Appendices for Translucent windows: how uncertainty in competitive interactions impacts detection of community pattern

Rafael D'Andrea[1,*], Annette Ostling[2,**] & James O'Dwyer[1,†]
[1]Dept. Plant Biology, University of Illinois at Urbana-Champaign, IL, USA
[2]Dept. Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA
[*]rdandrea@illinois.edu, [**]aostling@umich.edu, [†]jodwyer@illinois.edu

A Niche space

Clusters emerge in models where species belong in a niche space, typically a one-dimensional line, and their competitive impact on one another is a function of their mutual distances in that space. The definition of niche space here is abstract and operational: it is the set of aspects of organism performance that affect competition; species sufficiently far apart in that space can stably coexist.

For example, in a model of birds competing for two substitutable food types, say pollen and insects, a species' niche is the amount of time it spends foraging for each food. Assuming the total foraging time is fixed, the niche space is a one-dimensional axis, with points on that axis indicating how long the species spend on one of the resources—the time spent on the other food is then set, as both times need to add to a fixed amount. If there are three substitutable resources, say pollen, insects, and seeds, then niche space is a two-dimensional plane, and so on. The classical MacArthur-Levins (1967) model assumes a continuous resource axis with an indefinite number of food types—say, seeds falling on a size continuum—which in principle makes niche space infinite-dimensional. However, the number of dimensions drops if there are further constraints on the amount of time a bird can spend on each food. MacArthur and Levins (1967) considered the extreme scenario where the relative time spent on one of the resources fixes the time spent on all others: they assumed every species share a resource utilization curve that differs only by what resource it peaks on, i.e. the seed they eat the most. In that case, niche space collapses onto a single axis.

Under this operational definition, a species' niche is not necessarily a resource, measurable trait, or the environment where it lives. Those are only associated with the niche to the extent that they affect competition. In the example of birds competing for seeds falling on a size continuum, the assumption of same-shape resource utilization curves across all species means a species' favorite resource suffices to determine its niche. In this case the niche itself can be equated with the favorite seed size. More generally the niche would be a list of numbers reflecting the amount of time spent on seeds of each size. A relevant trait such as beak size will correlate with its favorite food (Schluter and Grant, 1984), but even in this idealized example of a one-dimensional niche axis the correlation between niche and trait need not be one-to-one, as birds' beaks are also shaped by other functions such as singing (Huber and Podos, 2006).

B Competition for essential resources leads to tail noise, and minimum distances in niche space are a good predictor of competitive interactions

Consider a resource-consumer model with two resource axes. Resources grow logistically in the absence of consumers. Consumers' niche strategies determine their resource preference along each resource axis, but they need resources from both axes to grow. One example is animals that thrive on a diet containing both plant and animal foods. We implement this by having consumer growth rates depend multiplicatively on resources from both axes (Saito et al., 2008):

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mu - mA + \left(\sum_{i=1}^{R} e_x \gamma_{AX_i} X_i \sum_{i=1}^{R} e_y \gamma_{AY_i} Y_i\right) A \tag{1}$$

where A is the abundance of the consumer, m is its intrinsic mortality, μ is the immigration rate, R is the number of resources on each axis, X_i are the abundances of the x-resources, γ_{AX_i} quantifies the consumer preference for the i-th x-resource, and e_x is the conversion efficiency—similarly for the y resources. The product in the consumption derives from the law of mass action in chemistry, in this case representing a reaction that needs three components to occur (the consumer and the two resource types). In our biological context, the term reflects a situation where each resource type offers nutrients not found in the other type, and the consumption of both types are linked together because the consumer will lack the energy to forage if it lacks either.

The corresponding resource equations are

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \nu + X_i r\left(1 - \frac{X_i}{K}\right) - \left(\sum_{N=1}^{S} \gamma_{NX_i} N \sum_{j=1}^{S} \gamma_{NY_j} Y_j\right) X_i \tag{2}$$

$$\frac{\mathrm{d}Y_i}{\mathrm{d}t} = \nu + Y_i r\left(1 - \frac{Y_i}{K}\right) - \left(\sum_{N=1}^{S} \gamma_{NY_i} N \sum_{j=1}^{S} \gamma_{NX_j} X_j\right) Y_i \tag{3}$$

where r is the resource intrinsic growth rate, K the carrying capacity, ν the resource immigration rate, and S the number of consumers. Notice that these equations reflect the abovementioned linked consumption of both resource types.

We note that this type of resource-consumer interaction is not the only way to model colimitation by more than resource; another common approach is Liebig's law of the minimum[1] (Saito et al., 2008). Empirically, consumers show a variety of responses to addition of essential resources, compatible with different models (Harpole et al., 2011). Our model predicts that consumers will respond to addition of either resource in isolation, and will respond maximally to addition of both resources. This has been observed in communities of bacteria (Danger et al., 2008), plankton (Harpole et al., 2011), and plants (Craine and Jackson, 2010).

For simplicity, we set μ, ν, r, K, e_x, e_y identical across species and resources, so that we can focus on the competitive effect of differences in resources preferences. We set the preference of consumer A for resource x_i to be a declining function of the difference between the resource and the consumer's favorite resource x_A along that axis, $\gamma_{AX_i} = \exp(-((x_A - x_i)/w_x)^4)$. Similarly for y-resources.

[1] Our multiplicative model corresponds to chemicals in a vat where the reaction is C + A + B → C + C, i.e. both substances A and B must be present at the same time for C to generate a new C; in contrast, the Liebig scenario corresponds to a reaction like C + A → CA; CA + B → C + C, i.e. where there is a stable intermediate state such that C does not need to encounter A and B at the same time, but still needs both to generate a new C. In biological terms, Liebig's law corresponds to a case where the consumer acquires each resource independently of the other but will not grow without both, whereas in the multiplicative model the consumer's ability to acquire one of the resources depends on having access to the other.

Competition kernel

Consumer-resource interactions do not map trivially to Lotka-Volterra competition coefficients that depend only on trait distances. In general, this mapping might only be valid in the vicinity of an equilibrium (Schoener, 1974, 1976). In fact, expansions of small perturbations around equilibria is the traditional approach to converting resource-consumer models to Lotka-Volterra form, and thus extracting competition coefficients valid in that regime (MacArthur, 1972; Abrams, 1980). This is the approach we adopt here.

The competition kernel, which quantifies the competitive impact of consumers on one another, is comprised of the matrix elements $\alpha_{AB} = -\frac{\partial}{\partial B}\left(\frac{1}{A}\frac{\mathrm{d}A}{\mathrm{d}t}\right)\Big|_{A^*B^*X^*Y^*}$, where the asterisk denotes equilibrium abundances. We calculate it explicitly via linearization around the equilibrium:

$$A(t) = A^* + a(t); \;\; X_i(t) = X_i^* + x_i(t); \;\; Y_j(t) = Y_j^* + y_j(t)$$

Substituting in the consumer equation and discarding higher order terms:

$$\begin{aligned}\frac{\mathrm{d}A}{\mathrm{d}t} &= \mu - (A^* + a)m + (A^* + a)e_x e_y \sum_i \sum_j \gamma_{AX_i}\gamma_{AY_j}(X_i^* + x_i)(Y_j^* + y_j) \\ &\approx -a\frac{\mu}{A^*} + A^* e_x e_y \sum_{ij} \gamma_{AX_i}\gamma_{AY_j}(Y_j^* x_i + X_i^* y_j)\end{aligned}$$

The competition coefficients are then

$$\begin{aligned}\alpha_{AB} &= -\frac{\partial}{\partial B}\left(\frac{1}{A}\frac{\mathrm{d}A}{\mathrm{d}t}\right) \\ &= \delta_{AB}\frac{\mu}{A^{*2}} - e_x e_y \sum_{ij} \gamma_{AX_i}\gamma_{AY_j}\left(Y_j^*\frac{\partial x_i}{\partial B} + X_i^*\frac{\partial y_i}{\partial B}\right) \\ &= \delta_{AB}\frac{\mu}{A^{*2}} + e_x e_y\left((A\mathbf{Y}^*)\cdot\left(-\frac{\partial \mathbf{x}_i}{\partial B}\right) + (A^T\mathbf{X}^*)\cdot\left(-\frac{\partial \mathbf{y}_i}{\partial B}\right)\right) \qquad (4)\end{aligned}$$

where in the last line we used compact matrix notation, with $A_{ij} \equiv \gamma_{AX_i}\gamma_{AY_j}$, and the dot indicating internal product between the vectors.

To proceed we must write $\mathbf{x}$ and $\mathbf{y}$ in terms of the consumer abundances. We do so by assuming resources undergo faster dynamics than consumers:

$$\begin{aligned}0 = \frac{\mathrm{d}X_i}{\mathrm{d}t} &= \nu + r(X_i^* + x_i)\left(1 - \frac{X_i^* + x_i}{K}\right) - (X_i^* + x_i)\sum_N \sum_j \gamma_{NX_i}\gamma_{NY_j}(N^* + n)(Y_j^* + y_j) \\ &= -\frac{\nu}{X_i^*}x_i - rX_i^*\frac{x_i}{K} - X_i^*\sum_N \sum_j \gamma_{NX_i}\gamma_{NY_j}(N^* y_j + Y_j^* n) \\ \rightarrow x_i &= -\frac{1}{\frac{\nu}{X_i^{*2}} + \frac{r}{K}}\sum_N \sum_j \gamma_{NX_i}\gamma_{NY_j}(N^* y_j + Y_j^* n)\end{aligned}$$

We can write this expression in compact matrix notation:

$$\mathbf{x} = P^x\mathbf{y} + Q^x\mathbf{Y}^*$$

where $P_{ij}^x = -\epsilon_i^x \sum_N N^*\gamma_{NX_i}\gamma_{NY_j}$, $Q_{ij}^x = -\epsilon_i^x \sum_N n\gamma_{NX_i}\gamma_{NY_j}$, and $\epsilon_i^x = 1/(\nu/X_i^{*2} + r/K)$. Notice that matrix P^x is a constant whereas matrix Q^x depends on the consumer perturbations.

We get an analogous expression for the y resources:

$$\mathbf{y} = P^y\mathbf{x} + Q^y\mathbf{X}^*$$

with parameters mirroring those of the x resources.

Solving this linear system for $\mathbf{x}$ and $\mathbf{y}$, we get

$$\mathbf{x} = (\mathbf{1} - P^xP^y)^{-1}(P^xQ^y\mathbf{X}^* + Q^x\mathbf{Y}^*)$$
$$\mathbf{y} = (\mathbf{1} - P^yP^x)^{-1}(P^yQ^x\mathbf{Y}^* + Q^y\mathbf{X}^*)$$

Substituting these in Eq. (4) and defining $B^x_{ij} \equiv \epsilon^x_i\gamma_{BX_i}\gamma_{BY_j}$ and $B^y_{ij} \equiv \epsilon^y_i\gamma_{BY_i}\gamma_{BX_j}$, we obtain our competition kernel:

$$\begin{aligned}\alpha_{AB} = \delta_{AB}\frac{\mu}{A^{*2}} &\\ + e_xe_y(A\mathbf{Y}^*)\cdot((\mathbf{1} - P^xP^y)^{-1}(P^xB^y\mathbf{X}^* + B^x\mathbf{Y}^*)) &\\ + e_xe_y(A^T\mathbf{X}^*)\cdot((\mathbf{1} - P^yP^x)^{-1}(P^yB^x\mathbf{Y}^* + B^y\mathbf{X}^*)) &\end{aligned} \tag{5}$$

Results

First, we note that if we unplug the y resources from consumer dynamics, such that only the x resources regulate those species, then differences in x-niche strategies are almost perfect predictors of pairwise competitive relations. In other words, if niche space is one-dimensional, niche strategies fully determine competitive interactions, and we therefore have no process noise.

Next, we add the y resources back in. Fig. S1A shows a community obtained by numerical simulation of this model, with species arranged by their x-niche. Despite niche space being two-dimensional, with both x and y resources modulating species dynamics, the clustered abundance pattern is visible along the x-niche axis, as confirmed by our clustering metric (results summarized in the legend of the plot). Thus, even if the y-niche strategies of these species were entirely unknown, we would still be able to infer niche dynamics from the x-axis alone. The competition kernel, as calculated via Eq. (5), correlates with x-niche distances (Fig. S1C), but the relationship is noisy because of the contribution of the y-niche. The hidden y-niche adds process noise.

To determine whether the noise was higher in the core or tail of the kernel, we fitted a cubic spline to each row of the kernel—which for each species in the community describes the competitive impact of all other species—and compared the expected kernel values from the spline against the observed values. Species pairs were deemed to be in the core or tail based on whether they were below or above the niche distance at which the splined kernel is at the community-level mean. Using the chi-squared statistic $\chi^2_i = \sum_j(\alpha_{ij} - \hat{\alpha}_{ij})^2/\hat{\alpha}_{ij}$, we found that the noise is higher in the tail ($\chi^2_{\text{core}} = 0.04, \chi^2_{\text{tail}} = 0.07$).

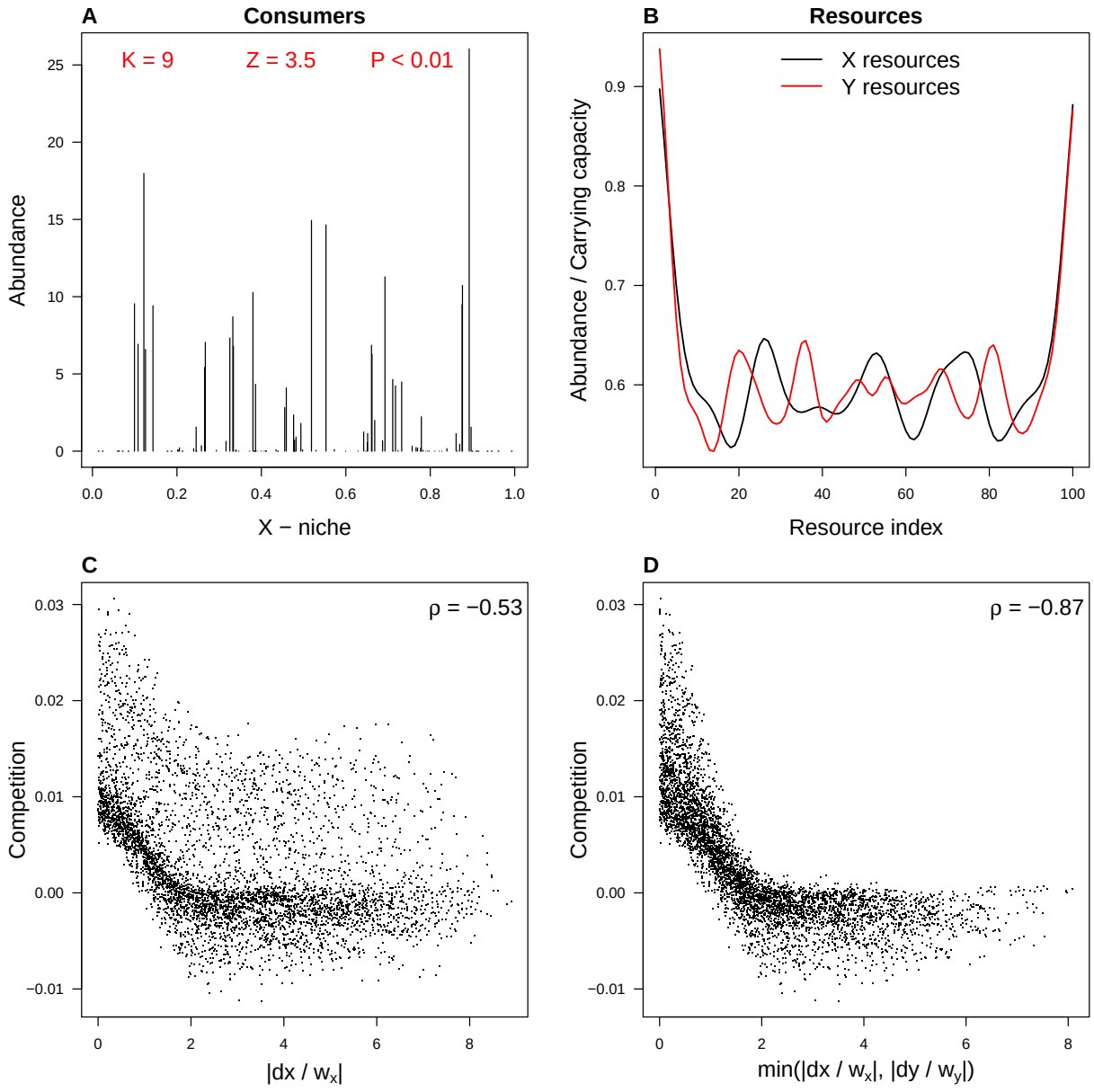
The minimum of the niche distances on either axis is a better predictor of the competition coefficients (Fig. S1D). In particular, the minimum distance is a much better predictor of the kernel's tail than the x-distance: unlike the x-distance, a large minimum distance is a reliable sign of low competitive interactions.

These results support our assumptions in the main text that essential resource axes may lead to noise in the tail of the kernel, and that the pairwise competitive relations are strongly associated with the minimum distance in the full niche space.

To summarize our findings regarding competition for essential resources: when niche space is one-dimensional, niche strategies fully determine competitive interactions—i.e. there is no process noise. When niche space is multidimensional, the map from competition to niche strategies is noisy,

but the minimum distance in niche space is a good approximation. If we have no information on one of the niche dimensions, this process noise is compounded, particularly in the tail.

Figure S1: Generalized process noise. **A**. Species abundances by their x-niche. Legend shows number of clusters (K), z-score of the gap statistic (Z), and corresponding p-value (P); **B**. Resource abundances along each resource axis; **C**. Pairwise competition coefficients calculated from consumer and resource abundances using Eq. (5), plotted against species differences in x-niche strategies. There is a visible negative correlation. Legend shows rank correlation (Spearman's ρ); **D**. Competition coefficients plotted against the minimum between dx and dy (scaled by the width of the respective resource use functions γ). Parameters used in the simulation: $S = R = 100$; $e_x = e_y = 1$; $K = 0.1$; $r = 100$; $m = 1$; $\mu = 10^{-5}$; $\nu = 0.1$; $w_x = 0.11$; $w_y = 0.08$. All species and resources start with identical abundances, and we run the simulation until abundances are no longer changing.



C Competition for available space in two niche dimensions leads to core noise, and Euclidean distances are a good predictor of competitive interactions

Consider a community of plants in a heterogeneous landscape. Plants compete for space, and the landscape is composed of patches, each of which can contain one individual. The local environment in each patch is characterized by two independent indices (for instance, soil humidity and salinity), and different species are optimally adapted to different suites of environmental conditions (i.e. different combinations of humidity and salinity). In this situation, the resource is available space with specific environmental conditions. Niche space is two-dimensional, in the sense that species niche strategies pertain to their affinities for these two independent environmental properties.

We can model such a scenario as follows:

$$\frac{\mathrm{d}N_{ikl}}{\mathrm{d}t} = -mN_{ikl} + b\,T_{ikl}X_{kl}N_i \tag{6}$$

$$X_{kl} = c - \sum_i N_{ikl} \tag{7}$$

where N_{ikl} is the abundance of species i in patches with humidity k and salinity l, and X_{kl} is the number of currently unoccupied patches with those conditions. The first term in Eq. 6 represents death, which for simplicity we assume occurs at the same rate m for all individuals in all patches. The second represents births, which are mediated by the species affinity for the local environmental conditions, T_{ikl}. We assume no dispersal limitation: all individuals of species i contribute offspring equally to all patches (hence the factor N_i in the birth term, as opposed to a weighted combination of N_{ikl}). The birth rate b is the same for all species and patches, but varies over time depending on abundances and site occupancy, such that the total number of births balances the total number of deaths. This portrays a community with a seed bank that enables newly available sites to be quickly colonized, but where births are limited by available space, which is in turn provided by deaths. The number of available patches of humidity k and salinity l is equal to the total number of patches c in the community with those conditions, minus the total occupancy in these sites. For simplicity, we assume c is the same for all environmental conditions.

We can sum Eq. 6 over the patches to find the dynamic equation for species abundances:

$$\sum_{kl} \frac{\mathrm{d}N_{ikl}}{\mathrm{d}t} = \frac{\mathrm{d}N_i}{\mathrm{d}t} = -mN_i + b\tau_i N_i \tag{8}$$

where $\tau_i = \sum_{kl} T_{ikl}X_{kl}$. Note that in order to match total births and deaths, $b = m(\sum_i N_i)/(\sum_i \tau_i N_i)$. This dependence of the birth rate on deaths reflects offspring's need for available gaps to recruit, and provides a stabilizing mechanism that prevents populations from growing indefinitely.

Substituting Eq. 8 in the usual definition of the competition coefficients, $\alpha_{ij} = -\frac{\partial}{\partial N_j}\left(\frac{1}{N_i}\frac{\mathrm{d}N_i}{\mathrm{d}t}\right)\Big|_{\mathbf{N}^*}$, we get

$$\alpha_{ij} = -\left(\frac{\partial \tau_i}{\partial N_j} b + \tau_i \frac{\partial b}{\partial N_j}\right) \tag{9}$$

Before we can calculate those derivatives, we must note that τ_i and b are defined in terms of patch-specific subpopulations. Therefore we must convert $\partial/\partial N_j$ into $\partial/\partial N_{jkl}$. This conversion depends on how an increment to the population N_j is distributed across the different patches. The natural assumption is that it will occur in proportion to the species affinity for the local environment in these

patches, reflecting local filters. In other words, $dN_{jkl} = \frac{1}{\sum_{k'l'} T_{jk'l'}} T_{jkl} dN_j$. The proportionality constant ensures that $\sum_{kl} dN_{jkl} = dN_j$. It then follows that $\frac{\partial}{\partial N_j} = \frac{1}{\sum_{k'l'} T_{jk'l'}} \sum_{kl} T_{jkl} \frac{\partial}{\partial N_{jkl}}$.

Applying the derivatives in Eq. 9 and using Eq. 7, we obtain

$$\alpha_{ij} = b \sum_{kl} T_{ikl} \frac{T_{jkl}}{\sum_{k'l'} T_{jk'l'}} - \frac{m}{\sum_{i'} N_{i'} \tau_{i'}} + \frac{m \sum_{i'} N_{i'}}{(\sum_{i'} N_{i'} \tau_{i'})^2} \left(\tau_j - \sum_{i'} N_{i'} \sum_{kl} T_{ikl} \frac{T_{jkl}}{\sum_{k'l'} T_{jk'l'}} \right) \tag{10}$$

This expression is not particularly illuminating, but we note that it contains terms like $\sum_{kl} T_{ikl} T_{jkl}$, which represent the overlap in environmental preferences between species i and j. This is in the same spirit as the classical work by MacArthur and others (MacArthur and Levins, 1967; MacArthur, 1972).

To complete the model, we must specify the environmental preferences of our different species. Suppose species i has a preferred humidity level x_i and salinity level y_i. Those preferences define the species niche strategy. Its affinity for a patch with humidity x_k and salinity y_l will then be

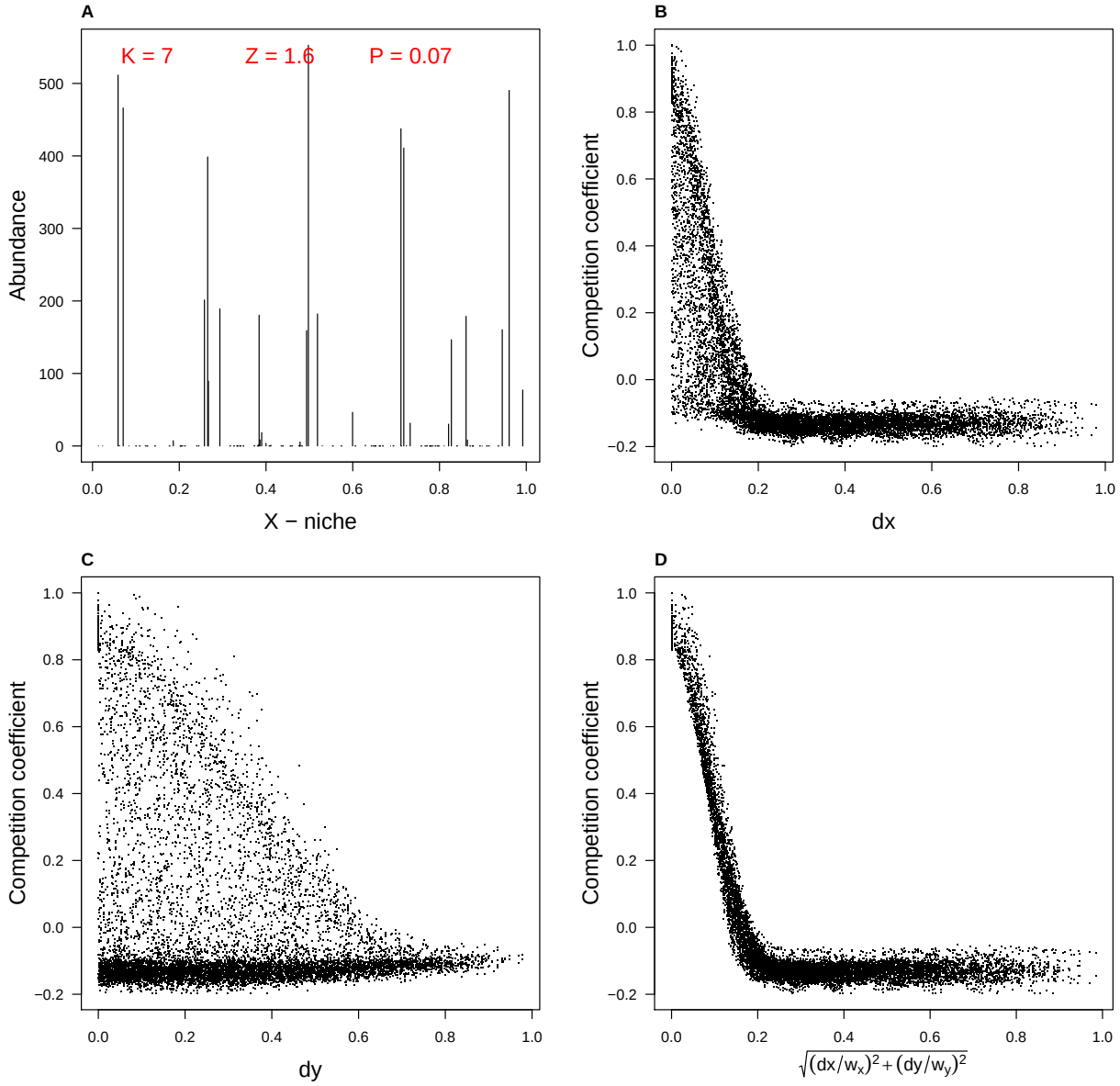
$$T_{ikl} = \exp\left[-\left(\left| \frac{x_i - x_k}{w_x} \right|^2 + \left| \frac{y_i - y_l}{w_y} \right|^2 \right)^2 \right]$$

In words, the affinity is a declining function of the (two-dimensional) difference between the local environment and the species' preferred environment. The weighting factors w_x and w_y set the relative scales of the x- and y-environmental dimensions. For instance in the limit where $w_y \gg w_x$, species are indifferent to the y-dimension relative to the x-dimension. This will occur for example if they have similar tolerance to all salinity levels within the landscape's range. In that case, niche space is one-dimensional.

Results from numerical implementation of this model are shown in Fig. S2. Species cluster along their x-environmental preferences (the x-niche, Fig. S2A), even though competition does not correlate perfectly with differences in x-preference (Fig. S2B). The noise, visible as a scatter in the cloud of points, occurs mostly between species with similar x-preferences. In other words, the noise is greater in the core than in the tail. (Contrast this with the tail-heavy noise from the resource-consumer model of competition for essential resources, Appendix B.) The correlation with differences in y-preference is also imperfect (Fig. S2C). It is only when we plot competition coefficients against the Euclidean distance that we see a tight relationship (Fig. S2D).

In summary, when we ignore either niche dimension, we face noise in the correspondence between competitive interactions and the observed niche strategies. That noise is concentrated in the core of the competition kernel. This is because if all we know is that two species are similar with respect to their x-preference, we may overestimate how strongly they compete because they may compensate for their x-similarity with very dissimilar y-niche requirements. On the other hand, the Euclidean distance in the two-dimensional niche space is an excellent predictor of competition. These results support our simple Lotka-Volterra model for complementary resources presented in the main text.

Figure S2: Competition for available space in two niche dimensions leads to core noise. **A**. Species abundances by their x-niche. Legend shows number of clusters (K), z-score of the gap statistic (Z), and corresponding p-value (P); **B**. Pairwise competition coefficients calculated from consumer abundances using Eq. (10), plotted against species differences in x-niche strategies. There is a visible negative correlation, but it is noisy, especially in the core; **C**. Competition coefficients plotted against differences in y-niche strategies show a similar pattern. **D**. Plotting competition against the Euclidean distance (with dx and dy scaled by the respective weighting factors) resolves the core noise. Parameters used in the simulation: $S = 100$ species; $R = 10$ different environmental levels in each dimension; $m = 0.01$; $w_x = 0.1$; $w_y = 0.38$. Both x- and y-niche strategies range from 0 to 1, as do environmental levels. All species start with identical abundances, totaling the carrying capacity of the community, such that community size does not vary. We then run the simulation until abundances are no longer changing.



D Generalized process noise

In the main text we showed how process noise arises from incomplete access to full niche space in an example where niche space is two-dimensional. Here we generalize this model, assuming an indeterminate number of unseen niche axes and other contributing factors which collectively add noise to the relationship between competition and x-niche differences. We thus set

$$A_{ij} = \exp(-(|x_i - x_j|/w)^4) + \varepsilon_{ij},$$

where the noise term ε_{ij} is a normally distributed random variable with mean 0. We vary the magnitude or amplitude of the noise by tuning the variance parameter. By analogy with the two-dimensional scenario, where we observed that different types of relationship between the niche axes (complementary or essential) lead to more noise in specific regions of the kernel, here we consider three cases: noise can be stronger between species with small x-niche differences, large x-niche differences, or irrespective of x-niche differences. We call these core noise, tail noise, and uniform noise, in reference to the part of the kernel where most of the noise is concentrated. We achieve these different cases by allowing the variance of the noise to depend on distances d_{ij} on the x-niche axis (see Fig. S3B-D for visual representations of each). Because a noisy kernel is unique, we simulated 100 replicates (runs) for each noise type and magnitude to ensure robust results.

We impose nonnegative competition coefficients between all species pairs, and positive intraspecific competition for all species in order to ensure self-regulation and thus avoid unchecked growth. (If the noise term brings the coefficient below zero, we set it to zero if off-diagonal, or redraw it otherwise.) Once all competition coefficients are drawn we rescale the matrix by its mean, $A \to \mu_A \; A/\overline{A}$, in order to ensure that the average competitive impact between species pairs is the same across runs and scenarios. The rescaled mean μ_A affects final community size. We use $\mu_A = 0.2$, which combined with our other parameter choices leads to community sizes between 600 and 35,000 individuals, depending on the noise scenario and magnitude. We use a circular niche axis to avoid edge effects. This isolates the contribution of niche differences to pattern (further testing revealed this assumption is not critical for the validity of our results).

Fig. S3 shows examples of kernels and simulation outcomes in each case. When there is no noise, the kernel is perfectly monotonic on x-niche distances (Fig. S3A), and the resulting community is visibly clustered (Fig. S3E). As process noise is added in the form of core, tail, and uniform noise, the kernel lose monotonicity in the respective region while retaining a statistical trend (Fig. S3B-D). The resulting communities, despite the noise, are still clustered, albeit less distinctively so (Fig. S3F-H). Notice the parallels between core (tail) noise in this generalized process noise model and complementary (essential) niche axes in the two-dimensional model (cf. Fig. S3 and Fig. 3 in the main text).

The probability of clustering decreases as the noise increases (Fig. S4). For the same degree of overall monotonic relationship between competition and x-niche differences, core noise seems to degrade pattern the most, and tail noise the least, with uniform noise in between (cf. blue, red, and green curves in Fig. S4A). Again, this is similar to what we observed in the two-dimensional case regarding complementary and essential niche axes. The core-tail structure predicts the probability of clustering consistently, whether the noise is in the core, tail, or across the kernel (Fig. S4B). This is also similar to our observations for the two-dimensional model (cf. Fig. S4 and Fig. 4 in the main text).

The similarity between the results of our two-dimensional niche model and our generalized noise model indicates that higher niche dimensions add no qualitatively new phenomena to pattern along a single niche axis. The fact that uniform noise shows intermediate behavior between core and tail noise suggests that process noise originating from a combination of complementary and essential

"hidden" niche axes will produce intermediate effects between strictly complementary and strictly essential niche axes.

Figure S3: Generalized process noise. **A-D**: Competition coefficients between a focal species near the center of the niche axis and all other species, plotted against niche difference between them. In the noise-free case (A), the kernel peaks at zero niche difference and decreases monotonically with absolute niche differences. In the noise scenarios the noise is concentrated in the core (B), tail (C), or uniform across the kernel (D). All kernels have the same average, $\overline{A} = S^{-2} \sum_{ij} A_{ij} = 0.2$. Legend numbers show the noise scores according to each of our two metrics, namely Spearman's ρ and $\pi(A)$ (see main text). **E-H**: Species abundance plotted against niche values. (y-axis clipped at $N = 15$ to emphasize abundance structure in the gaps between high-abundance species.) Clusters are quite distinctive in the noise-free scenario (E). In the noisy scenarios (F-H) the pattern is less regular but clustering still occurs—all examples shown are significant at statistical level $p \leq 0.01$. Legend shows the estimated number of clusters and the z-score of the clustering statistic.

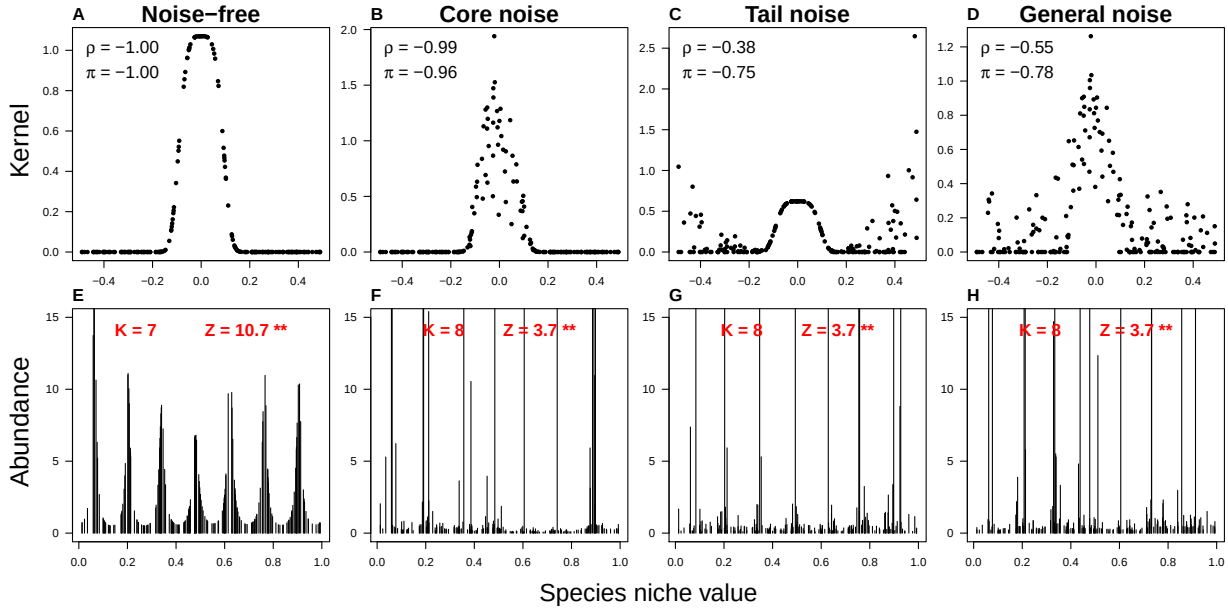
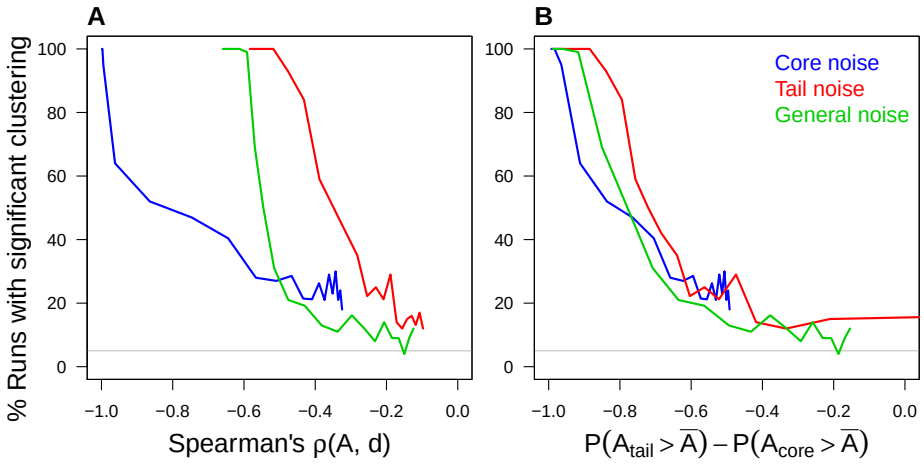


Figure S4: Probability of clustering by magnitude of noise. The probability of clustering is estimated as the percentage of runs of each noise scenario where clustering was statistically significant ($p \leq 0.05$), out of a total 100 replicates each. The noise is measured as the degree of randomness between the kernel and the x-niche axis (process noise) or the trait axis (measurement noise), using two indices: **A**: Monotonic decline of competition with species differences, measured with Spearman's rank correlation coefficient ρ. **B**: Core-tail structure, defined as the difference between the proportion of tail and core elements that exceed the kernel mean. Scenarios shown are process noise with complementary niche axes (blue curves), process noise with essential niche axes (red), and measurement noise (black). Compare with Fig. 3 in the main text.



E Estimating the niche value through multiple proxy traits

In the main text we wrote

$$t_i^a = b^a x_i + c^a + \varepsilon_i^a, \tag{11}$$

where t_i^a is species i's trait value on trait axis a, x_i is species i's niche strategy, b^a and c^a are constants, and ε_i^a is the measurement noise. Written this way, it looks like we are assuming the niche causes the trait. However, this is not so.

In general, species i's niche strategy along a niche axis will be a function of its trait values along n functional trait axes: $x_i = f(T_i^1, T_i^2, \ldots, T_i^n)$. Our example in the main text corresponds to the particular case where that function is linear in all of the trait values, $x_i = k^0 + \sum_{j=1}^n k^j T_i^j$. This expression can be easily inverted to write the species trait along a given trait axis as a linear function of its niche strategy and all other trait values: $T_i^a = 1/k^a x_i - k^0/k^a - \sum_{j \neq a} k^j/k^a T_i^j$. This makes it clear that the measurement error ε_i^a between trait axis a and the niche axis in Eq. 11 stands for the combined contribution of the unmeasured trait axes. Assuming no correlation between trait values along different trait axes, we can find a linear transformation of T_i^a that recovers Eq. 11, with $\underset{i}{\text{Mean}}(\varepsilon_i^a) = 0$ and $\underset{i}{\text{Var}}(\varepsilon_i^a) = \sigma_a^2$. Parameter σ_a reflects the amplitude of the noise between trait axis a and the niche axis x. This noise amplitude will be small when the contribution of the unmeasured traits is small.

In the main text we assume furthermore that the noise is normally distributed, unbiased, and uncorrelated. In other words, $\varepsilon_i^a \sim N(0, \sigma_a)$, $\text{cov}(x_i, \varepsilon_i^a) = 0$, and $\text{cov}(\varepsilon_i^a, \varepsilon_j^b) = 1$ if $i = j$ and $a = b$, and 0 otherwise. In these circumstances, the first principal component of such a set of proxy traits will typically predict species niche values better than most proxy traits individually. An example is shown in Fig. S5.

Consider the special case where the expected value of each trait is the true niche value ($b^a = 1$ and $c^a = 0$, for each $a = 1, 2, \ldots, L$). The coefficient of determination between the niche axis and

trait axis a is defined as $R^2_{x,a} = \frac{\mathrm{cov}^2(x,t^a)}{\sigma_x^2\sigma_a^2}$, where σ_x^2, σ_a^2, and $\mathrm{cov}(x,t^a)$ are respectively the variance in niche values, variance in trait values, and covariance between niche and trait values across species. Substituting we get

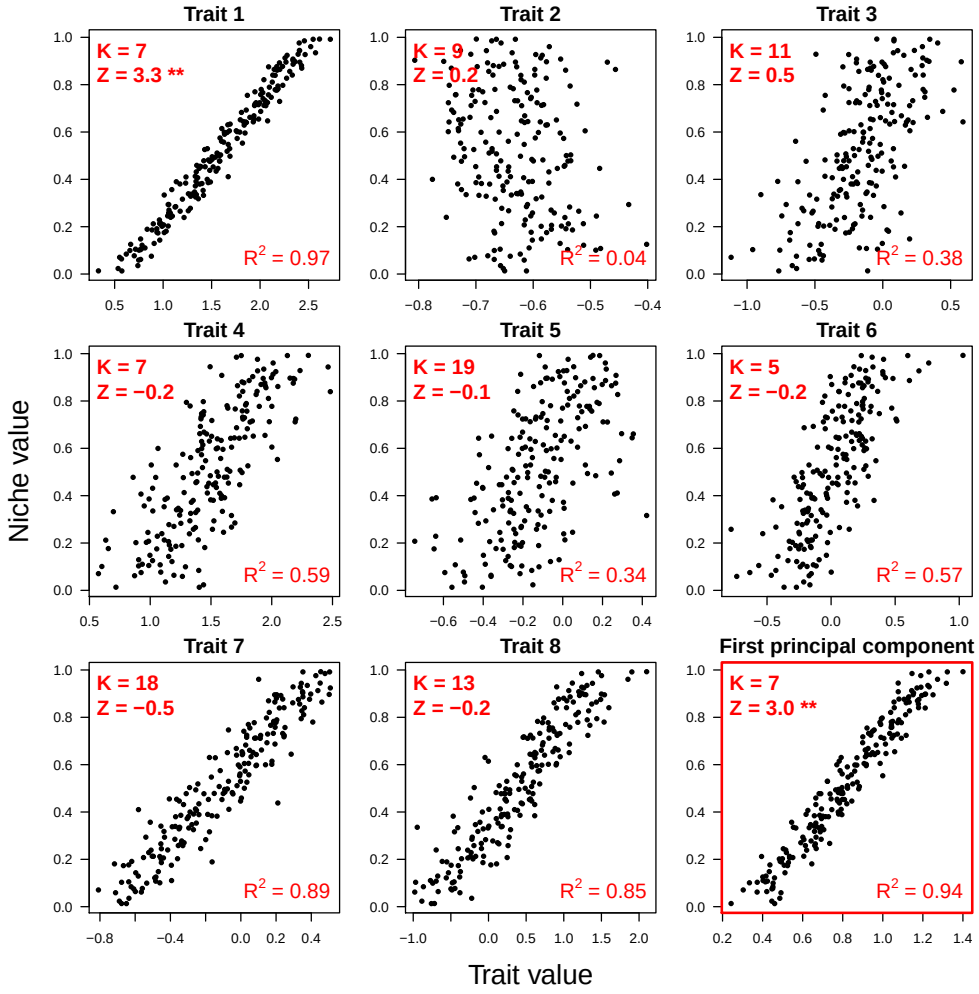
$$\begin{aligned} R^2_{x,a} &= \frac{\mathrm{cov}^2(x, x+\varepsilon^a)}{\sigma_x^2\sigma_a^2} \\ &= \frac{\mathrm{cov}^2(x,x)}{\sigma_x^2(\sigma_x^2+\sigma_a^2)} \\ &= \frac{\sigma_x^2}{\sigma_x^2+\sigma_a^2} \end{aligned}$$

We define a species' consensus trait z_i as the average across its proxy traits: $z_i = 1/L\sum_a t_i^a = x_i + \sum_a \varepsilon_i^a/L = x_i + \overline{\varepsilon_i^a}$. It is also a random variable whose expected value is the true niche value, with the advantage that it has a lower-magnitude error: $\overline{\varepsilon_i^a} \sim N(0, \sqrt{\overline{\sigma_a^2}/L})$. By analogy with the calculation above, the R^2 between the niche values and the consensus trait is $R^2_{x,z} = \frac{\sigma_x^2}{\sigma_x^2+\overline{\sigma_a^2}/L}$. Therefore the consensus trait will have a higher R^2 against the niche axis than a given trait t^a if $\overline{\sigma_a^2}/L < \sigma_a^2$. If all traits have the same noise amplitude, then $\overline{\sigma_a^2} = \sigma_a^2$ and the condition is trivially met for any number of trait axes $L > 1$. When the noise amplitudes differ across traits, the consensus trait will still have a higher coefficient of determination that then "typical" trait with noise amplitude $\overline{\sigma_a^2}$.

In the more general case where b^a and c^a differ across proxy traits, the first principal component replaces the simple trait average as the best predictor of the niche value. In Fig. S5 we show results for an example with 8 proxy traits where the constants b^a and c^a were respectively drawn from a normal distribution centered at 1 and 0: $b^a \sim N(1,\sigma_b)$, $c^a \sim N(0,\sigma_c)$. We used $\sigma_b = \sigma_c = 0.5$. The amplitude of the measurement noise σ_a for each trait was drawn from a uniform distribution between 0.05 and 0.25.

In the example shown in Fig. S5, most trait axes fail to reveal the existing clustering pattern along the niche axis, even those with high R^2 (Trait 7 and Trait 8). In contrast, the first principal component is clustered (clustering z-score = 3.0, p-value < 0.01). Note that in this example the first principal component is a better predictor of the niche than seven of the eight traits ($R^2 = 0.94$), but is a poorer predictor than Trait 1, which is tightly linked to the niche ($R^2 = 0.99$) and is the only clustered trait (z-score = 3.3, p-value < 0.01). In general, the first principal component will typically have a higher coefficient of determination with the niche axis than individual trait axes, and is also more likely to be clustered.

Figure S5: Example showing relationship between the trait value of each species (x-axis) and the niche value (y-axis) for 8 proxy traits that predict the niche value with varying degrees of precision, plus the first principal component (ninth panel). The legend in each panel shows the coefficient of determination between the niche and trait values R^2, the estimated number of species clusters along that trait axis K, and the clustering z-score Z, with the number of stars indicating significance (2 stars for $p \leq 0.01$, 1 star for $p \leq 0.05$, zero stars for $p > 0.05$).



F Clustering metric — the gap statistic via k-means clustering

Here we describe a metric to identify and quantify clustering between species on a niche or trait axis[2], based on a method developed by Tibshirani et al. (2001). The metric receives as input the species abundances and their niche (trait) values, and then estimates the number of clusters that best fits the data. It also provides an index—the gap statistic—that quantifies the degree to which the species assemblage fits that number of clusters better than a set of reference (null) assemblages.

For each candidate number of clusters k within a provided range $[k_{\min}, k_{\max}]$, we calculate the goodness of fit of k clusters to our species assemblage, f_k, as well as the goodness of fit of k clusters to each of the null assemblages $\tilde{f}_{k,i}$, where $i = 1, 2, \ldots, N_{\text{nulls}}$ and N_{nulls} is the total number of

[2]The method also applies in general to a niche or trait space with any number of dimensions, as long as that space allows a meaningful definition of "distance" between species—such as Euclidean distance.

null assemblages. We then record the difference (the gap) between the former and the average of the latter: $\Delta_k = f_k - \overline{\tilde{f}_k}$. The estimated number of clusters $\hat{k}$ is the k that maximizes the gap, $\Delta_{\hat{k}} = \max_k(\Delta_k)$, and the maximum gap itself is the gap statistic, $Gap = \Delta_{\hat{k}}$.

The method is quite general, and both the null assemblages and the measure of goodness of fit must be specified. We generate our null assemblages by shuffling abundances among species, i.e. we randomly redistribute observed species abundances to observed niche (trait) values. This keeps both the set of abundances and the set of traits, while randomizing the map between them.

As for the goodness of fit, following (Tibshirani et al., 2001) we use the k-means clustering algorithm (see MacQueen 1967). This method partitions individuals into a specified number of groups such that the so-called dispersion between individuals within the same group is minimized. Let D_r be the total pairwise squared distance between members of group r, that is, $D_r = \sum_{i,j}(x_i - x_j)^2$, where x_i is the niche or trait value of individual i. The dispersion W_k is then the sum of those numbers across all k groups, $W_k = \sum_{r=1}^{k} D_r$. The k-means algorithm consists in assigning each individual to one of the k groups so that W_k is as low as possible[3,4].

We then set $f_k = \log(1/W_k)$ as the goodness of fit, and feed it into the gap statistic algorithm to estimate the number of clusters k. Notice that it does not make sense to compare the goodness of fit f_k directly across different k's because f_k necessarily increases with k, as the average within-cluster distance is always lower for higher numbers of clusters. By comparing against null assemblages, the gap method finds the biggest increase in goodness of fit *beyond* what is expected from the increase in k. The reason we use $\log(1/W_k)$ rather than simply $1/W_k$ is that the expected increments in $1/W_k$ with increasing k are multiplicative rather than additive. See (Tibshirani et al., 2001) for more details. Our metric is then defined as $Gap = \max_k\Big(\log(1/W_k) - \overline{\log(1/\tilde{W}_k)}\Big)$, and the value of k at which the quantity in the parenthesis is maximal is the estimated number of clusters $\hat{k}$.

If we knew the expected distribution of this statistic under the null hypothesis where species niches (traits) and abundances are unrelated to each other, we could stop here, as the value of Gap would tell us about both degree of clustering and statistical significance. Because we do not know that distribtution, we must compare it with a null set of values. We therefore calculate the metric on each of our null assemblages, and from that we obtain a z-score and a p-value. Those are defined as $z = (Gap - \overline{\tilde{Gap}}) \,/\, \sigma(\tilde{Gap})$ and $p = \frac{1}{N_{\text{nulls}}} \sum_{i=1}^{N_{\text{nulls}}} I(\tilde{Gap}_i \geq Gap)$ where $\tilde{Gap}_i$ is the gap statistic obtained for the i-th null assemblage, $\overline{\tilde{Gap}}$ and $\sigma(\tilde{Gap})$ are the mean and standard deviation across those values, and I is the index function, equal to 1 if its argument is true and zero otherwise. The z-score and p-value quantify respectively the degree to which the gap statistic returned by our species assemblage exceeds null expectations, and whether that result is statistically significant.

G Kernel structure in the trait-noise scenario

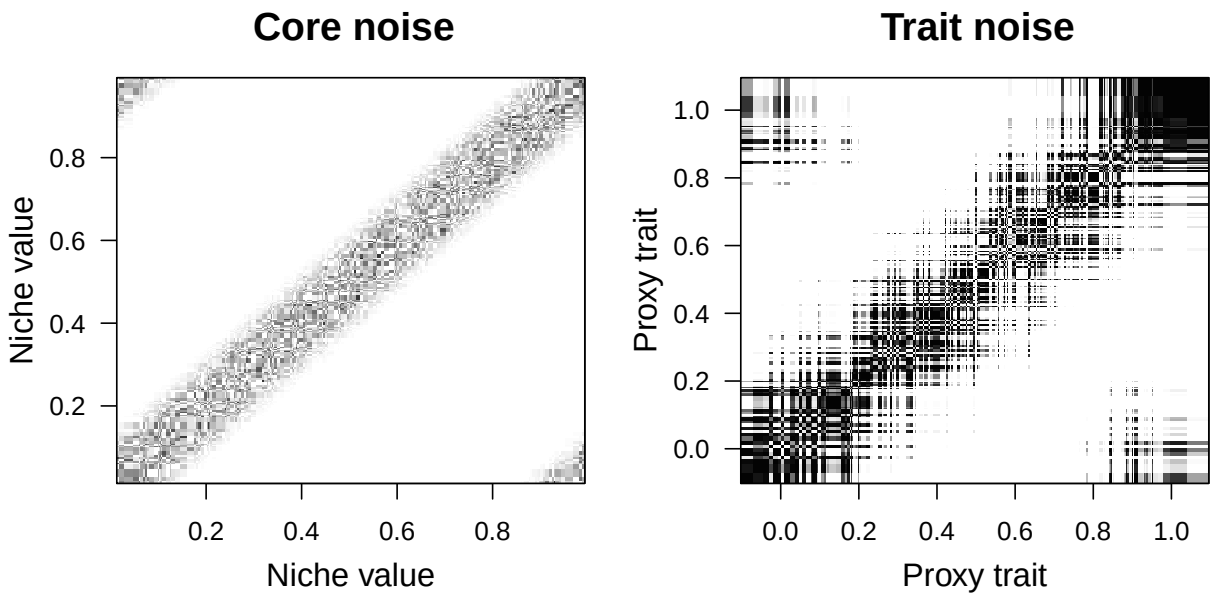
In the trait-noise scenario, the competition coefficients plotted against proxy trait values resemble the core-noise kernel, with most of the noise occurring for pairs with small trait difference (compare Figs. 3A and 5H in the main text). However, a closer look reveals that the measurement-noise

[3]Dispersion can be defined in different ways, depending on how one wishes to weigh the number of members per group. While Tibshirani et al. (2001) choose to scale the squared distances by the number of individuals in the group, Yan and Ye (2007) argue that scaling it by the number of pairs in the group is a a better choice in some cases. In our study we saw no qualitative difference between those different scaling options, and opted to use the simpler unweighted sum of squared distances.

[4]In our model, all conspecific individuals have the same niche (trait) value, and therefore necessarily belong in the same cluster. Species are therefore trivial clusters of individuals. Ecologists are of course interested in clustering beyond the obvious grouping of individuals into species.

kernel is more structured (Fig. S6): unlike the core-noise scenario, the matrix is symmetric and the noise is autocorrelated. We note that there is a permutation of the rows and columns of the competition matrix in the trait noise scenario that restores the noise-free matrix. This is not the case with any of the kernel noise scenarios presented in the main text.

Figure S6: Comparison between the competition kernel from the core-noise scenario and the trait-noise scenario. **Left**: Core-noise competition coefficients between species pairs plotted against their respective niche values, with darker shades of gray representing stronger competition. Coefficients tend to be higher when species have similar niche values (the top-left and bottom-right corners are continuations of the diagonal, as the niche axis is circular). The noise is manifested in the fine-grained pixelation of the grey band, and lacks any visible structure. **Right**: Trait-noise competition coefficients plotted against trait values. As in the core-noise scenario, the noise is concentrated in the diagonal band where competition is stronger. However, unlike the core-noise scenario, the matrix is symmetric and the pixel sizes are larger, indicating autocorrelation in the noise.



References

Abrams, P. (1980). Are Competition Coefficients Constant? Inductive Versus Deductive Approaches. *The American Naturalist*, 116(5):730–735.

Craine, J. M. and Jackson, R. D. (2010). Plant nitrogen and phosphorus limitation in 98 North American grassland soils. *Plant and Soil*, 334(1):73–84.

Danger, M., Daufresne, T., Lucas, F., Pissard, S., and Lacroix, G. (2008). Does Liebig's law of the minimum scale up from species to communities? *Oikos*, 117(11):1741–1751.

Harpole, W. S., Ngai, J. T., Cleland, E. E., Seabloom, E. W., Borer, E. T., Bracken, M. E., Elser, J. J., Gruner, D. S., Hillebrand, H., Shurin, J. B., and Smith, J. E. (2011). Nutrient co-limitation of primary producer communities. *Ecology Letters*, 14(9):852–862.

Huber, S. K. and Podos, J. (2006). Beak morphology and song features covary in a population of Darwin's finches (Geospiza fortis). *Biological Journal of the Linnean Society*, 88(3):489–498.

MacArthur, R. H. (1972). *Geographical ecology; patterns in the distribution of species*. Harper & Row.

MacArthur, R. H. and Levins, R. (1967). The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*, 101(921):377–385.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *5th Berkeley Symposium on Mathematical Statistics and Probability 1967*, 1(233):281–297.

Saito, M. A., Goepfert, T. J., and Ritt, J. T. (2008). Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography*, 53(1):276–290.

Schluter, D. and Grant, P. R. (1984). Determinants of Morphological Patterns in Communities of Darwin's Finches. *The American Naturalist*, 123(2):175–196.

Schoener, T. W. (1974). Resource partitioning in ecological communities. *Science (New York, N.Y.)*, 185(4145):27–39.

Schoener, T. W. (1976). Alternatives to Lotka-Volterra competition: models of intermediate complexity. *Theoretical population biology*, 10(3):309–33.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:411–423.

Yan, M. and Ye, K. (2007). Determining the number of clusters using the weighted gap statistic. *Biometrics*, 63(4):1031–1037.