Ecological Archives Exxx-xxx-A1

Appendix F. The functional form of the ASED (*Nu*) as derived from a discrete cumulative distribution function.

Because the SAD is a discrete distribution, the theoretical ASED distribution will reflect that discrete nature. However, as a function of a continuous variable, $\bar{\varepsilon}$, the empirical ASED may take on any value in the range of 1 to ε_{max}. The METE-predicted ASED is only defined for $\bar{\varepsilon}$ values given by $\bar{\varepsilon} = (1+(1/(n \cdot \lambda_2))$, where *n* (the number of individuals of a species) is in the range [1,*N*] (and where $N = n_{max}$). Although we would ultimately like to model metabolic rate as a continuous distribution (conforming to our understanding of metabolic "energy" as a continuous variable in nature), there is no closed form rank-metabolic rate distribution for the ASED. The method suggested in Harte, 2011 (eq. 7.44, p. 155), allows us to calculate the ASED numerically and create a model that is close to a continuous distribution. We use this numerically-estimated continuous formulation of the ASED for the all analyses in this paper, and discuss the problems of doing so below.

Various artifacts arise from trying to predict a continuous variable, metabolic "energy" (or metabolic rates, defined in units of power), from a rank-metabolic rate distribution constructed from a discrete probability distribution. For example, when using a discrete ASED graphed on rank-metabolic rate and rank-log(metabolic rate) graphs, predicted values for multiple species with large average metabolic rates have a maximum predicted value, and values in the mid-range may be repeated. A continuous probability distribution would instead predict the ASED to be a strictly decreasing function (rather than a stepped function) of ranked average metabolic rates.

Graphically, the values $\bar{\varepsilon}_1$, $\bar{\varepsilon}_2$, ... $\bar{\varepsilon}_N$ (where $\bar{\varepsilon}_1$ is the smallest value and $\bar{\varepsilon}_N$ is the largest value) get progressively more spread out approaching $\bar{\varepsilon}_N$. With this in mind, we build a rank-metabolic rate distribution using the cumulative distribution function (CDF) through the following steps: (1) calculate an observed CDF for the observed energy values, and (2) use the cumulative probability values in the observed CDF and find the $\bar{\varepsilon}$ in the METE-predicted CDF with the same cumulative probability value. This is the predicted metabolic rate value. Therefore, constructing a CDF from this type of probability distribution causes large steps on the x-axis between each gain in probability at large $\bar{\varepsilon}$.

For example, as in Fig. F1, for three observed probability values of 0.81, 0.88, and 0.92, the continuous CDF (blue line) estimates a different $\bar{\varepsilon}$ for each probability value, while a discrete CDF (green line) has $\bar{\varepsilon}$ falling in the gap between two defined metabolic rate values and therefore must be assigned to either the lower or higher energy value. Therefore, different CDF values may end up with the same predicted $\bar{\varepsilon}$, leading to a step-like, artifactual pattern that may not resemble patterns of natural systems. This explains why it is a general feature of the predicted, discrete ASED that we see a flat line of predicted values for the first few ranks, *i.e.* the highest $\bar{\varepsilon}$ values, on rank abundance graphs, and may see repeated values elsewhere in the distribution.
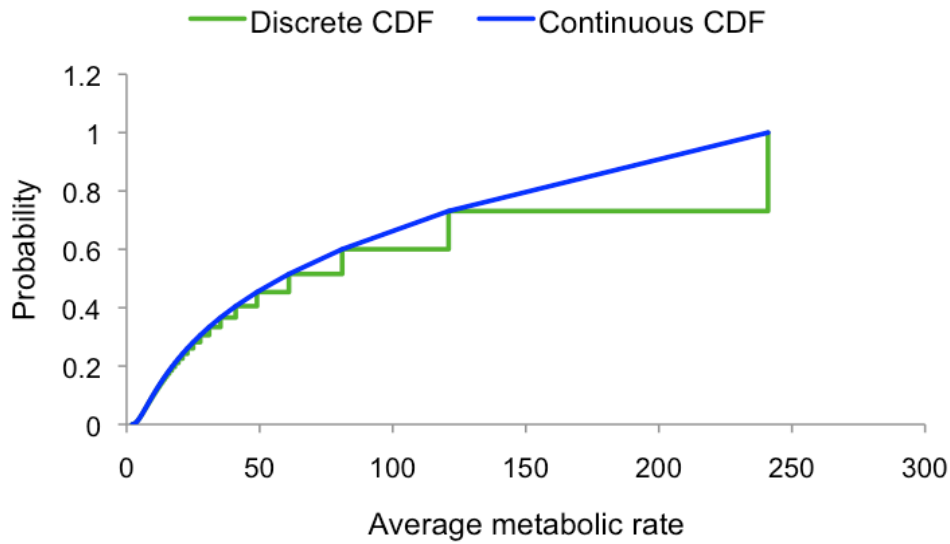
FIG. F1. Discrete predictions of *Nu* distribution compared to idealized continuous curve. These example distributions were generated with $N = 200$, $S = 20$, and total energy $E = 5000$.

LITERATURE CITED

Harte, J. 2011. Maximum Entropy and Ecology: A Theory of abundance, distribution, and energetics. Oxford University Press, Oxford, UK.