CALIBRATING DIATOM INDICATORS: HOW MANY SAMPLES ARE ENOUGH?

Euan D. Reavie[1] & Steve Juggins[2]

[1]Natural Resources Research Institute, University of Minnesota Duluth
[2]School of Geography, Politics and Sociology, Newcastle University

Diatom-based transfer functions have become popular indicators of aquatic condition for monitoring and paleolimnology programs. Assuming rigorous sampling of the environmental gradient, it is expected that training sets with more samples will be "better." I.e., they will provide better definition of environmental and diatom assemblage conditions in the region of interest, and as a result a model based on a larger sample set should provide more reliable inferences of condition. The substantial effort involved in developing a training set necessitates that an optimal sample size be estimated using model performance criteria.

Three large, diatom-based training sets were investigated to determine optimal sample sizes for inference models. The sample sets included (1) assemblages from Great Lakes coastlines, (2) phytoplankton from the pelagic Great Lakes and (3) surface sediment assemblages from Minnesota lakes. Diatom-based weighted average models to infer nutrient concentrations were developed for each training set. Training set sample sizes ranging from 10 to the maximum number of samples were created through random sample selection, and performance of each model was evaluated. For each model iteration, diatom-inferred (DI) nutrient data were related to stressor data (e.g., adjacent agricultural or urban development) to characterize the ability of each model to track human activities. The relationships between model performance parameters (DI-stressor correlations and model r^2, error and bias) and sample size were used to determine the minimum sample size needed to optimize models for each region. Depending on the training set, at least 40-80 samples were needed to capture the variation in diatom assemblages and environmental conditions to such a degree that non-analogue situations should be rare, and so should provide an unambiguous result if the model was applied to any sample assemblage from the region. It is recommended that one exercises caution when dealing with smaller training sets unless there is certainty that the selected samples reflect the regional variability in diatom assemblages and environmental conditions. Further, we advise that our findings for minimum required sample size may not necessarily extend to other regions and environmental variables. We encourage training set users to employ a similar evaluation to determine whether they have effectively sampled their region of interest.