Chun-Wei Chang, Takeshi Miki, Fuh-Kwo Shiah, Shuh-ji Kao, Jiunn-Tzong Wu, Akash R. Sastri and Chih-hao Hsieh. Year. **Linking secondary structure of individual size distribution with nonlinear size-trophic level relationship in food webs.**

APPENDIX A

Mathematical derivation of the generalized MTE model

We proposed a generalized metabolic theory of ecology (MTE) to explain the structure of individual size distribution (ISD). In the generalized MTE model, we relax the assumption of linear size-TL relationship. (Size is measured as log body mass or log body volume; hereafter, we refer to $\log M$ as the body size. TL is the trophic level.) In the original metabolic theory (Brown et al. 2004), the ISD was derived by comparing the ratio between the total metabolism and individual metabolism of organisms at any given size M under the consideration of energy loss due to trophic transfer. The trophic level critically depends on the size of organisms. Consider the total metabolism $I_{tot}(M)$ for organisms of body size M,

$$I_{tot}(M) = i_0 N_0 M_0^{3/4} e^{-E/kT} \alpha^{\tau(M)} \tag{A.1}$$

where i_0 is the normalization constant, N_0 is the primary producer abundance, M_0 is the primary producer body size, and α is the transfer efficiency across trophic levels. Here, $\tau(M)$ is the trophic level at size M, which represents the size-TL relationship. In the original MTE, the size-TL relationship is assumed to be a linear function of $\log M$:

$$\tau(M) = b_0 + b_1 \log M \quad,$$

where $b_0 = -\dfrac{\log M_0}{\log \beta}$ and $b_1 = 1/\log \beta > 0$, and β is the predator-prey mass ratio (PPMR).

Note that this assumption implies that the PPMR is a constant and independent of body size.

In the general MTE model, we relax this assumption of linear size-TL relationship; we assume that $\tau(M)$ is a smooth function of body size which can be approximated by a mth order polynomial function, i.e.

$$\tau(M) = b_0 + b_1 \log M + b_2 \log^2 M + \ldots + b_m \log^m M = \sum_{i=0}^{m} b_i \log^i M$$

This phenomenological approximation is motivated by empirical observations (see main text).

With this setting, $\alpha^{\tau(M)}$ in the original MTE model should be replaced by the mth order polynomial of $\log M$ as follows:

$$\alpha^{\tau(M)} = \alpha^{\sum_{i=0}^{m} b_i \log^i M} = \alpha^{b_0} \left(\alpha^{\log M} \right)^{\sum_{i=1}^{m} b_i \log^{i-1} M} = \alpha^{b_0} M^{\log \alpha \sum_{i=1}^{m} b_i \log^{i-1} M} \tag{A.2}$$

Because the abundance N of organisms with size M is equal to the total metabolism I_{tot} divided by individual metabolism I_{ind} of size M, we derive the following relationship:

$$N = \frac{I_{tot}(M)}{I_{ind}(M)} = \frac{i_0 N_0 M_0^{3/4} e^{-E/kT} \alpha^{b_0} M^{\log \alpha \sum_{i=1}^{m} b_i \log^{i-1} M}}{i_0 M^{\lambda} e^{-E/kT}}$$

$$= N_0 M_0^{3/4} \alpha^{b_0} M^{-\lambda + \log \alpha \sum_{i=1}^{m} b_i \log^{i-1} M}$$

where λ is the metabolic exponent, which is usually taken as 3/4 as the original MTE (Brown et al. 2004) or can be other values derived from other models (Pawar et al. 2012). For simplicity, we assumed that trophic efficiency α is equal to 10% as suggested in previous studies (Lindeman 1942, Brown and Gillooly 2003). Thus, the derived ISD is further

simplified to:

$$N = aM^{-\lambda - \sum_{i=1}^{m} b_i \log^{i-1} M}$$

$$\log N = \log a - (\lambda + b_1) \log M - \sum_{i=2}^{m} b_i \log^i M \qquad \textbf{(A.3)}$$

where a is equal to $0.1^{b_0} N_0 M_0^{3/4}$ and is the constant that determines the intercept of the ISD.

According to this equation, the log transformed ISD can be decomposed into two major components: the power-law core distribution, $-(\lambda+b_1)\log M$, and the polynomial secondary structure, $-\sum_{i=2}^{m} b_i \log^i M$. Here, the power-law componenet has the same structure as predicted by the original MTE (Brown et al. 2004). While in this study, we additionally identify the secondary structure, which has not been quantitatively examined before. Here, we aim to emphasize and derive the theoretical properties of the secondary strucutre in this study. Based on eqn A.3, the magnitude of the secondary strucutre, b_i, is much smaller than the power-law exponent because the power exponent, $-(\lambda+b_1)$, contains a large metabolic exponent (i.e. ¾, according to Brown et al 2004). Therefore, we can see that the secondary structure is subtle and has a much smaller value than the power-law component. However importantly, the power-law distribution and secondary structure of ISD come from different ecological processes acting at different scales. The power-law compoenent is the result of size-metabolism scaling and average predator-prey mass ratio, while, the secondary strcture comes from the complexity of trophic links (the nonlinearity of size-TL relationship) in size-based food webs. In addition, the eqn A.3 also indicates that the secondary structure can

be separated from the power-law component in log ISD; that is, one can isolate the secondary structure from the power-law distribution in empirical measurements. More specifically, after partitioning out the power-law component from the log ISD by statistical fitting, the residuals deviating from such power-law fitting will leave only secondary structure (plus small random noise). Therefore in empirical measurements, in order to extract the secondary structure of ISD, we suggest to obtain the residuals of the log ISD from fitting the power-law, instead of simultaneously fitting both the power-law exponent and the polynomial secondary structure. Such a two-step fitting procedure (excluding the power-law first and then obtaining the secondary structure from the residuals) allows one to extract the secondary structure (See justification in Appendix C). Because the residuals contain information of secondary structure, we practically define the secondary structure of ISD as the polynomial function of ISD residuals versus body size. In the following section, we will discuss the theoretical properties of the residuals.

As explained above, the secondary strucutre of ISD appears when the size-TL relationhip is nonlinear. More importantly, the residuals deviating from the power-law fitted log ISD contain determinstic secondary structure, a footprint of the nonlinear size-TL relationhip. Consider a ISD data set (N_i, M_i), i=1, 2, ..., n, collected under the size-TL relationhip following m^{th} order polynomial relationship. After been log transformed, the relationship between N_i and M_i can be presented in a matrix form:

$$\mathbf{Y} = \begin{pmatrix} \log N_1 \\ \log N_2 \\ \vdots \\ \log N_n \end{pmatrix} = \begin{bmatrix} 1 & \log M_1 & \cdots & \log^m M_1 \\ 1 & \log M_2 & \cdots & \log^m M_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \log M_n & \cdots & \log^m M_n \end{bmatrix} \begin{pmatrix} a \\ -b_1 - \lambda \\ -b_2 \\ \vdots \\ -b_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

where $\varepsilon_i \sim (0, \sigma^2)$, all ε_i are independent identical distributed. The expectation of $\mathbf{Y}$ can be expressed as

$$E(\mathbf{Y} \mid \mathbf{M}) = (\mathbf{1} \quad \log \mathbf{M} \quad \log^2 \mathbf{M} \quad \cdots \quad \log^m \mathbf{M}) \begin{pmatrix} a \\ -b_1 - \lambda \\ -b_2 \\ \vdots \\ -b_m \end{pmatrix},$$

where $\mathbf{M} = (M_1 \quad M_2 \quad \cdots \quad M_n)$.

Partitioning the matrix, one would get

$$E(Y \mid M) = [\mathbf{X} \quad \mathbf{P}](\mathbf{b}_x \quad -\mathbf{b}_p)^{\mathrm{T}}$$

$$= \mathbf{X}\mathbf{b}_x^{\mathrm{T}} - \mathbf{P}\mathbf{b}_p^{\mathrm{T}} \tag{A.4}$$

where $\mathbf{X}_{n\times 2} = (\mathbf{1} \quad \log \mathbf{M})$, $\mathbf{P}_{n\times(m-1)} = (\log^2 \mathbf{M} \quad \log^3 \mathbf{M} \quad \cdots \quad \log^m \mathbf{M})$, $\mathbf{b}_x = (a \quad -b_1 - \lambda)$,

and $\mathbf{b}_p = (b_2 \quad b_3 \quad \cdots \quad b_m)$.

Accordding to the statistical theory of linear models, if we fit the ISD (fitting **Y** by **X**) by a simple power-law model using a least squares approach, the residuals **e** of such a fitting would demonstrate higher order determinstic structure.

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y},$$

where **I** is n-dimensional identiy matrix; $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$ is the hat matrix that projects **Y** onto the linear space spanned by **X**. Taking the expectation for the residuals and substituting

the expectration of **Y** by eqn A.4, one obtains the expectation of residuals **e** as

$$E(\mathbf{e}\,|\,\mathbf{M}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}})E(\mathbf{Y}\,|\,\mathbf{M})$$

$$= (\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}})(\mathbf{X}\mathbf{b}_x^{\mathrm{T}} - \mathbf{P}\mathbf{b}_p^{\mathrm{T}})$$

$$= \mathbf{X}\mathbf{b}_x^{\mathrm{T}} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}_x^{\mathrm{T}} - \mathbf{P}\mathbf{b}_p^{\mathrm{T}} + \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{b}_p^{\mathrm{T}}$$

$$= -\mathbf{P}\mathbf{b}_p^{\mathrm{T}} + \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{b}_p^{\mathrm{T}}$$

$$= -(\mathbf{I} - \mathbf{H})\mathbf{P}\mathbf{b}_p^{\mathrm{T}}$$

$$= -\mathbf{Q}\mathbf{b}_p^{\mathrm{T}} \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{(A.5)}$$

where $\mathbf{Q}_{n\times(m-1)} = (\mathbf{I} - \mathbf{H})\mathbf{P} = ((\mathbf{I} - \mathbf{H})\log^2\mathbf{M} \quad (\mathbf{I} - \mathbf{H})\log^3\mathbf{M} \quad ... \quad (\mathbf{I} - \mathbf{H})\log^m\mathbf{M})$.

Each column of **Q**, $\mathbf{Q}_j = (\mathbf{I} - \mathbf{H})\log^j\mathbf{M}$, is the least squares residual of $\log^j\mathbf{M}$ fitted by **X**,

j=2, ..., m. This theoretical property is useful. It indicates that the residuals of ISD deviating

from the power-law fit can be used to predict the polynomial coefficients of size-TL

relationship, $\mathbf{b}_p$, quantitatively. Practically, this prediction can be implemented if we fit the

residuals of ISD (deviating from the linear fit in log space) by a multiple regression model

with the covariates $-\mathbf{Q}_j$ which are the negative value of least square residuals of $\log^j\mathbf{M}$ fitted

by $\log\mathbf{M}$, j=2, ..., m. In this multiple regression, the regression coefficients can be used to

predict the same order polynomial coefficients of size-TL relationship. Ecologically, such

prediction allows us to link the dynamics of ISD with the dynamics of a complex size-based

food web. We note that there is no theoretical value of m (the order of polynomial function);

rather, m can be determined empirically according to some statistical criteria such as AIC.

Throughout this research, we used the order *m*=3 in empirical analyses; this highest possible order is limited by the data resolution and supported by the AIC criterion in examining our empirical size-TL relationships.

Conversely, we can use the size-TL relationship to predict the secondary structure of ISD. In eqn A.3, $N = aM^{-\lambda - \sum_{i=1}^{m} b_i \log^{i-1} M}$, the parameters b_i, can be substituted by the estimates of the ith order polynomial coefficients of the empirical size-TL relationship. Note that assigning any intercept, *a*, does not change our prediction on secondary structure. After simulating the ISD based on eqn A.3, we can extract the secondary structure from the residuals of the log ISD fitted by power-law. As a result, by following the above procedures, we can predict the secondary structure of ISD in log space by the empirical size-TL relationship.

LITERATURE CITED

Brown, J. H., and J. F. Gillooly. 2003. Ecological food webs: High-quality data facilitate theoretical unification. Proceedings of the National Academy of Sciences, USA 100:1467–1468.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology 85:1771–1789.

Lindeman, R. L. 1942. The trophic-dynamic aspect of ecology. Ecology 23:399–417.

Pawar, S., A. I. Dell, and V. M. Savage. 2012. Dimensionality of consumer search space drives trophic interaction strengths. Nature 486:485–489.