 DOI: 10.1086/679576

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

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

Alternative Analyses for the Individual Size Distribution (ISD) and the Intraspecific ISD (iISD)

In our analyses in the main text, we converted all three probability distributions (species abundance distribution [SAD], ISD, and iISD) into distributions of rank and compared the predicted values at each rank against the observed values. While this approach has been widely adopted (Harte et al. 2008; Harte 2011; White et al. 2012*a*), it may not be entirely adequate for continuous distributions such as the ISD and the iISD, where empirical measurements are usually rounded off to decimals and thus may not be directly comparable to the truly continuous values obtained from the predicted distributions of rank. Here, we conduct additional analyses for the ISD and the iISD with alternative approaches applied directly on the probability distributions without converting them to distributions of rank to demonstrate the robustness of our results.

For the ISD, we grouped the scaled individual metabolic rates into log(1.7) bins (1–1.7, 1.7–2.89, 2.89–4.913, etc.), which resulted in 10–21 bins for each forest community. The predicted frequency for each bin was then calculated from the cumulative distribution of $\Psi(\varepsilon)$ (eq. [4]) and compared with the observed frequency. The predictive power of the maximum entropy theory of ecology (METE) for the ISD does not change qualitatively when the ISD is analyzed as frequencies ($R^2 = 0.93$; fig. C1) instead of as ranked metabolic rates ($R^2 = 0.89$; fig. 2*B*).

The iISDs for most species contain too few individuals for the above-described analysis with binned frequencies. Instead, we directly looked at the shape of the distribution. METE predicts that the iISD for each species within a community follows an exponential distribution left truncated at 1, with the parameter of the distribution proportional to the abundance of the species (see eq. [5]). Deviation from METE's prediction can occur in one or both of two ways: (1) the observed iISDs are not well characterized by exponential distributions; and (2) assuming that the iISDs can be characterized by exponential distributions (which may or may not be true), the parameter of the distributions that best capture the observed iISDs differ from those predicted by METE (eq. [5]). Here, we show that METE's prediction for iISD fails in both aspects, which is consistent with our results in the main text (fig. 2*D*).

Characterizing iISDs with Exponential Distributions

In each community, we fit an exponential distribution left truncated at 1 (the minimal rescaled metabolic rate within each community) to rescaled individual metabolic rates for each species with at least 5 individuals and obtained the maximum likelihood (MLE) parameter of the distribution. For each species, 5,000 independent samples were drawn from a left-truncated exponential distribution with the MLE parameter, where the sample size was equal to the abundance of the species. The two-sample Kolmogorov-Smirnov test was then applied to evaluate whether the empirical iISD differs significantly from each sample drawn from the left-truncated exponential distribution. If the proportion of tests (among all 5,000) where the empirical iISD and the randomly generated sample differ in distribution is higher than the significance level (α) of the tests, the empirical iISD for the focal species does not conform to a left-truncated exponential distribution.

Figure C2 shows a histogram of proportions of Kolmogorov-Smirnov tests that are significant at $\alpha = 0.05$ among species (with abundance ≥ 5) across all 60 communities. Overall, the iISDs for more than half the species are deemed to be significantly different from the left-truncated exponential distribution, which implies that the form of iISD predicted by METE does not hold.

Comparing the MLE Parameter with METE's Predicted Parameter

We further compared the MLE parameter of the left-truncated exponential distribution for each species to the parameter predicted by METE ($\lambda_2 n$; see eq. [5]; fig. C3). Note that this analysis is biased in favor of METE, as we have already shown that left-truncated exponential distribution does not provide a good characterization of empirical iISD for most species (fig. C2). That the R^2 value for the iISD is below 0 even when METE is evaluated with this biased analysis further strengthens our conclusion that METE is unable to meaningfully capture any variation in the iISD.

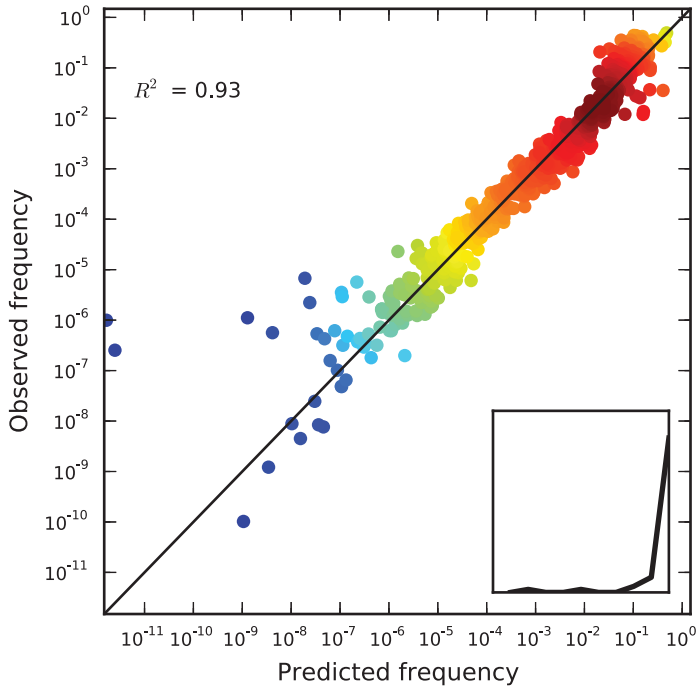

Figure C1: Plot of maximum entropy theory of ecology's predictions against empirical observations across 60 communities for the individual size distribution, which is analyzed as binned frequencies. The diagonal black line is the 1 : 1 line. The points are color coded to reflect the density of neighboring points, with warm (red) colors representing higher densities and cold (blue) colors representing lower densities. The inset in the lower right corner shows the distribution of R^2 among individual communities from below 0 (*left*) to 1 (*right*).

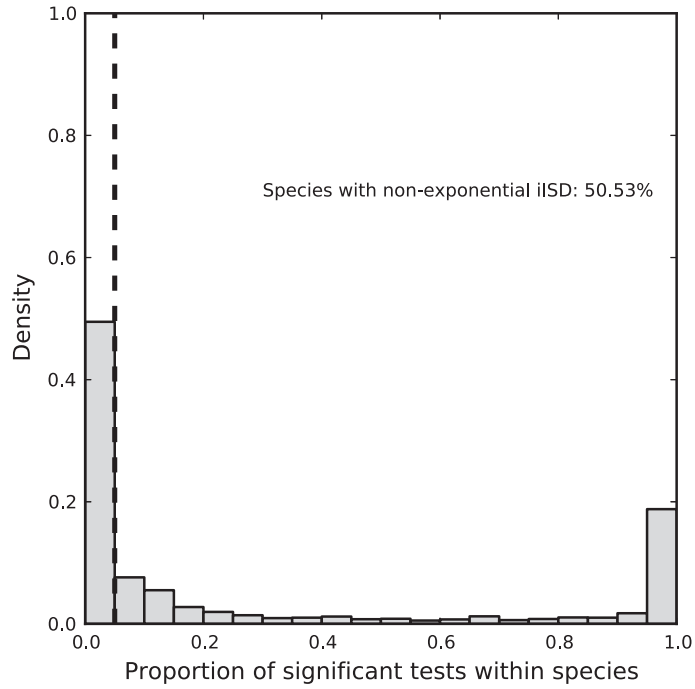

Figure C2: Histogram of the proportion of Kolmogorov-Smirnov tests that are significant for each species. The dashed vertical line represents the significance level of the tests ($\alpha = 0.05$). Species for which the proportion of tests (among 5,000) with significant results is higher than 0.05 have intraspecific individual size distributions (iISDs) that differ significantly from the left-truncated exponential distribution.

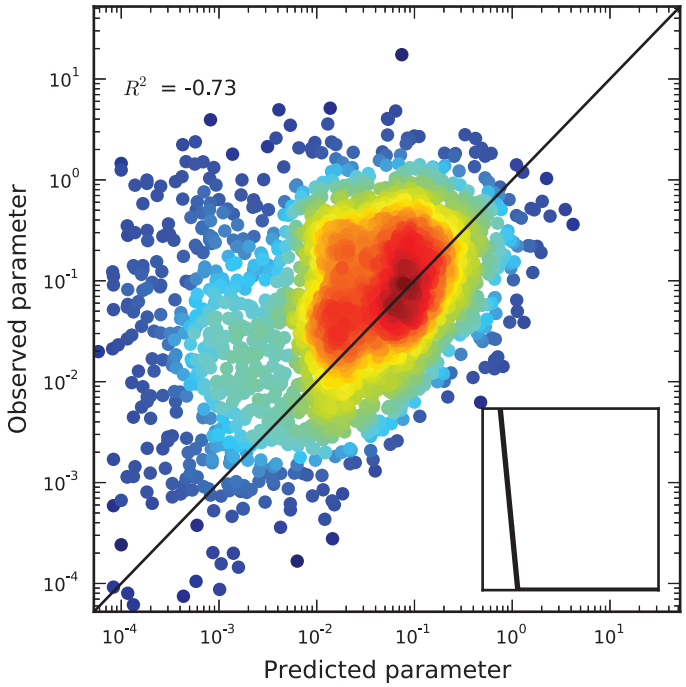

Figure C3: Intraspecific individual size distribution (iISD) parameter predicted by the maximum entropy theory of ecology plotted against maximum likelihood parameter for the empirical distribution for each species (with no fewer than five individuals) in each of the 60 communities. The diagonal black line is the 1 : 1 line. The points are color coded to reflect the density of neighboring points, with warm (red) colors representing higher densities and cold (blue) colors representing lower densities. The inset reflects the distribution of R^2 among 60 communities from negative (*left*) to 1 (*right*).