Online resource 2

Electronic supplementary material to:

Journal: Theoretical Ecology

Title: Trophic niche-space imaging, using resource and consumer traits

Authors: Leopold A. J. Nagelkerke [1,*] and Axel G. Rossberg [2,3]

[1] Aquaculture and Fisheries Group, Wageningen University, Wageningen Institute of Animal Sciences (WIAS), De Elst 1, 6708 WD Wageningen, The Netherlands, e-mail: Leo.Nagelkerke@wur.nl, Tel: +31 317 483940, fax: +31 317 483937

[2] Lowestoft Laboratory, Centre for Environment, Fisheries and Aquaculture Science (Cefas), Pakefield Road, Lowestoft, Suffolk NR33 0HT, United Kingdom

[3] School of Biological Sciences, Queen's University Belfast, Belfast BT9 7BL, Northern Ireland, United Kingdom, e-mail: axel@rossberg.net, Tel: +44 28 9097 5859, fax: +44 28 9097 5877

[*] Author for correspondence

Online resource 2

Electronic supplementary material to:

Journal: Theoretical Ecology

Title: Trophic niche-space imaging, using resource and consumer traits

Authors: Leopold A. J. Nagelkerke and Axel G. Rossberg

A   Details of data analysis

Our procedure to obtain the trophic trait vectors of consumers and resources from field data has three main aspects: (1) parametrization of the trophic trait model (TTM) based on observed feeding behaviour for a given set of empirical trait variables, (2) cross-validation of the model to assess its predictive power, and (3) selection an appropriate set of trait variables such as to achieve good predictive power. Variable selection (3) requires repeated cross validation (2), and cross-validation repeated model fitting (1).

A.1   Parametrization of the TTM

Input trait variables   The procedure for parametrizing the TTM starts from two given sets of phenotypic traits, one for resources and one for consumers. The values of the resource trait variables are known for each of R kinds of resource, here diet categories, and the values of the consumer trait variables are known for each of Q kinds of consumers, here labeobarb species. For each resource category, the values of n trait variables are given, and for each labeobarb species the values of m trait variables. The trait values for resource i are given by an n-component vector of real values $\tilde{\mathbf{t}}_i$, that is $\tilde{\mathbf{t}}_i$ and trait values for consumer j by an m-component vector $\tilde{\mathbf{s}}_j$. The trait values of labeobarbs were log-transformed. By taking logarithms, the problem of how to best account for dependencies of trait variables on body size is mitigated, because different normalizations in terms of body size correspond to different linear combinations of logarithmic traits. The best-fitting abstract foraging traits, which are linear combinations of logarithmic phenotypic traits, will automatically combine these traits in such a way as to achieve the most appropriate normalization.

Trait standardization   The input trait values were now standardized, by transforming them into equivalent sets of trait values that were statistically uncorrelated and had mean zero and variance one. This was achieved as follows: from the vectors of phenotypic traits, we first computed the means over the sets of consumers and resources considered, i.e. $\bar{\mathbf{t}} = R^{-1} \sum_{i=1}^{R} \tilde{\mathbf{t}}_i$ and $\bar{\mathbf{s}} = Q^{-1} \sum_{j=1}^{Q} \tilde{\mathbf{s}}_j$, and then formed the raw resource trait matrix defined by $\tilde{\mathbf{T}} = (\tilde{\mathbf{t}}_1 - \bar{\mathbf{t}}, ..., \tilde{\mathbf{t}}_R - \bar{\mathbf{t}})$ and raw consumer trait matrix defined by $\tilde{\mathbf{S}} = (\tilde{\mathbf{s}}_1 - \bar{\mathbf{s}}, ..., \tilde{\mathbf{s}}_Q - \bar{\mathbf{s}})$. Then we sought matrices $\mathbf{A}$ and $\mathbf{B}$ such that $(R-1)^{-1}\mathbf{A}\tilde{\mathbf{T}}\tilde{\mathbf{T}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = I$, and $(Q-1)^{-1}\mathbf{B}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} = I$, and defined new, scaled trait matrices as $\mathbf{T} = \mathbf{A}\tilde{\mathbf{T}}$, $\mathbf{S} = \mathbf{B}\tilde{\mathbf{S}}$.[1] The columns of $\mathbf{T} = (\mathbf{t}_1, ..., \mathbf{t}_R)$ and $\mathbf{S} = (\mathbf{s}_1, ..., \mathbf{s}_Q)$ are equivalent representations of the original phenotypic traits. The original trait values can be recovered by making use of $\mathbf{A}$, $\mathbf{B}$, $\bar{\mathbf{s}}$, and $\bar{\mathbf{t}}$.

---

[1]To find an appropriate transformations $\mathbf{A}$, we computed the eigendecomposition of the covariance matrix $(R-1)^{-1}\tilde{\mathbf{T}}\tilde{\mathbf{T}}^{\mathrm{T}} = \mathbf{VDV}^{\mathrm{T}}$, with the diagonal matrix $\mathbf{D}$ containing the eigenvalues and the columns of $\mathbf{V}$ the eigenvectors, and set $\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{V}^{\mathrm{T}}$. We proceeded analogously for consumer traits to obtain $\mathbf{B}$.

cov.normalize [functions.R]

1

One might be concerned that this standardization procedure enhances measurement errors for trait variables that show little variability over the data set considered or for combinations of trait variables that are strongly correlated. However, selections of trait variables that include such cases will be identified by the cross-validation procedure that we apply to each candidate combination, and the variable-selection procedure will reject them.

Relation between link-strength function stomach-content data   There are various ways in which trophic link strength can be defined. Here we use a particularly simple definition: We interpret $a(\mathbf{t},\mathbf{s})\Delta t$ as the probability that a consumer individual with traits $\mathbf{s}$, upon encountering a resource item with traits $\mathbf{t}$, will feed on it within a small time interval of length Δt. Mathematically, this is the conditional probability $a(\mathbf{t},\mathbf{s})\Delta t = P[\mathbf{s}$ eats $\mathbf{t}$ within $\Delta t|\mathbf{s}$ meets $\mathbf{t}]$. By basic probability theory and the assumption that encounters occur at random, a feeding interaction will then occur with probability

$$P[\mathbf{s} \text{ meets and eats } \mathbf{t} \text{ within } \Delta t] = a(\mathbf{t},\mathbf{s})P[\mathbf{s} \text{ meets } \mathbf{t}]\Delta t$$
$$\propto a(\mathbf{t},\mathbf{s}) \times (\text{density of } \mathbf{s}) \times (\text{density of } \mathbf{t}). \tag{1}$$

The precise meaning of Eq. (1) depends on the statistical ensemble considered. Here, we choose the population of one labeobarb species at random with uniform probability and randomly pick one of its individuals. Within this ensemble, the density of labeobarbs with traits $\mathbf{s}$ is a fixed constant when a species with traits $\mathbf{s}$ exists, and zero otherwise.[2] Resources we consider sampled at random from the volume of the lake. The probability of encounter with prey that has traits $\mathbf{t}$ is proportional to the volumetric abundance of prey with traits $\mathbf{t}$ in the lake if such prey exists, and zero otherwise. If, upon encounter, a labeobarb individual eats the prey item, this ends up in its gut. The probability on the left hand side of Eq.(1) is therefore approximately proportional to the volumetric proportion in the gut content: $P[\mathbf{t}$ meets and eats $\mathbf{s}$ within $\Delta t] \propto$ (proportion of $\mathbf{s}$ in gut of $\mathbf{t}$) if labeobarbs with traits $\mathbf{t}$ exists, and zero otherwise.

Estimation of the link-strength function   To estimate $a(\mathbf{t},\mathbf{s})$ by Eq. (1), we approximated the three densities entering Eq. (1) by normal distributions. To determine the parameters of these normal distributions, we computed the mean $\boldsymbol{\mu_t}$ and covariance matrix $\mathbf{C_t}$ of occurrences of resource traits $\mathbf{t}$ in the lake, the mean $\boldsymbol{\mu_s}$ and covariance matrix $\mathbf{C_s}$ of consumer traits $\mathbf{s}$ over all labeobarb species with equal weights, and the mean $\boldsymbol{\mu}_f$ and covariance matrix $\mathbf{C}_f$ of pairs $(\mathbf{t},\mathbf{s})$ in feeding interactions. Since, unfortunately, reliable direct measurements of the densities of resource items are not available, we use the mean proportion of prey items in labeobarb gut content as a rough surrogate, as discussed in the main text. With f_{rc} denoting the proportion of item r in the gut of labeobarb species c (*diet proportion*) this gives estimated densities $\rho_r = Q^{-1}\sum_{c=1}^{Q} f_{rc}$. Abbreviating further $N = \sum_{rc} f_{rc}$, we obtain

---

[2]Mathematically inclined readers will notice that we gloss over a few Dirac-delta functionals and questions regarding proportionality constants here. Discussing these subtleties would not yield any additional insights.

$$\boldsymbol{\mu_t} = \left(\sum_{r=1}^{R} \rho_r\right)^{-1} \sum_{r=1}^{R} \mathbf{t}_r \rho_r = \frac{1}{N} \sum_{c=1}^{Q} \sum_{r=1}^{R} \mathbf{t}_r f_{rc}$$

$$\boldsymbol{\mu_s} = \frac{1}{Q} \sum_{c=1}^{Q} \mathbf{s}_c, \tag{2}$$

$$\boldsymbol{\mu}_f = \frac{1}{N} \sum_{c=1}^{Q} \sum_{r=1}^{R} \begin{pmatrix} \mathbf{t}_r \\ \mathbf{s}_c \end{pmatrix} f_{rc},$$

and

$$\mathbf{C_t} = \frac{1}{N} \sum_{c=1}^{Q} \sum_{r=1}^{R} (\mathbf{t}_r - \boldsymbol{\mu_t})(\mathbf{t}_r - \boldsymbol{\mu_t})^{\mathrm{T}} f_{rc},$$

$$\mathbf{C_s} = \frac{1}{Q} \sum_{c=1}^{Q} (\mathbf{s}_c - \boldsymbol{\mu_s})(\mathbf{s}_c - \boldsymbol{\mu_s})^{\mathrm{T}}, \tag{3}$$

$$\mathbf{C}_f = \frac{1}{N} \sum_{c=1}^{Q} \sum_{r=1}^{R} \left[\begin{pmatrix} \mathbf{t}_r \\ \mathbf{s}_c \end{pmatrix} - \boldsymbol{\mu}_f \right] \left[\begin{pmatrix} \mathbf{t}_r \\ \mathbf{s}_c \end{pmatrix} - \boldsymbol{\mu}_f \right]^{\mathrm{T}} f_{rc}.$$

Three short technical notes help understanding details of calculations in Eqs. (2) and (3). Note 1: The diet proportions of a consumer do not always add up to 1, e.g., when resource items are omitted from the statistics or when portions of diets remain unclassified. The normalization constant N takes this into account. Note 2: Because we computed weighted co-variances, no attempt of a Bessel correction (a denominator of "$n-1$" rather than "n") was made. Note 3: Due to the initial data standardization $\boldsymbol{\mu_s} = 0$, and $\mathbf{C_s}$ is proportional to an identity matrix.

Putting the estimated normal distributions into Eq. (1) gives

$$\exp\left\{ -\frac{1}{2} \left[\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} - \boldsymbol{\mu}_f \right]^{\mathrm{T}} \mathbf{C}_f^{-1} \left[\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} - \boldsymbol{\mu}_f \right] \right\} \propto$$

$$a(\mathbf{t},\mathbf{s}) \exp\left[-\frac{1}{2} (\mathbf{t} - \boldsymbol{\mu}_t)^{\mathrm{T}} \mathbf{C_t}^{-1} (\mathbf{t} - \boldsymbol{\mu}_t) \right] \exp\left[-\frac{1}{2} (\mathbf{s} - \boldsymbol{\mu}_s)^{\mathrm{T}} \mathbf{C_s}^{-1} (\mathbf{s} - \boldsymbol{\mu}_s) \right]. \tag{4}$$

Solving this for $a(\mathbf{t},\mathbf{s})$ yields

$$a(\mathbf{t},\mathbf{s}) = a_1 \exp\left[\mathbf{b}^{\mathrm{T}} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix}^{\mathrm{T}} \mathbf{C} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} \right] \tag{5}$$

with

$$\mathbf{b} = \mathbf{C}_f^{-1} \boldsymbol{\mu}_f - \begin{pmatrix} \mathbf{C_t}^{-1} \boldsymbol{\mu_t} \\ \mathbf{C_s}^{-1} \boldsymbol{\mu_s} \end{pmatrix} \tag{6}$$

and

$$\mathbf{C} = -\mathbf{C}_f^{-1} + \begin{pmatrix} \mathbf{C_t}^{-1} & 0 \\ 0 & \mathbf{C_s}^{-1} \end{pmatrix}. \tag{7}$$

3

We measure time in units of stomach turnover time, assuming this does not vary much between consumer species. The constant of proportionally a_1 can then be chosen such that the mean of Eq. (5) over all consumer-resource pairs, weighted by resource abundance ρ_r, approximates the actual weighted mean link strength $(Q \sum_r \rho_r)^{-1} \sum_{rc} f_{rc} \rho_r$.

Transformation to trophic traits The symmetric matrix $\mathbf{C}$ defined by Eq. (7) can be represent in terms of a complete orthonormal set of $n + m$ eigenvectors $\mathbf{e}_k$ and the corresponding real eigenvalues λ_k,

$$\mathbf{C} = \sum_{k=1}^{n+m} \mathbf{e}_k \lambda_k \mathbf{e}_k^{\mathrm{T}}. \tag{8}$$

Without loss of generality, we assume that the eigenvalues are ordered such that $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_{n+m}|$.

Now define, for each $k = 1, ..., n + m$, partial eigenvectors $\mathbf{e}_k'$, $\mathbf{e}_k''$ consisting of the first n and remaining m components of $\mathbf{e}_k$, i.e. the components corresponding to resource and consumer (Rossberg et al., 2010) respectively. With this definition:

$$\mathbf{e}_k = \begin{pmatrix} \mathbf{e}_k' \\ \mathbf{e}_k'' \end{pmatrix}. \tag{9}$$

Since the eigenvectors $\mathbf{e}_k$ form a complete orthonormal basis, we can decompose any $(n + m)$-component vector $\mathbf{u}$ as

$$\mathbf{u} = \sum_{k=1}^{n+m} \mathbf{e}_k \left(\mathbf{e}_k^{\mathrm{T}} \mathbf{u} \right). \tag{10}$$

If $\mathbf{t}$ represents the phenotypic traits of some resource item, and $\mathbf{s}$ the phenotypic traits of a labeobarb species, we therefore get, using Eqs. (9) and (10),

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \sum_{k=1}^{n+m} \mathbf{e}_k \left(\mathbf{e}_k'^{\mathrm{T}} \mathbf{t} + \mathbf{e}_k''^{\mathrm{T}} \mathbf{s} \right) = \sum_{k=1}^{n+m} \mathbf{e}_k |\lambda_k|^{-1/2} \left(w^{(k)} - f^{(k)} \right) \tag{11}$$

where

$$w^{(k)} = |\lambda_k|^{1/2} \mathbf{e}_k'^{\mathrm{T}} \mathbf{t}, \quad f^{(k)} = -|\lambda_k|^{1/2} \mathbf{e}_k''^{\mathrm{T}} \mathbf{s}. \tag{12}$$

Putting Eq. (11) into the quadratic term in expression (5) and taking Eq. (8) and the orthogonality of the vectors $\mathbf{e}_k$ into account yields

$$a(\mathbf{t}, \mathbf{s}) = a_1 \exp \left[\sum_{k=1}^{n+m} \mathbf{b}^{\mathrm{T}} \mathbf{e}_k \mathbf{e}_k^{\mathrm{T}} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} - \frac{1}{2} \sum_{k=1}^{n+m} \sigma_k \left(w^{(k)} - f^{(k)} \right)^2 \right], \tag{13}$$

where

$$\sigma_k = -\operatorname{sign} \lambda_k. \tag{14}$$

4

The variable-selection procedure will provide a dimensionality $D \geq 1$ of the trophic niche space to be used or tested. We therefore keep only the first D terms in the sum over $(w^{(k)} - f^{(k)})^2$. This is legitimate, because, as a result of the initial data standardization, the vectors $\mathbf{t}$ and $\mathbf{s}$ have the same variance in all directions. A large eigenvalue of $\mathbf{C}$ therefore immediately implies that trait variations in the direction of the corresponding eigenvector are more important than those directions corresponding to smaller eigenvalues. An eigenvalue $\lambda_k = 0$ corresponds to a direction in trait space that does not matter for the quadratic term at all. It can therefore always be dropped. The case that $\lambda_k = 0$ and $\sigma_k = 0$ is therefore excluded in the following.

To simplify the first sum in Eq. (13), we define the trophic baseline traits as

$$V^* = \sum_{k=D+1}^{n+m} \mathbf{b}^{\mathrm{T}} \mathbf{e}_k \mathbf{e}_k'^{\,\mathrm{T}} \mathbf{t} \quad \text{and} \quad F^* = \sum_{k=D+1}^{n+m} \mathbf{b}^{\mathrm{T}} \mathbf{e}_k \mathbf{e}_k''^{\,\mathrm{T}} \mathbf{s}, \tag{15}$$

accounting for the last $n + m - D$ terms of this sum, and the constants

$$d^{(k)} = \sigma_k |\lambda_k|^{-1/2} \mathbf{b}^{\mathrm{T}} \mathbf{e}_k \quad (k = 1, \ldots, D), \tag{16}$$

encapsulating the information from the first D terms. The constants $d^{(k)}$ are then combined with the raw vulnerability traits to obtain the (final) vulnerability traits as

$$v^{(k)} = w^{(k)} - d^{(k)} \quad (k = 1, \ldots, D). \tag{17}$$

It can be verified using Eqs. (12) to (17) that Eq. (5) is equivalent to

$$a(\mathbf{t}, \mathbf{s}) = a_0 \exp\left[V^* + F^* - \frac{1}{2} \sum_{k=1}^{D} \sigma_k \left(v^{(k)} - f^{(k)} \right)^2 + X \right], \tag{18}$$

with X denoting the remaining $n + m - D$ terms in the sum over k, and

$$a_0 = a_1 \exp\left[-\sum_{k=1}^{D} \frac{(\mathbf{b}^{\mathrm{T}} \mathbf{e}_k)^2}{2\lambda_k} \right]. \tag{19}$$

The approximation that niche space is D-dimensional is equivalent to setting $X = 0$. With $a_{ij} = a(\mathbf{t}_i, \mathbf{s}_j)$, this yields Eq. (2) of the main text.

Formulae (12), (15), and (14) define a transformation from standardized physical/phenotypic traits to trophic traits. For any potential resource-consumer pair characterized by raw physical traits $\tilde{\mathbf{t}}$ and $\tilde{\mathbf{s}}$, the vectors $\mathbf{t}$ and $\mathbf{s}$ entering these equations are obtained as

$$\mathbf{t} = \mathbf{A} \left(\tilde{\mathbf{t}} - \bar{\mathbf{t}} \right), \quad \mathbf{s} = \mathbf{B} \left(\tilde{\mathbf{s}} - \bar{\mathbf{s}} \right), \tag{20}$$

where $\mathbf{A}$, $\mathbf{B}$, $\bar{\mathbf{t}}$, and $\bar{\mathbf{s}}$ are the coefficients of the initial standardization transformation of the data used for determining $\mathbf{b}$ and $\mathbf{C}$. By combining this transformations with the transformation from standardized trait vectors to trophic traits, Eqs. (12), (15), (17), the final result in Eq. (18) can be used to predict trophic link strengths from phenotypic traits.

A.2 Cross-validation of the TTM

To assess, for a given set of physical/phenotypic trait variables, the predictive power of the TTM as obtained above, we performed a two-fold leave-one-out cross validation over all resource categories and labeobarb species. That is, running over all possible resource–labeobarb pairings (p, q), the TTM was parametrized with the given set of phenotypic trait variables, and the predicted link strength of the omitted resource category p with the omitted labeobarb species q using this parametrization was correlated with the actual link strength.

The procedure is complicated by the fact that link strength is more difficult to measure for rare diet items than for abundant ones, leading to larger measurement errors. In computing the correlation coefficient, link strengths were therefore weighted with the square of our proxy for resource abundance, that is, its mean contribution to consumer diets: $\overline{f}_r = Q^{-1} \sum_{c=1}^{Q} f_{rc}$. Hence, the correlation was computed as

$$\rho = \frac{\sum_{p,q}(\hat{a}_{pq} - c_1)(a_{pq} - c_2)\overline{f}_p^2}{\left[\sum_{p,q}(\hat{a}_{pq} - c_1)^2\overline{f}_p^2\right]^{1/2}\left[\sum_{p,q}(a_{pq} - c_2)^2\overline{f}_p^2\right]^{1/2}} \tag{21}$$

with averaged link strengths

$$c_1 = \frac{\sum_{p,q}\hat{a}_{pq}\overline{f}_p^2}{\sum_{p,q}\overline{f}_p^2}, \quad c_2 = \frac{\sum_{p,q}a_{pq}\overline{f}_p^2}{\sum_{p,q}\overline{f}_p^2}. \tag{22}$$

This is very similar to directly computing the correlation between predicted and observed diet proportions, but, because of the quadratic weights entering Eq. (22), not the same.

The value of ρ was used as a measure for the predictive power of the TTM in the cross-validation procedure.

A.3 Selection of phenotypic trait variables

The estimation of the trophic link strength function $a(\mathbf{t}, \mathbf{s})$ requires estimation of $1 + (n + m)/2 + (n+m)^2/2$ real numbers entering the coefficients a_1, $\mathbf{b}$, and $\mathbf{C}$ in Eq. (5) from $R \times Q$ measured link strengths. As a rule of thumb, the numbers of physical resource traits n and phenotypic consumer traits m for which $a(\mathbf{t}, \mathbf{s})$ is estimated should therefore be small compared to both the number of resource categories R and the number of consumer species Q. In our case, many more trait variables have been measured. A systematic method for choosing appropriate sets of trait variables is therefore required. The number of trait variables can be reduced at two levels: immediately at the level of the raw trait variables, or at the level of trophic traits by choosing the number D of dimensions of trophic trait space to be retained. Here, both approaches are combined. We seek combinations of trait variables and a value of D that yield good predictive power according to the cross-validation procedure, i.e., combinations that give a large value of ρ defined by Eq. (21) above.

We conducted an exhaustive search through all combinations of up to $n = 2$ and $m = 2$ trait variables and values of D between 1 and 4. Reasons for limiting the search space to this range were given in the main text: It reduces the risk of unnoticed overfitting (Reunanen, 2003) and we find that dimension 4 makes only a small contribution to improving the fit of our model. There may be procedures that lead to good model fits also for larger values of n, m, or D. To support the conclusions of the present study, however, our conservative choices were sufficient.

References

Reunanen, J. 2003. Overfitting in making comparisons between variable selection methods. The Journal of Machine Learning Research 3:1371–1382.

Rossberg, A. G., Å. Brännström, and U. Dieckmann. 2010. How trophic interaction strength depends on traits — A conceptual framework for representing multidimensional trophic niche spaces. Theoretical Ecology 3:13–24.