 DOI: 10.1086/679576

Appendix D from X. Xiao et al., "A Strong Test of the Maximum Entropy Theory of Ecology"

(Am. Nat., vol. 185, no. 3, p. E70)

Validation of R^2 with Simulations

In the main text, we adopted the coefficient of determination R^2 between the empirical data and the values predicted by the maximum entropy theory of ecology (METE) to evaluate the performance of the theory. For the three patterns that are probability distributions (species abundance distribution [SAD], individual size distribution [ISD], and intraspecific ISD [iISD]), we ranked both the observed and the predicted values for comparison, resulting in two monotonically nondecreasing sequences (see "Analyses"). Such ranking could, in concept, lead to spuriously high coefficients of determination between the predictions and the observations, resulting in models appearing to perform well at prediction when in fact they did not.

We explored this possibility using simulations. For each community with state variables S_0, N_0, and E_0, we first constructed a simulated SAD by sampling S_0 abundance values from a discrete uniform distribution between 1 and $2(N_0–S_0)/S_0$, where the upper bound was chosen so that the expected total abundance of the simulated community equaled N_0. Then, for such a simulated community with S_0 species and N_0' individuals (with N_0' centered around but in most cases not equal to N_0), we further constructed a simulated ISD by sampling N_0' size values from a continuous uniform distribution between 1 and $2(E_0–N_0')/N_0'$. Again, the upper bound was chosen so that the expected total metabolic rate of the simulated community equaled E_0. This simulation procedure thus largely preserved the values of the state variables for each community, while the shape of the SAD and the ISD differed markedly from METE's predictions.

We conducted 100 simulations for each empirical community and applied METE to each simulated community with the values of the state variables S_0, N_0', and E_0' that resulted from the simulation. The performance of METE on the SAD and the ISD was evaluated in the same way as in the main text, with R^2 between the "observed" (simulated) and predicted rank values. If the high R^2 values that we obtained for the SAD and the ISD in empirical communities are an artifact of ranking, we would expect equally high R^2 values for the simulated communities. In contrast, we found that R^2 values for the SAD and the ISD in the vast majority of the simulated communities were below 0, and the R^2 for the two patterns in real communities were higher than those in any of the 100 simulated communities (fig. D1). This shows that METE's predictive power for the SAD and the ISD is not an artifact of ranking, consistent with previous studies of the SAD (White et al. 2012*a*) as well as our alternative analysis for the ISD where the distribution is not converted to ranks (app. C, fig. C1).

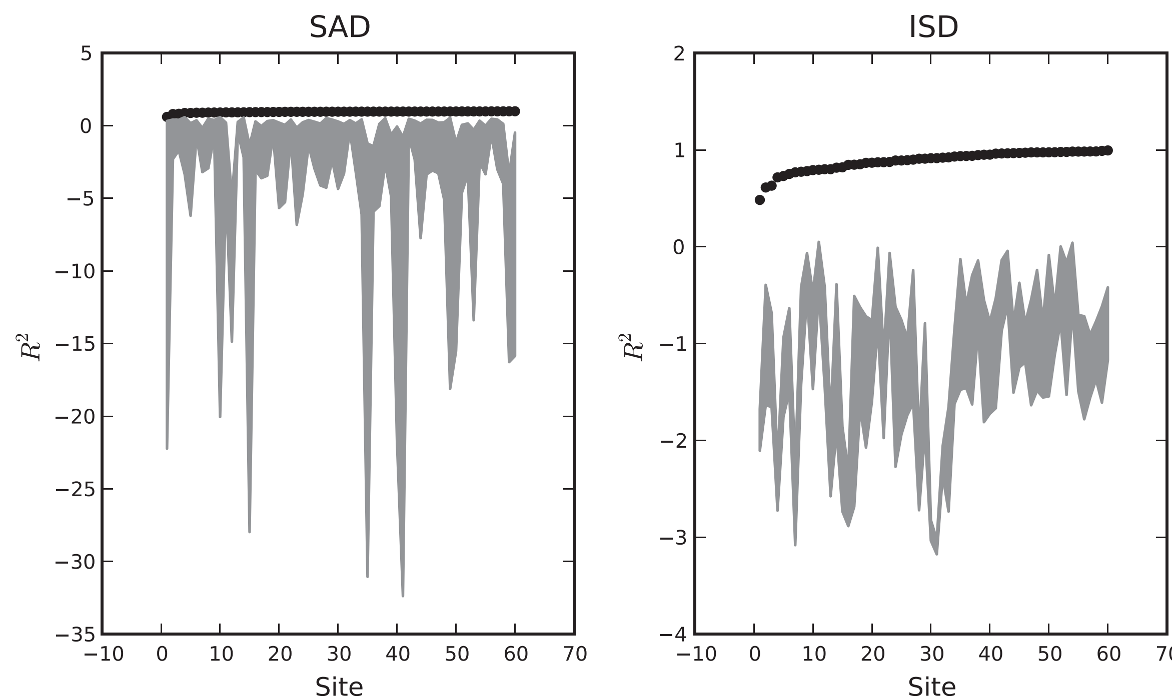

Figure D1: R^2 for the species abundance distribution (SAD) and the individual size distribution (ISD) in the empirical communities (black circles) versus the full range of R^2 for the two patterns in 100 simulated communities (gray area), where both the SAD and the ISD were generated from uniform distributions.