response function models include a large number of predictor variables. The objective of this analysis is to determine if these response function models are over-fitted to the data used to calibrate them, resulting in over-estimation of strength of associations. PRECON was used to produce response functions for white oak chronologies from n = 149 sites, with separate response functions using 34 monthly climate variables or 10 seasonal climate variables. An analysis of goodness-of-fit statistics for response function calibration provided strong evidence of over-estimation of strength of associations. The degree of over-estimation was greater when 34 monthly climate variables were included in the models compared to models with10 season variables. There was much less evidence of over-fitting for the R-verif statistic that reflects strength of association between predicted and actual tree-ring indices that were not included in model calibration. The PRECON R-verif statistic is the best measure of the strength of multivariate growth-climate associations currently available.

Keywords: response function analysis, *Quercus alba*, white oak, dendroclimatology.

INTRODUCTION

Statistical analyses used to evaluate associations between tree radial growth indices and climate variables fall into three categories: (1) correlation functions that display bivariate associations between growth and climate variables, (2) multiple regression analyses between growth and climate variables, and (3) bootstrapped multivariate response functions between growth and climate variables. Bivariate correlation analysis and standard multiple regression analyses do not account for multicollinearity among climate predictor variables and are invalid when data are not normally distributed. Bootstrapped response function analysis, as implemented by both PRECON (Fritts 1976; Guiot 1991) and DENDROCLIM2002 (Biondi and Waikul 2004), addresses the multicollinearity problem by performing regression with orthogonal principal components of the original climate variables. The bootstrap approach to significance testing calculates empirical 95% confidence intervals for estimated regression coefficients associated with each climate variable, eliminating the need to assume the variables are normally distributed. If the confidence interval for the mean (PRECON) or median (DENDROCLIM2002) of a bootstrapped estimate of a regression coefficient does not include zero, the association between growth and that climate variable is deemed statistically significant. However, the bootstrapped regression models computed by both PRECON and DENDROCLIM2002 include most climate principal components (those principal components that account for 90% to 95% of total variance in the climate data). Including many candidate climate predictor variables increases the risk that the model will be over-fitted to the specific data set and will not generalize well to independent data.

Table 1. Goodness-of-fit statistics produced by PRECON.

Fit Statistic	Definition
RSQ	Proportion of variance in radial growth indices explained by the overall (climate + prior growth) response function model.
RSQ CL	Proportion of variance in radial growth indices explained by only the climate variables in the response function model.
RSQ GR	Proportion of variance in radial growth indices explained by only the prior growth variables in the response function model.
R-calib	Correlation between predicted and actual growth indices for those years of data randomly selected for inclusion in the bootstrap repetition and used to calibrate the model coefficients. Calculated as a running average of all bootstrap repetitions.
R-verif	Correlation between predicted and actual growth for those years of independent data NOT selected for inclusion in each repetition of the bootstrapped calibration of model coefficients. Calculated as a running average of all bootstrap repetitions.

Coefficients of determination (R^2) and multiple correlation coefficients (multiple r) are used to evaluate the strength of multivariate associations. However, the calculation and interpretation of such statistics for bootstrapped response function models is problematic. For this reason, DENDROCLIM2002 does not report goodness-of-fit statistics, but only lists which climate variables are significantly associated with radial growth (Biondi, pers. comm.). However, PRECON reports several such statistics (Table 1).

The PRECON multiple correlation statistics (R-calib and R-verif) are computed for each repetition of the bootstrap process, and the tests of significance for these correlations are based on bootstrapped 95% confidence intervals. Because R-verif reflects the ability of the response function model to predict tree-ring indices using data not included in the calibration of the regression coefficients, this statistic should not be affected by over-fitting the model to the calibration data. The coefficients of determination calculated by PRECON are based on actual and predicted ring-index values computed with the final response function model that uses the final mean bootstrapped regression coefficients for candidate predictor variables. All climate and lagged growth (if any) predictor variables are included in the response function, regardless of whether or not their associated regression coefficients are significantly different from zero. When many candidate climate predictor variables are included in the response function analysis, this process is likely to produce predicted values that are over-fit to the calibration data set and these goodness-of-fit statistics are likely to over-estimate the actual strength of growth-climate associations. The coefficients of determination reported by PRECON are not adjusted for the degrees of freedom associated with the model (*i.e.* the number of predictor variables).

The objective of this study was to evaluate the validity of RSQ, RSQ CL, R-calib and R-verif statistics reported by PRECON for assessing strength of multivariate tree growth-climate associations. If the inclusion of a large number of climate predictor variables in the bootstrapped response function analysis results in over-fitting of the model to the data, I predicted that (1) values for R-calib should be substantially larger than values for R-verif for the same response function model, and RSQ and RSQ CL should be substantially greater than R-verif squared, (2) these discrepancies between R-verif and other PRECON goodness-of-fit statistics should be reduced when fewer candidate climate predictor variables are included in the response function analysis, and (3) R-calib, RSQ and RSQ CL should be substantially greater than zero even when there are no significant regression coefficients in the model. If there are no significant regression coefficients in a response function, all of these goodness-of-fit statistics should have values close to zero.

METHODS

PRECON was used to analyze growth-climate associations for n = 149 white oak (*Quercus alba* L.) chronologies from sites distributed throughout eastern North America, from Texas to New York and Iowa to Florida. This database included both

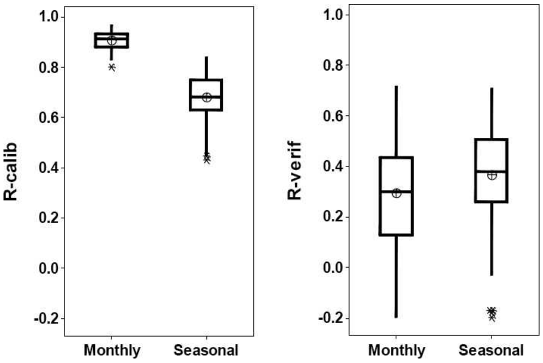

Figure 1. R-calib *vs.* R-verif for PRECON response functions using 34 monthly climate variables *vs.* 10 seasonal climate variables, as performed for n = 149 white oak sites.

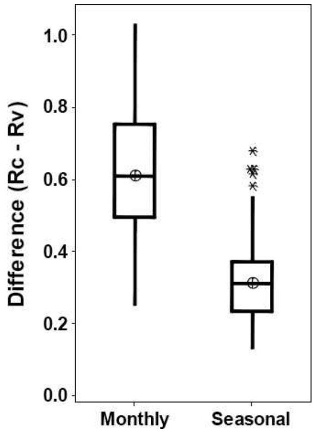

Figure 2. Difference between R-calibration and R-verification (Rc−Rv) for PRECON response functions using 34 monthly climate variables *vs.* 10 seasonal climate variables (n = 149 sites).

climate-sensitive chronologies from prairie border sites in Iowa and Illinois and chronologies from warm, mesic sites in the southeastern U.S. that had no significant correlations with any climate variables. By summarizing the results of many PRECON analyses for a wide range of growth-climate associations I believed that consistent patterns would reflect the performance of the program rather than idiosyncrasies of any particular site. All chronologies were crossdated, standardized to white noise residuals using ARSTAN, and truncated to a common time period from 1930 to 1980. Two PRECON analyses with 999 repetitions each were performed for each site using (1) monthly mean maximum temperature and precipitation data for prior year June through current year October, for a total of 34 climate variables, and (2) seasonal mean maximum temperature and precipitation for prior Summer (June, July, August), prior Autumn (September, October, November), prior Winter (December, January, February), current year Spring (March, April) and current early growing season (May, June, July), for a total of 10 climate variables. R-calib and R-verif, as reported for the final bootstrap repetition, as well as final model RSQ and RSQ CL, were compared using paired t-tests (by site) to determine if the predicted results of model over-fitting described in the Introduction were manifested.

RESULTS

R-calib for PRECON models using 34 monthly climate variables was greater than R-calib for models using 10 seasonal variables (Figure 1, mean difference = 0.22, $p < 0.001$). However, R-verif for models using monthly climate variables was less than R-verif for models using seasonal variables (Figure 1, mean difference = −0.07, $p < 0.001$), indicating that response functions using fewer seasonal variables had slightly better predictive ability with independent data than response functions using many monthly variables.

The mean difference between R-calib and R-verif (Rc – Rv) for the same site and response function was greater than zero for both monthly and seasonal climate models (Figure 2, one-sample t-test, $p < 0.001$). The difference (Rc−Rv) was larger for models that included a large number of monthly climate variables compared to models that included fewer seasonal climate variables ($p < 0.001$).

For those PRECON models that did not have any significant regression coefficients (*i.e.* the overall model was not significant) mean R-calib was much greater than zero for both monthly and seasonal models (Figure 3, one-sample t-test, $p < 0.001$). All R-calib values from 298 PRECON runs (monthly and seasonal models combined) were significant, based on the fact that their bootstrapped 95% confidence interval did not include zero, including the 74 models that had no significant regression coefficients. Mean R-verif was also slightly greater than zero for non-

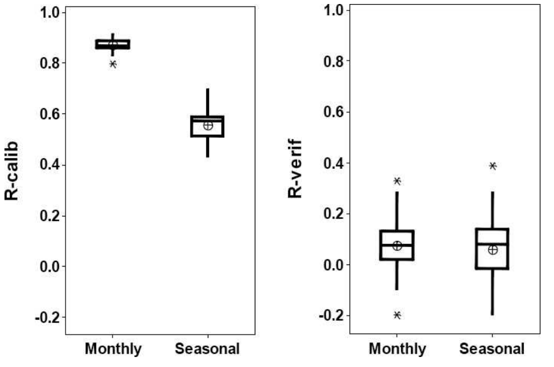

Figure 3. R-calib *vs.* R-verif for PRECON response function models that had no significant regression coefficients (n = 50 and n = 24 sites for monthly and seasonal variable models, respectively).

significant PRECON models (Figure 3), including both monthly climate models ($p < 0.001$) and seasonal climate models ($p = 0.053$). However, in only 3 out of 298 PRECON runs used for this analysis was R-verif significant when there were no significant regression coefficients in the model.

The coefficients of determination reported for final PRECON response functions (RSQ and RSQ CL) had values that were similar to the squared value of R-calib (Rc-SQ, Figures 4 and 5). These three goodness-of-fit statistics were more inflated relative to the squared value of R-verif (Rv-SQ) when many monthly climate variables were entered in the analysis (Figure 4) than when fewer seasonal climate variables were entered (Fig. 5). Because R-calib is computed during the bootstrap process whereas RSQ and RSQ CL are based on the final response function model, it was expected that Rc-SQ would differ somewhat from RSQ and RSQ CL.

DISCUSSION

Results of this analysis indicated that all PRECON goodness-of-fit statistics except R-verif are inflated because of over-fitting of the model to the calibration data. The magnitude of this problem increased as the number of climate predictor variables entered into the response function analysis increased. Based on the analyses presented here, I believe that R-verif is a better measure of the strength of multivariate growth-

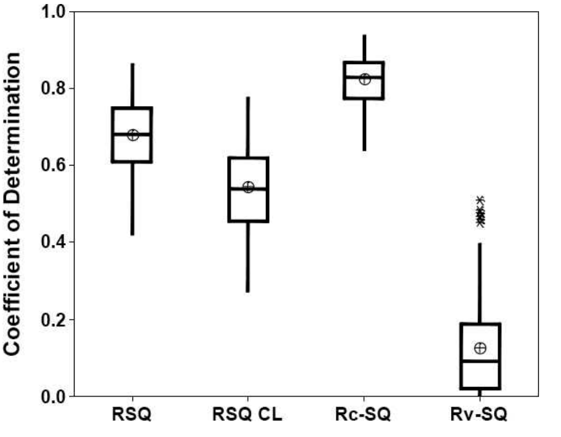

Figure 4. Comparison of goodness-of-fit statistics for response function models with 34 monthly climate variables (n = 149 sites). R-calib and R-verif were squared to make them comparable to RSQ and RSQ CL.

climate associations than the other PRECON goodness-of-fit statistics because (1) R-verif was much less influenced by the number of climate predictor variables entered into the model, as indicated by the small difference between R-verif values from models that included 10 *vs.* 34 climate variables, and (2) R-verif was very rarely significant (based on the bootstrapped 95% confidence interval) when there were no significant regression coefficients in the model. However, because R-verif reflects ability to predict independent data, this is a more conservative measure of strength of

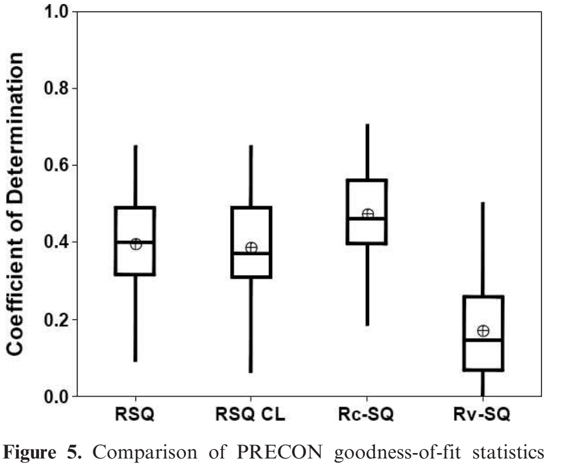

Figure 5. Comparison of PRECON goodness-of-fit statistics for response function models with 10 seasonal climate variables (n = 149 sites). R-calib and R-verif were squared to make them comparable to RSQ and RSQ CL.

association than standard correlation or R^2 statistics. R-verif is probably best used for comparing strength of growth-climate associations among multiple PRECON runs for different sites or different species. Given that values for R-verif were not much different between models with many monthly *versus* fewer seasonal climate variables, this statistic could also be used to compare response function models based on different types or numbers of climate predictor variables.

The "over-fitting" problem I have described here for PRECON is likely caused by the liberal criterion used to select which (and how many) climate principal components to include as predictors in the bootstrapped response function analysis. In PRECON runs that used 34 monthly climate variables, 20 to 23 climate principal components were retained in the response functions. In analyses that used 10 seasonal climate variables, 8 to 9 principal components were retained in the response functions. I believe this reflects an underlying philosophy by the authors of PRECON that the program should be used as an exploratory analysis tool for identifying climate variables that may influence tree growth. A similar philosophy and criterion for variable entry into response function analysis was used in DENDROCLIM2002 (Biondi and Waikul 2004). Nonetheless, when many climate predictor variables are included in response function analysis there is a high probability that chance coincident variation between some climate variables and tree-ring indices will result in spurious correlations that inflate goodness-of-fit statistics for calibration models. This problem can be mitigated if investigators use a more focused approach to studying tree growth-climate associations, including only those climate variables for which there is a substantiated biological reason for suspecting a cause-effect relationship with tree growth. Also, the use of seasonal climate variables in place of monthly variables can reduce the number of candidate predictors and consequent over-fitting of the model to the calibration data. The analyses described in this paper indicated that the use of fewer seasonal variables may actually increase the ability of resulting response functions to predict tree-ring indices for independent verification data.

How best to measure the strength of multivariate growth-climate associations remains an unresolved question. Many researchers are investigating tree growth-climate relationships to better understand forest responses to changing climate, including which tree species or trees growing on which site types are more or less strongly affected by climate variables. Comparison of the strength of growth-climate associations among species or sites is one approach for assessing sensitivity of trees and forests to changing climate. I believe that the question of how best to measure strength of multivariate growth-climate associations is worthy of a collaborative effort to develop a standard procedure that will be recognized as useful and valid. Until this question is formally resolved, I believe that the R-verif statistic computed by PRECON is the best measure of the strength of multivariate growth-climate associations currently available.

REFERENCES CITED

Biondi, F., and Waikul, K., 2004. DENDROCLIM2002: A C++ program for statistical calibration of climate signals in tree-ring chronologies. *Computers & Geosciences* 30:303–311.

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

Guiot, J., 1991. The bootstrapped response function. *Tree-Ring Bulletin* 51:39–41.

Received 9 January 2009; accepted 31 October 2009.