Comment on “From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity”

Stephen H. Roxburgh1* and Karel Mokany2

Shipley et al. (Reports, 3 November 2006, p. 812) developed a quantitative method for predicting the relative abundance of species from measured traits. We show that the method can have high explanatory power even when all trait and abundance data are randomly and independently generated, because of a mathematical dependence between the observations and predictions. We also suggest a potential solution to this problem.

Shipley et al. (1) developed a quantitative method, using tools borrowed from statistical mechanics, to predict plant relative abundances from field-measured species traits. When the authors applied the method to their field data (from 12 vineyards along a 42-year chronosequence of secondary succession in southern France), they found that up to 94% of the variation in the observed relative abundances could be accounted for by their model. The method is elegant and appealing (2, 3), particularly when considered against the backdrop of a highly successful field application. However, we show that (i) the model can generate high levels of explanatory power even when all trait and abundance data are randomly and independently generated and (ii) this spurious explanatory power arises because the model predictions and the field observations that are being compared are not statistically independent; they are functionally related within the model equations.

This functional dependence can be seen through inspection of the equation that is solved to generate the predictions of the species relative abundances (4)

\[ f(p_i) = \ln(W) - \lambda_0(1 - \sum p_i - \sum_{j=1}^{T} \lambda_i \left( 1 - \sum p_i \right) ) \]

where \( \lambda_i = \sum p_i / t_j \) and is the observed community-aggregate value for trait \( j \) summed over \( S \) species. The symbol \( p_i \) is the observed relative abundance of species \( i \), and \( p_i \) is the model-predicted relative abundance of the same species. The remaining symbols in the above equations are defined in (1).

Because the observed relative abundances are used both to test the model-predicted abundances and in the calculation of those predictions, the two quantities cannot be considered independent, and a relationship between them would a priori be expected.

We explored the implications of this dependence by Monte-Carlo analysis. Random communities were constructed with numbers of traits and species similar to those reported by Shipley et al. (1). Each random community comprised a matrix of \( T = 8 \) trait values with varying species richness (\( S = 10, 15, 20, 30, 50, \) or 100) species. Traits for each species were drawn independently and at random (5) based on the eight measured traits analyzed in (1), using previously published trait data for the intermediate-succession field ages (6). Random sets of “observed” relative abundances were drawn from the standard exponential distribution and normalized to sum to unity. An exponential distribution was used to simulate the typical pattern of dominance by a few abundant species. We refer to these abundances as “pseudo-observed” to prevent confusion with the actual field abundances. For each richness level, the number of traits was held constant at eight.

![Fig. 1. (A) Pseudo-observed versus model-predicted relative abundances for six different levels of community richness (\( S = 10, 15, 20, 30, 50, \) and 100 species). Each panel shows the combined scatter of 25 random communities for each richness level and the least-squares line of best fit for each of the 25 communities. For clarity, the figure is limited to the first 25 of 5000 random communities. (B) Frequency histograms of \( r^2 \) for 5000 random communities for each of the six levels of richness. For all simulations \( T \), the number of traits was held constant at eight.](http://science.sciencemag.org/content/full/316/5830/1425b)

---

1School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, Australia, and Bushfire Co-operative Research Centre, Equis, Post Office Box E4008, Kingston, ACT 2604, Australia. 2School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia.

*To whom correspondence should be addressed. E-mail: stephen.roxburgh@envis.comau
level, 5000 random communities were generated and provided as input to the algorithm (7).

In Fig. 1A, the pseudo-observed and model-predicted relative abundances are plotted for the first 25 random communities of each simulation, together with the least-squares line of best fit. Figure 1B displays frequency histograms of the coefficient of determination ($r^2$) for the pseudo-observed versus model-predicted abundances for the 5000 random communities. With 8 traits and 10 species, on average 90% of the variation is explained, even though the underlying pseudo-observed abundances and trait data were independently and randomly generated. As richness increases, the mean variation explained by the model declines (but always remains greater than zero). This reflects a change in the degrees of freedom of the analysis, with correspondingly fewer constraints on the optimization as the number of species increases relative to the number of traits measured (2).

Although it is tempting to compare our simulation results directly with the analysis of field data by Shipley et al. (1), such comparisons may be misleading for two reasons. First, the aim of our analysis was to explore the overall sensitivity of the theoretical framework to the lack of independence, using $r^2$ values calculated from simulated communities, each containing $S$ species. However, the results reported by Shipley et al. (1) are calculated from the combined abundance predictions from 12 separate analyses, one for each of 12 successional communities. Second, Shipley et al. analyzed each of their 12 communities with a matrix size of $S = 30$ species (the global richness across all 12 communities), even though the observed richness for any individual community was usually less than 20, declining to just two species toward the end of the succession (1, 8). Shipley et al.’s analysis therefore included matrices with a number of species with zero relative abundance, whereas all species in our simulations had positive abundances. A specific test of the validity of Shipley et al.’s field results would therefore require an analysis tailored to their field data set, that is, by aggregating predictions from multiple optimizations and by allowing species to have zero relative abundance.

Our analyses show that care must be taken when interpreting the results from Shipley et al.’s (1) method, as the variation explained by the model is confounded by at least two sources of variability. The first is the source of variation that is of primary interest, that is, the signature of nonrandom ecological patterns resulting from the process of community assembly. The second is variation due to the mathematical dependence of the observed and predicted abundances, as discussed above. To be of practical utility, the method requires modification to separate these two sources of variation. The randomization approach illustrated here, which quantifies the second source of variation through a null model–based approach, provides one option for separating these effects.

References and Notes
4. This equation is not numbered in (2), but can be found between Eqs. 4 and 5 on p. 813. The dependency between the observed and predicted relative abundances can also be seen in the equation at step (i) of the description of the algorithm used to estimate the predicted abundances, in the online document http://pages.usherbrooke.ca/shipley/recherche.
5. Random trait values were generated as follows. Using frequency distributions constructed from the observed trait data in (5) as a guide, we drew the three random biomass traits (V_m1, Stm, and Lm) from an exponential distribution, with mean equal to that in the observations. For Lifespan, the observed probability of perenniality was used to generate the binary data for this trait, and the remaining four traits (S_n, Smat, SLA, and H) were drawn from a normal distribution, with the observed mean and standard deviation.
7. To provide a focus for discussion, we based our randomizations on the characteristics of the actual field data; however, equivalent results can be obtained by relaxing this constraint and drawing random numbers freely from any appropriate distribution. An R program for conducting this analysis, and for the other analyses presented in this comment, is available from the lead author.
9. We thank P. Chesson, J. B. Wilson, and two anonymous reviewers for comments. We also thank B. Shipley for clarifying various aspects of the methodology.

13 December 2006; accepted 23 March 2007
10.1126/science.1138810
Comment on "From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity"

Stephen H. Roxburgh and Karel Mokany

*Science* **316** (5830), 1425.
DOI: 10.1126/science.1138810

http://science.sciencemag.org/content/316/5830/1425.2

http://science.sciencemag.org/content/sci/316/5830/1425.3.full
http://science.sciencemag.org/content/sci/314/5800/812.full

http://science.sciencemag.org/content/sci/316/5830/1425.4.full

This article cites 5 articles, 2 of which you can access for free
http://science.sciencemag.org/content/316/5830/1425.2#BIBL

http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service