Limnol. Oceanogr.: Methods 5, 2007

A statistical procedure for unsupervised classification of nutrient limitation bioassay experiments with natural phytoplankton communities

Tom Andersen[1], Tuomo M. Saloranta[2], and Timo Tamminen[3]
[1]Dept. Biology, University of Oslo, P.O. box 1066, Blindern, 0316 Oslo, Norway
[2]Norwegian Institute for Water Research, Gaustadalléeu 21, 0349 Oslo, Norway
[3]Finnish Environment Institute, Mechelininkatu 34a, P.O. box 140, FIN-00251, Helsinki, Finland

Web Appendix 1

Implementation details—Software for performing the statistical nutrient limitation classification procedure described here is distributed as two multimedia appendices, containing modifiable source code and a ready-to-run software application. The modifiable source code (multimedia appendix 1) requires access to and some familiarity with Matlab (http://www.mathworks.com). It consists of a function AIClim.m and a small script (AIClimdemo.m) demonstrating the use of the function. The users are expected to be fluent enough in Matlab to write their own scripts for importing and arranging input data before calling AIClim, and presenting output data returned by the function. This approach is clearly the most powerful for analyzing a large number of experiments and for experiments that do not conform to the ready-made solution below.

AIClim is called with the following input arguments, which, with the exception of factors, should be vectors and matrices with the same number of rows:

- *Block:* column vector with treatment replication block numbers coded as successive integers starting from 1. An error message results if less than 2 treatment replicates are encountered.
- *Time:* column vector with sampling times coded as successive integers with time 0 denoting the start of the experiment.
- *Treat*: two-column matrix coding the treatments as row values [0 0], [1 0], [0 1], or [1 1], corresponding to control, P alone, N alone, P and N combined for a P and N limitation assay. An error message results if any treatment combinations are missing.
- *Response:* column vector containing the measured response parameter. An error message results if missing values are encountered.
- *Factors:* optional 2-character string symbolizing the treatment factors, with 'PN' as the default value.

The output includes the following variables:

- *Pattern:* the limitation pattern of the model with the lowest AIC (the selected model), as a 2-character string code followed by the time effect order in parenthesis (e.g., XP(2) for exclusive P limitation with quadratic time effect).
- *Model:* a composite object (Matlab structure) containing the minimum AIC, the contrast and time effect matrices of the selected model, as well as the resulting design matrix. It also includes an index array, which makes the original response vector sorted according to replication block, time, and treatment.
- *Graph:* a handle to a graph of observations and predictions for the selected model.

We also made a stand-alone application (multimedia appendix 2) in the form of a MS Excel plug-in for users who do not want to struggle with Matlab. The application is made from Matlab code using the Matlab Excel Builder from Mathworks, Inc., which means that it can be distributed freely, without any need for having Matlab installed. Excel plug-ins produced will unfortunately only work under the MS Windows operating system. This application is more limited than the source code version, but its use does not require any programming skills. The plug-in is distributed as a self-extracting zip-file ('AIClim.exe') that will install itself in a chosen directory upon double-clicking. Successful installation depends on being logged on the computer with administrator privileges. The extraction directory will contain an Excel file called 'AIClim.xls.' Opening this file will reveal the following worksheet (Fig. 1; execution of macros must of course be enabled in order for the application to work).

The measurement data from the nutrient limitation experiment is entered into the 25 by 2 cell yellow area. Pay attention to the sequence of treatments and incubation times indicated to the left of this area: if data are entered in the wrong order, the whole analysis will be invalid. The AIC-based classification procedure is run on the present data set by clicking on the 'Calculate' button, which updates the two graphs and some hidden cells in the worksheet. The upper line graph on the screen shows the fit between observations and predictions for the model with the lowest AIC. The lower bar graph shows the distribution of classifications from a bootstrap analysis of replicated sub-samples. The number of random permutations is set by changing the number in the blue cell below the button. Normally 100 permutations will be sufficient but the number can be increased at the expense of longer execution times. The total height of the bars indicate probabilities of different classifications, while the shading on

the bars represent the frequencies of time effects of different order within that particular treatment contrast. In the example above, all resamples gave the same classification ('XP' – exclusive P limitation without any minor effect of N), 15% of which had a third order time effect while in the remainder the time effects were second order.

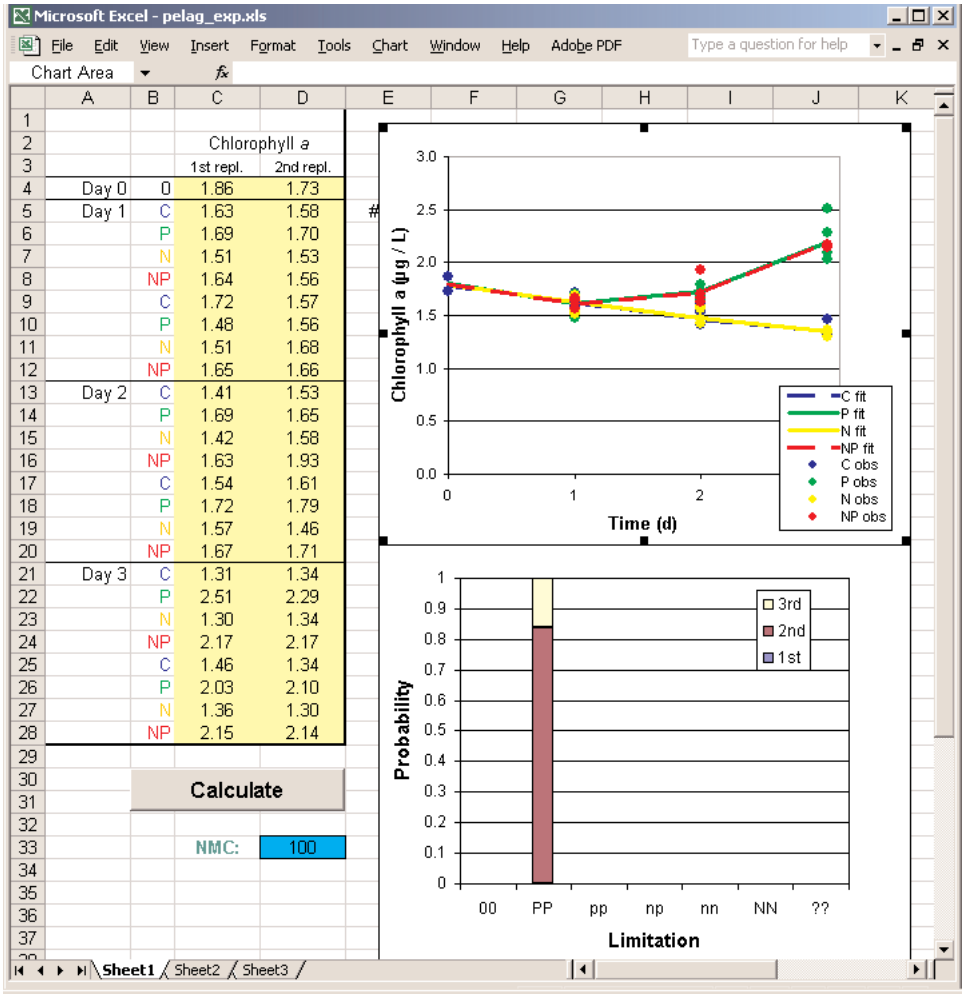

Fig. 1. Screen dump of MS Excel file "AIClim.xls" from multimedia appendix 2 (stand-alone MS Excel plug-in).