Clark, A.T., Neuhauser, C. (2018):
Harnessing uncertainty to approximate mechanistic models of interspecific interactions
Theor. Popul. Biol. **123** , 35 - 44

http://dx.doi.org/10.1016/j.tpb.2018.05.002

Harnessing Uncertainty to Approximate Mechanistic Models of Interspecific Interactions

Running Head: Harnessing Uncertainty

Abstract Word Count: 163
Main Text Word Count: 5897
Number of References: 50
Number of Figures: 5

Key words: Lotka-Volterra competitive equations; process noise; model uncertainty; interspecific competition; model abstraction; interspecific tradeoff

Adam Thomas Clark[1,2,3,4*] and Claudia Neuhauser[1,5,6]
[*]Corresponding author adam.tclark@gmail.com; [1]University of Minnesota, Dept. of Ecology, Evolution, and Behavior, 1987 Upper Buford Circle, Saint Paul, MN 55108, USA; [2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Synthesis Centre for Biodiversity Sciences (sDiv), Deutscher Platz 5e, 04103 Leipzig, Germany (current address); [3]Department of Physiological Diversity, Helmholtz Center for Environmental Research (UFZ), Permoserstrasse 15, Leipzig 04318, Germany; [4]Leipzig University, Ritterstrasse 26, 04109 Leipzig, Germany; [5]University of Minnesota, University of Minnesota Informatics Institute, Minneapolis, MN, 55455, USA; [6]Division of Research, University of Houston, Houston TX 77204 (current address).

Highlights:

- Lotka-Volterra models are commonly used to study interspecific interactions
- However, these models can fail when interaction strengths among species are uncertain
- We show that underlying mechanisms constrain relationships among model parameters
- We approximate these constraints by deriving expected covariance among parameters
- We find that covariance significantly improves predictions in diverse communities

Abstract:

Because the Lotka-Volterra competitive equations posit no specific competitive mechanisms, they are exceedingly general, and can theoretically approximate any underlying mechanism of competition near equilibrium. In practice, however, these models rarely generate accurate predictions in diverse communities. We propose that this difference between theory and practice may be caused by how uncertainty propagates through Lotka-Volterra systems. In approximating mechanistic relationships with Lotka-Volterra models, associations among parameters are lost, and small variation can correspond to large and unrealistic changes in predictions. We demonstrate that constraining Lotka-Volterra models using correlations among parameters expected from hypothesized underlying mechanisms can reintroduce some of the underlying structure imposed by those mechanisms, thereby improving model predictions by both reducing bias and increasing precision. Our results suggest that this hybrid approach may combine some of the generality of phenomenological models with the broader applicability and meaningful interpretability of mechanistic approaches. These methods could be useful in poorly understood systems for identifying important coexistence mechanisms, or for making more accurate predictions.

1. Introduction:

Of all of the ecological models used to describe species dynamics, none is so ubiquitous as the Lotka-Volterra competitive equations (Lotka 1932, MacArthur and Levins 1967, Wangersky 1978). These equations approximate interactions among species as a series of simple, linear functions describing the per-capita effect of each species on other species' growth rates. Lotka-Volterra competition models are therefore considered the simplest possible abstraction of competition (MacArthur 1970, Tilman 1982). They are typically one of the first models of interspecific interactions taught to students, and underpin an enormous span of ecological theory, ranging from the criteria for coexistence and competitive displacement (MacArthur and Levins 1967, Chesson 1990, 2000), to relationships among community diversity, productivity, and stability (May 1973, Lehman and Tilman 2000, Loreau 2004).

Because of their relatively simple nature, Lotka-Volterra competition models make no specific assumptions about the mechanisms underlying competitive interactions – i.e. they are entirely "phenomenological." Models that include specific mechanisms necessarily restrict the kinds of interactions that can take place among species. Because Lotka-Volterra models include no such restrictions, they can be parameterized in ways that approximate any combination of underlying mechanisms, at least locally around equilibrium (MacArthur 1970). This can be quite valuable, as it means that Lotka-Volterra models can be applied to a broad array of systems, as opposed to more mechanistic approaches for which the "correct" underlying mechanism may need to be identified in order to make accurate predictions (but see Schaffer 1981). Indeed, there are many classic examples of ecological systems in which dynamics are consistent with the qualitative expectations of Lotka-Volterra models, including aquatic microbial communities (Gause 1934), flour beetles (Park 1936), and warblers (MacArthur 1958), and even some examples where quantitative estimates from these models accurately predict abundances in multi-species communities (Vandermeer 1969, Carrara et al. 2015).

Despite this broad applicability, there are substantially more examples of cases where Lotka-Volterra models have failed to accurately predict outcomes of competition. This frequently manifests as a phenomenon known as “non-additivity” or “higher-order interactions,” in which the effects of two species on one another’s growth rates vary depending on the presence of a third species, or of some other factor such as environmental variation (Wilbur 1972, Roxburgh and Wilson 2000, Dormann and Roxburgh 2005, Weigelt et al. 2007, Michalet et al. 2015, Bairey et al. 2016, Mayfield and Stouffer 2017, Levine et al. 2017, Grilli et al. 2017). While predictions can occasionally be improved by augmenting models to include parameters that describe these higher-order interactions (Gause 1934, Wilbur 1972, Weigelt et al. 2007, Mayfield and Stouffer 2017), predictive power generally remains low, and the changes in model form make them more difficult to generalize to other systems.

The poor performance of Lotka-Volterra models in many real-world ecological systems is hardly surprising. A hallmark of ecological data is that it is highly uncertain – because of both observation error, which results from imperfect measurement of a system, and process noise, which describes natural variability in the characteristics of the system. Because Lotka-Volterra models abstract systems into a series of linear interactions regardless of the underlying mechanism, they are disproportionately influenced by this uncertainty, particularly in systems with many interacting species (Dormann 2008, Palamara et al. 2016). For coexistence to be stable in the Lotka-Volterra framework, a primary requirement is that the dominant eigenvalue of the Jacobian matrix (i.e. the most positive eigenvalue) must have a negative real part when evaluated at equilibrium (n.b. stability also requires the system to be “feasible”, i.e. all species must have abundance greater than or equal to zero). This requires that the matrix summarizing all pairwise interactions among coexisting species be of full rank (i.e. all rows and columns must be linearly independent) (Levin 1970, Chesson 1990, Haygood 2002, Meszéna et al. 2006). However, even infinitesimal additions of error or noise will lead to all interactions "appearing" as though they are linearly independent, and can distort the original model in ways that are difficult to predict or reverse (Anderson et al. 2010).

In contrast, constraints caused by particular mechanisms of interaction typically limit the kinds of effects that stochasticity can impart. For example, in a model of competition for a single limiting resource, variability in a species' competitive ability for that resource will alter its interactions with all other species in the community in similar ways. Furthermore, coexistence in systems with mechanistic constraints usually implies the existence of interspecific tradeoffs (i.e. strategies that unavoidably come at a cost to one another) (Tilman 1990, 2011, Chesson 2000). These relationships cause species to differentiate along several trait axes simultaneously and prevent any single species from dominating under all possible circumstances. Provided that error and noise are sufficiently small or constrained that they do not drive species away from these tradeoffs, then predictions from mechanism-based models can be relatively robust to uncertainty (Clark et al. 2018). Importantly, there is abundant evidence for such tradeoffs in many empirical systems (Wright et al. 2004, Litchman and Klausmeier 2008, Reich 2014).

In many cases, observation error can be mitigated through proper replication and experimental controls (but see *3.3. Some notes on observation error*). In contrast, process noise results from variation that is inherent to the system itself (e.g. within-species trait variability, environmental heterogeneity), and cannot be so easily mitigated. However, because process noise actually influences the dynamics of systems, it may be possible to use this variation to learn about the system's dynamic structure. This is, after all, the primary insight behind regression analysis: if variability is shared among several components of a system, this may indicate that they are meaningfully related. Importantly, because individual mechanisms can only recreate a subset of the types of interactions that can be expressed in Lotka-Volterra models, this suggests that the mechanisms that actually underlie coexistence in a system should leave behind a specific "signature" in the parameters of the Lotka-Volterra model (Tilman 1982).

One of the simplest ways to identify this mechanism-driven correspondence among parameters is through covariance. Covariance measures the degree of correspondence between two variables (n.b. correlation is simply covariance standardized by univariate variance). As we will show, the covariance among interaction parameters in the Lotka-Volterra competitive equations can be derived directly from a hypothesized underlying mechanistic model, and doing so often requires relatively little information about the specific parameter values and functional form of this mechanistic model. Once calculated, this covariance can be used to constrain the Lotka-Volterra system. Though covariance alone will not capture the full dynamics of the underlying mechanistic model, it may be that it can reintroduce some aspects of the mechanistic structure and interspecific tradeoffs that regulate coexistence in real-world systems. Modeling covariance could therefore be useful for improving predictions, and identifying signatures left behind by influential coexistence mechanisms. This method is similar to that employed in spatial moment models, which utilize information about the spatial covariance of observed patterns to analytically approximate species dynamics in spatial systems (Bolker and Pacala 1999, Detto and Muller-Landau 2013).

In this manuscript, we demonstrate the potential utility of harnessing covariance to detect and integrate components of mechanistic models into the phenomenological Lotka-Volterra framework. As a worked example, we use a model developed by MacArthur (1970), which relates a mechanistic resource competition model to the classical Lotka-Volterra framework. After *(i)* introducing this model, we will use it to *(ii)* demonstrate how process noise in the mechanistic parameters propagates through the system and changes model predictions, *(iii)* derive the covariance relationship among phenomenological parameters that results from this process noise, *(iv)* show how this information can reduce distortion of the model, and *(v)* explore how this theoretical insight can be applied in the analyses of empirical data to identify underlying mechanisms and improve model predictions.

2. Model and Results:

2.1. MacArthur's resource competition model

MacArthur's 1970 resource competition model relates a mechanism-based model of interspecific competition for perfectly substitutable resources to the classical Lotka-Volterra competition framework (MacArthur 1970, Tilman 1982). MacArthur's simplifying assumption was to propose that resource dynamics occurred much faster than consumer dynamics, which allowed resource concentrations to be estimated as a simple function of consumer abundance. Because of this simplification, the model can be re-written in a way that is mathematically identical to the classical Lotka-Volterra competitive equations, despite the fact that it includes parameters that can be interpreted mechanistically (though see caveats in *3.1. Potential limitations*). In this model, interspecific interaction terms of the Lotka-Volterra competitive equations (a_{ij}, which describe the per-capita effect of species j on the growth rate of species i) can be directly related to a function of the mechanistic parameters (c_{il}, which describe the ability of species i to acquire resource l). Note that we use the notation of Chesson (1990).

MacArthur's resource competition model has been extensively developed in many subsequent studies (e.g. Schoener 1974, Chesson 1990, 2000, Haygood 2002), and has been found to be particularly tractable for a number of reasons. First, given any combination of n species and $l_{max} \geq n$ limiting resources for which an "interior" equilibrium exists (i.e. all species have positive abundances), the model is globally stable, meaning that it approaches a single equilibrium from any starting point (Haygood 2002). Second, competition is symmetric between species (i.e. $a_{ij} = a_{ji}$), which provides a specific structure to competitive interactions and reduces the number of parameters. Lastly, and most importantly for our purposes, the model provides a simple link between "mechanistic" parameters describing species consumption rates (c_{il}), and "phenomenological" parameters describing the outcomes of competitive interactions (i.e. a_{ij}). These parameters are typically grouped into matrices $\boldsymbol{c}$ and $\boldsymbol{a}$, respectively.

To facilitate the process of tracking error propagation, we further simplify MacArthur's original model (full derivation of our model, and a more detailed discussion of how the parameters relate to other ways of expressing the Lotka-Volterra system, are available in Appendix A.I in the supplement). In our framework, $\boldsymbol{c}$ and $\boldsymbol{a}$ are related following

$$a_{ij} = \sum_{l=1}^{l_{max}} c_{il}c_{jl} \qquad \text{Eq. (1)}$$

Dynamics in the abundance of species i, X_i, depend only on the species growth rate, b_i, the maximum amount of resource taken up in the absence of other competitors, k_i, and competitive interactions with other species, following the form

$$\frac{1}{X_i}\frac{dX_i}{dt} = b_i\left(k_i - \sum_{j=1}^{n} a_{ij}X_j\right) \qquad \text{Eq. (2)}$$

Eq. (2) is mathematically identical to the classical Lotka-Volterra competitive equations, though with mechanistic constraints on the values of $\boldsymbol{a}$. Note, however, that we use a different parameterization in Eq. (2) than is presented in many ecological textbooks. Unlike these "classical" forms, carrying capacities and interaction coefficients in our model are not standardized by the strength of species self-inhibition. Thus, the classical carrying capacity K_i (i.e. species abundance in the absence of competitors) is equal to k_i/a_{ii} in our model's parameterization, and self-inhibition in our model a_{ii} (i.e. the effect of species *i* on itself) is not necessarily equal to one (unlike the classical α terms, for which $\alpha_{ij} = a_{ij}/a_{ii}$) (Chesson 2000).

For a stable equilibrium to occur in this system, there must be a tradeoff among species consumption vectors such that each species is able to access a unique combination of resources, which implies $l_{max} \geq n$ and that the columns of $\boldsymbol{c}$ must be linearly independent (MacArthur 1970, Chesson 1990). Because $\boldsymbol{a}$ is composed of pairwise products of $\boldsymbol{c}$, satisfying these conditions ensures that all rows and columns of $\boldsymbol{a}$ are also linearly independent (i.e. that matrix $\boldsymbol{a}$ is of "full rank") (Chesson 1990). Recall, however, that not all communities that meet these criteria will coexist stably, as the equilibrium must also be interior (Chesson 1990, Haygood 2002). A discussion of the full stability criteria is available in Appendix A.II in the supplement.

2.2. Effects of process noise on model predictions

We define "process noise" as variation in the mechanistic consumption parameters of $\boldsymbol{c}$ that alters the consumption vectors themselves, rather than merely changing our ability to accurately measure these parameters (i.e. in contrast to observation error). This variation might be caused by genetic differences, trait plasticity, spatial heterogeneity, or any other such process that drives within-species trait differences. Process noise therefore can lead to changes in species dynamics, equilibrium population abundances, and even persistence.

As demonstrated in Eq. (1), process noise also leads to variation in $\boldsymbol{a}$. Consider a system where process noise is normally and independently distributed around each of the elements of $\boldsymbol{c}$, with standard deviation σ_c. Based on the mechanistic relationship between $\boldsymbol{c}$ and $\boldsymbol{a}$ in Eq. (1), process noise in $\boldsymbol{c}$ can be analytically related to variation in the terms of $\boldsymbol{a}$ as

$$\sigma_{a_{ii}}^2 = 2{\sigma_c}^2\left(l_{max}{\sigma_c}^2 + 2a_{ii}\right) \qquad \text{Eq. (3a)}$$

$$\sigma_{a_{ij}}^2 = \sigma_c{}^2\left(l_{max}\sigma_c{}^2 + a_{ii} + a_{jj}\right) \quad \text{Eq. (3b)}$$

where σ_{aii} and σ_{aij} are the standard deviation in the terms of $\boldsymbol{a}$ resulting from σ_c, i and j are both assumed to be less than n, and $i \neq j$. These terms increase roughly as a linear function of σ_c (Fig. 1a), and are of a similar magnitude as σ_c regardless of community size when scaled by their corresponding mean values a_{ii} and a_{ij}, respectively (Fig. 1b).

In Appendix B.I in the supplement, we include full derivations for these terms based on the mathematical properties of means and variances. However, intuitively, one can also think of Eqs. (3a-b) as the effect of a change in the consumption parameters $\boldsymbol{c}$ on the interaction matrix $\boldsymbol{a}$. Recall that variance is the average squared distance between a random variable and its mean. Thus, the square root of Eqs. (3a-b) represent the change in the interaction coefficients that is expected from the average observed deviation in the elements in $\boldsymbol{c}$. In other words, just as σ_c represents the mean deviation in species consumption rates caused by process noise, $\sigma_{a_{ii}}$ and $\sigma_{a_{ij}}$ represent the corresponding mean deviation in self-limitation or competitive effects, respectively.

For simplicity, let us suppose that we can perfectly measure all of the parameters in this system, and that the terms k_i and b_i are not subject to any kind of process noise. In this case, variability in the outcomes of competitive interactions can be entirely attributed to σ_c and its effects on the realized values of $\boldsymbol{c}$ and $\boldsymbol{a}$. For example, consider a case of two species competing for two limiting resources that are able to coexist in the absence of process noise (i.e. $\sigma_c = 0$). If we increase process noise in this system, the increasing variability in $\boldsymbol{c}$ and $\boldsymbol{a}$ will increase the probability of competitive exclusion between the two species (Fig. 2a). Now, suppose that we attempt to predict the outcome of competition by directly measuring the components of $\boldsymbol{a}$. Most commonly, this is accomplished by measuring a_{ii} and a_{jj} under circumstances where only one species is present, and a_{ij} and a_{ji} under circumstances where both species are jointly present. In this case, we would find something curious: even if we perfectly measure the components of ***a*** in the presence of process noise, we will overestimate the frequency of competitive exclusion (Fig. 2b). This is because we fail to take into account the correlation among the components of ***a*** that is caused by the mechanistic constraints and tradeoffs that govern coexistence in MacArthur's model. However, note that if we account for covariance among these terms (for example, by measuring the terms at the same time in the same system), we are able to much more closely match the outcome expected from the mechanistic model.

2.3. Model covariance relationship

The correlations between elements of ***a*** impose limitations on their structure, thereby restricting the very general form of the Lotka-Volterra competitive equations to behave more like the specific mechanisms posited in the MacArthur model. As the number of competing species grows, these relationships become increasingly complex and influential. It is therefore helpful to derive the expected covariance relationship among all terms in ***a*** based on their mechanistic relationships in the MacArthur model. This reveals six classes of elements, including the two variance terms in Eqs. (3a-b)

$$\mathrm{cov}\left(a_{ij}, a_{ji}\right) = {\sigma_c}^2\left(l_{max}{\sigma_c}^2 + a_{ii} + a_{jj}\right) \qquad \text{Eq. (3c)}$$

$$\mathrm{cov}\left(a_{ij}, a_{ii}\right) = 2{\sigma_c}^2 a_{ij} \qquad \text{Eq. (3d)}$$

$$\mathrm{cov}\left(a_{ij}, a_{ik}\right) = {\sigma_c}^2 a_{jk} \qquad \text{Eq. (3e)}$$

$$\mathrm{cov}\left(a_{ij}, a_{km}\right) = 0 \qquad \text{Eq. (3f)}$$

where i, j, k, and m represent any four distinct species (i.e. $i,j,k,m < n$, and $i \neq j \neq k \neq m$). Because of symmetry in ***a***, these classes include several kinds of associations (e.g. cov(a_{ij}, a_{ii}) = cov(a_{ij}, a_{jj})). Full derivations and details are available in Appendix B.I in the supplement.

Note that though the relationships in Eqs. (3a-f) constrain the parameters of $\boldsymbol{a}$, they are not guaranteed to perfectly mimic the underlying relationships imposed by the mechanistic model. This is because we include only the first two "statistical moments" describing the effects of process noise (i.e. mean and variance), and thus we omit more complex aspects of statistical distributions, such as asymmetry or "fat tails." Thus, just as models that ignore higher-order interactions will fail to characterize context-dependent changes in the elements of $\boldsymbol{a}$, ignoring higher-order statistical moments (e.g. skew or kurtosis) can lead to prediction error if stochasticity in a system is not normally distributed.

2.4. Testing the covariance model

We can demonstrate how the covariance relationships capture some aspects of the mechanistic model by simulating three types of models with added process noise: *(i)* the mechanistic model based on Eqs. (1-2); *(ii)* an uncorrelated model, in which variance in $\boldsymbol{a}$ is calculated following Eqs. (3a-b), but covariance is ignored; and *(iii)* a covariance model, which accounts for all variance and covariance relationships among components of $\boldsymbol{a}$ following Eqs. (3a-f). Thus, the mechanistic model represents the "true" process, the uncorrelated model represents a sampling design that ignores correlations among species interaction coefficients, and the covariance model demonstrates potential improvements in predictions resulting from properly accounting for the mechanistic model's effects on the correlation structure of $\boldsymbol{a}$.

To demonstrate the effects of community diversity and noise on these three models, we simulated each of them across five sizes of communities (n = 2, 4, 6, 8, 10), and four levels of process noise (σ_c = 0.01, 0.02, 0.05, 0.1). In each scenario, we generated species consumption vectors by drawing points from a tradeoff, such that $\Sigma_l \, c_{il} = l_{max}$ for all species (i.e. all points in $\boldsymbol{c}$

fell along a single $l_{max} - 1$ dimensional surface). We randomly selected points from along this tradeoff surface using a random normal variable with mean equal to 1 and standard deviation equal to 0.1 (n.b. we applied a logit transformation to ensure positive values – see Appendix C in the supplement for details). Note, $\sigma_c/0.1$ therefore corresponds to the standard deviation in the elements of $\boldsymbol{c}$ caused by process noise, relative to the standard deviation in consumption parameters observed between species. This tradeoff ensured that increased effectiveness of foraging for one resource (i.e. larger values of c_{il}) implies decrease effectiveness for others (i.e smaller values of c_{ik}), thereby helping to maintain coexistence in the mechanistic model. As discussed above, such a tradeoff is consistent with both empirical observations (Wright et al. 2004, Reich 2014) and theoretical expectations (Tilman 1990, 2011, Clark et al. 2018) for communities of coexisting species.

For all scenarios, we set $l_{max} = n/2$. Thus, in the mechanistic models, a maximum of $n/2$ species could coexist, whereas the uncorrelated and covariance model could potentially predict coexistence among a larger number of species. For each scenario, we identified stable equilibria and compared model predictions for species richness, species abundance, and for the elements of $\boldsymbol{a}$. We ran 20,000 iterations for each model to identify the distribution of potential outcomes. Detailed methods for these three models are available in Appendix C in the supplement.

For predictions of community richness, we find that the covariance model matches the mechanistic model much more closely than does the uncorrelated model, both in terms of increased precision (i.e. less variance around the mean prediction) and decreased bias (i.e. predictions are centered around the true values) (Fig. 3). Correlation between estimates of $\boldsymbol{a}$ from the three models remains relatively high across all models and scenarios. For large community sizes and large process noise (e.g. $n = 10$ and $\sigma_c = 0.1$), the uncorrelated model also tends to over-predict coexistence, while the covariance model does not. On average, the covariance model also provides better predictions of abundance, but the difference is relatively small.

2.5. Applying theoretical results

The results from Fig. 3 demonstrate that we can more accurately predict outcomes of the MacArthur model by incorporating the mechanistic signature of covariance imposed on the interaction matrix $\boldsymbol{a}$ than we can by measuring the components of $\boldsymbol{a}$ independently. Note that this procedure assumes that we already know the "correct" expected values of all of the terms in $\boldsymbol{a}$, and the magnitude of σ_c. However, under these circumstances we would often have enough information to reconstruct $\boldsymbol{c}$, at which point it would be more efficient to simply make predictions based on the underlying mechanistic model. To make our results more useful for real-world applications, we demonstrate a series of methods in the section below that could be used to parameterize the covariance model based on empirical observations of species communities, even when the true values of $\boldsymbol{a}$ and $\boldsymbol{c}$ are not known.

2.5.1. Detecting model dimensionality

The first challenge is to identify the number of limiting resources in the system (but see Appendix B.I for approximations that can sometimes avoid this necessity). In theory, the number of nonzero eigenvalues of $\boldsymbol{a}$ should indicate its rank. In practice, this is less straightforward because variation in the uncorrelated model will artificially increase the dimensionality of $\boldsymbol{a}$ to full rank regardless of l_{max}. However, if process noise is small, then it is possible that these added dimensions will only be weakly present in $\boldsymbol{a}$, in which case they might be identified as corresponding to particularly small eigenvalues, or as components of $\boldsymbol{a}$ that can be removed without worsening model predictions.

For example, in a simulated system with four species, but only two limiting resources, the "true" dimensionality of ***a*** should not exceed two, and the third and fourth eigenvalues should be equal to zero (n.b. these eigenvalues are calculated from the interaction matrix as a metric of dimensionality, not from the Jacobian matrix as a metric of stability). Encouragingly, we find that these eigenvalues from the uncorrelated model are small relative to the other two eigenvalues (Fig. 4a). Furthermore, if we transform this matrix to set the value of these eigenvalues to zero, which effectively removes some of the effects of the uncorrelated noise from the matrix, this leads to improved predictions of richness and abundance, but has very little effect on the actual values of the elements of ***a*** (Fig. 4b-d). Detailed methods are described in Appendix D.I in the supplement.

In a system with unknown dimensionality, similar results would be a good indication that the information associated with these eigenvalues was not mechanistically meaningful, and could be useful for detecting the true dimensionality of ***a***. Note, however, that this technique is probably not a good method for making predictions directly from the uncorrelated model. First, we find that the improvement in prediction power is smaller than that achieved by the covariance model. Second, as systems grow larger, this correction technique becomes more difficult to apply, because the transformation can introduce imaginary parts into ***a***, and because rounding errors tend to make it impossible to fully remove many small eigenvalues simultaneously.

2.5.2. Estimating covariance model parameters

Once the true dimensionality of the system is known, the second challenge is to determine whether the parameters of the covariance model can be properly estimated by fitting the model to observed data. If we parameterize the covariance model with values centered at the "true" expected values of ***a*** from the mechanistic model, we find that these correspond closely to the average predictions from the mechanistic model (Fig. 5a-g). Similarly, if we calculate the likelihood of observed outcomes, we find that likelihood increases as the estimated ***a*** terms approach their true values (Fig. 5g-h). Moreover, as outlined above, differences in the distribution of eigenvalues resulting from these parameterizations of ***a*** successfully identify systems with different numbers of limiting resources (Fig. 5i).

Jointly, these results show that by optimizing parameters in the covariance model such that they maximize the likelihood of observed data given the covariance model, parameter estimates should converge on the true values of ***a***. These analyses therefore provide a proof of concept that regression or optimization tools that are able to estimate components of a covariance matrix (e.g. generalized least squares, Bayesian hierarchical modeling) could be used to empirically parameterize the covariance model based on observed data. Full methods for this procedure are described in Appendix D.II in the supplement.

3. Discussion:

Our results demonstrate two points. First, we show that predictions of abundance, interaction coefficients, and coexistence from Lotka-Volterra models that do not account for mechanistic associations among model parameters can be inaccurate and biased, particularly for communities with many competing species. Second, we find that underlying mechanistic relationships among the interaction parameters of a Lotka-Volterra system can be successfully approximated using analytically derived covariance, which helps improve predictions and ameliorate bias. These results therefore suggest that naively parameterizing Lotka-Volterra models from field data will often lead to models with poor predictive power. Most importantly, our findings suggest that it may be relatively straightforward to parameterize a semi-mechanistic "hybrid model," which uses covariance to incorporate the rough skeleton of hypothesized coexistence mechanisms, but does not require as much detailed information as would a fully mechanistic approach.

Note that poor performance of the uncorrelated model is not contingent on uncertainty arising from process noise. Any variation that is not constrained by underlying mechanistic relationships will deform the interaction matrix ***a***, resulting in increased dimensionality (Anderson et al. 2010, Bates and Maechler 2016), and therefore poorer predictions (Dormann 2008). Thus, even if uncertainty arises from observation error rather than process noise, methods that do not constrain ***a*** to retain its mechanistic structure will likely fail when making predictions for communities of more than a few species (but see Barabás and Allesina 2015 for discussion of predictions from Lotka-Volterra models that can be robust to uncertainty).

Though models that incorporate true underlying mechanisms will likely generate predictions that are more accurate and generalizable than those from covariance approximations, there are nevertheless advantages to the hybrid approach we introduce here. Note that Eqs. (3a-f) do not require knowledge of species consumption rates, nor the identity of the resources for which species compete. Likewise, the number of limiting resources can often be estimated directly from ***a*** (Fig. 4, and Appendix D.II). Thus, even with relatively limited mechanistic information, it may be possible to derive and parameterize a covariance model such as the one we present here.

The potentially broad applicability of covariance models raises the question of whether covariances among elements in ***a*** could be measured entirely empirically, rather than deriving them from a hypothesized mechanistic model. We suspect not. There are n^4 terms describing covariance among the elements of ***a***, and measuring many of these components (e.g. Eqs. (3e-f)) would likely require replicated observations of all possible three-way combinations of species. While there are some examples of studies that realize this level of replication (Wilbur 1972, Miller 1994, Weigelt et al. 2007), these are limited to communities of relatively few species, and are generally only feasible for study organisms that are small and fast-growing. Thus, while empirically calibrated covariance might serve as a preliminary test to narrow down a large list of potential coexistence mechanisms, we think it is very unlikely that such a method could be used to improve model predictions without first specifying a mechanistic "backbone."

Interestingly, covariance could potentially explain higher-order interactions observed in some other studies (Wilbur 1972, Miller 1994, Dormann and Roxburgh 2005, Bairey et al. 2016, Mayfield and Stouffer 2017). For example, consider the terms a_{ij} and a_{ik}, describing the effect on species *i* of species *j* and *k*, respectively. Because $\mathrm{var}(a_{ij} + a_{ik}) = \mathrm{var}(a_{ij}) + \mathrm{var}(a_{ik}) + \mathrm{cov}(a_{ij}, a_{ik})$, individual observations of the joint effect of these species will either be weaker or more extreme than would be expected from two-way interactions. Unless replication of three-way interactions is sufficiently large, covariance would manifest as a change in the strength of competitive interactions depending on the presence of a third species. Thus, studies that have identified non-additive competitive interactions in the past may constitute further evidence for the important role of covariance in Lotka-Volterra models.

3.1. Potential limitations

Though simple and relatively easy to interpret, an important caveat for the implementation of MacArthur's resource competition that we utilize here is that it is only "semi-mechanistic." A major underlying assumption of this model is that resources are perfectly substitutable – i.e. any species can theoretically persist on a sufficient quantity of any single resource (Tilman 1982). Furthermore, the combination of traits that we generate for species (and that are usually utilized for this model) assumes that all species forage for multiple resources simultaneously. Lastly, the model assumes strictly linear relationships among resource requirements (i.e. the effect on individual species of adding one type of resource to the system is always a fixed fraction of the effect of adding another type of resource, regardless of the available concentration of either resource). These assumptions are relatively unrealistic, as few resources are actually perfectly substitutable, competition among substitutable resources tends to lead towards “switching” behavior (i.e. species specializing on harvesting a single type of resource), and species responses to most resources tend to saturate at higher availabilities, leading to nonlinear responses (Tilman 1982). Because of these caveats, we would expect evolution to drive species in our model to become specialists on individual resources. Thus, though helpful as a worked example, we would not necessarily expect this model to accurately predict community dynamics in most real-world systems.

A more general potential problem with the covariance approach that we use here is that in some cases, we may find that mechanistic systems cannot be well-approximated merely by incorporating covariance. Because we characterize the components of $\boldsymbol{a}$ entirely by their mean, variance and covariance, we effectively assume that these terms are drawn from a shared multivariate normal distribution. However, the terms for $\boldsymbol{a}$ in the MacArthur model are actually the result of product distributions (Grimmett and Stirzaker 2001). While an approximation based on normal distributions worked well in our tests, this will not always be the case, as product distributions include higher-order moments. Higher-order moments arise in any stochastic system where probability distributions are not perfectly described by mean, variance, and covariance, and can include features such as asymmetry for fat tails that cannot be approximated using normal distributions. Thus, for other types of mechanistic models, or even other realizations of MacArthur’s model, more complex distributions or higher-order moments may therefore be needed to accurately characterize dynamics.

Lastly, another potential problem with our approach is model identifiability. Because covariance includes relatively little information about mechanistic structure, there may be multiple models that generate similar covariance signatures. Nevertheless, this limitation could also be a useful property of our method. For example, mechanistic models of disease propagation (May and Anderson 1987) and plant metapopulation dynamics (Tilman 1994) can be developed with identical relationships among model parameters, despite obvious differences between the systems. Lotka-Volterra abstractions for these models would therefore also have identical covariance. In less obvious cases, such correspondences might be helpful for identifying seemingly disparate, but mechanistically related, classes of models.

3.2 Broader implications

Though we use one version of the MacArthur model as an example, the underlying methods we present here could be applied to other mechanistic models. Importantly, even where analytical derivation relating process noise to variation in phenomenologically observable parameters is not practical, the same results can be achieved by empirically calculating covariance from simulations of the posited underlying mechanistic model. Such a method could be particularly powerful, as it would allow inclusion of multiple sources of stochasticity (e.g. process error and observation error in multiple parameters), as well as more complex stochastic processes (e.g. non-Gaussian stochasticity). For example, a suite of individual based models could be constructed to characterize the dynamics of many different combinations of potential underlying mechanisms and types of stochasticity (Black and McKane 2012, Grimm et al. 2016). Based on simulations of these models, one could then calculate the relationships among species interaction strengths, generating a matrix of covariances among the elements of $\boldsymbol{a}$. Lastly, these relationships could be used to constrain an optimization algorithm in order to fit observed data, similar to the procedure we demonstrate in Fig. 5.

Estimating covariance directly from simulations could also help better integrate information from tradeoffs into our methods. In our model, we impose tradeoffs among species consumption vectors, but do not constrain process noise to adhere to this surface, as this would have produced covariance among the elements of $\boldsymbol{c}$, making the system analytically intractable (see Appendix C in the supplement). Because process noise is relatively small in our tests, species traits always fall closely around the tradeoff surface. This is equivalent to assuming that the underlying physiological tradeoffs that constrain species traits in our system are higher dimensional than is the ecological trait space that determines coexistence in our model (Roff and Fairbairn 2007). Nevertheless, more restrictive adherence of process noise to the tradeoff surface could be incorporated into computational estimates of covariance, which might generate more stable estimates of species abundances, and improved predictive ability.

3.3. Some notes on observation error

As previously explained, we primarily discuss the effects of process noise rather than observation error because observation error can often be mitigated. Nevertheless, some kinds of variables are by nature *pathological* – that is, they have no mean, and therefore no variance, regardless of sample size. Ratios of normally distributed variables, which commonly arise in Lotka-Volterra systems, are an important example of such a variable (Marsaglia 2006). For example, competition coefficients are typically estimated as $\alpha_{ij}=(K_i-X_i)/X_j$, growth rates are often calculated as $r_i=\ln(X_i(time=\tau)/X_i(time=0))(1/\tau)$, and even the Jacobian matrix that we use to determine stability includes ratios of potentially random variables (Eqs. (SA12a-b)).

There are some strategies for reducing the influence of these ratios. If the mean of the numerator and denominator are of a suitable magnitude relative to their variance, then their ratio may be roughly normally distributed, though the precise conditions for this are not trivial (Marsaglia 2006). Alternatively, by fitting models to dynamic data, it may be possible to estimate some of these parameters directly, rather than as a ratio of empirically measured variables (Carrara et al. 2015, Palamara et al. 2016). Nevertheless, some variables are necessarily the outcome of ratios. For example, coexistence in the MacArthur model depends on the relative consumption rates and carrying capacities of species, not their absolute magnitude (Chesson 1990, Haygood 2002). It therefore seems unavoidable that Lotka-Volterra-like methods (and likely a great many other models of species interactions) will be especially susceptible to the effects of observation error. We therefore advocate cautious testing for the effects of observation error before attributing uncertainty to process noise.

3.4. Conclusion:

Our findings suggest that a simple hybrid approach that tracks the propagation of uncertainty through ecological systems might be useful for identifying important coexistence mechanisms and predicting species abundances in poorly understood competitive communities. It remains to be seen whether the simple approach that we use based on covariance will be tractable and effective for other types of mechanisms and in real-world ecological systems. However, we hope that the methods and concepts that we introduce here both provide a warning of ways that Lotka-Volterra models can be mis-calibrated, and will help expand the utility of Lotka-Volterra approaches in diverse systems, and help guide how ecologists use these approaches in the future.

Acknowledgements: We thank E. Borer, G. Fury, F. Isbell, C. Lehman, D. Tilman, and D. Williams, and are grateful to our reviewer G. Barabás and an anonymous reviewer for helpful comments on earlier drafts of this manuscript. A.T.C. was supported by a NSF Graduate Research Fellowship, base award number 00006595, by a University of Minnesota Graduate Excellence Fellowship, and by an sDiv "catalyst" postdoctoral fellowship. Initial work by A.T.C. was conducted at the Sedgwick Reserve, hosted through the University of California Santa Barbara. Computing time was provided by the University of Minnesota Supercomputing Institute, iDiv, and UFZ.

References:

Anderson, G. W., A. Guionnet, and O. Zeitouni. 2010. Real Wigner matrices: traces, moments and combinatorics. Pages 6–34 An introduction to random matrices. Cambridge University Press, New York.

Bairey, E., E. D. Kelsic, and R. Kishony. 2016. High-order species interactions shape ecosystem diversity. Nature Communications 7:12285.

Barabás, G., and S. Allesina. 2015. Predicting global community properties from uncertain estimates of interaction strengths. Journal of The Royal Society Interface 12:20150218.

Bates, D., and M. Maechler. 2016. Matrix: Sparse and Dense Matrix Classes and Methods.

Black, A. J., and A. J. McKane. 2012. Stochastic formulation of ecological models and their applications. Trends in Ecology & Evolution 27:337–345.

Bolker, B. M., and S. W. Pacala. 1999. Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. American Naturalist 153:575–602.

Carrara, F., A. Giometto, M. Seymour, A. Rinaldo, and F. Altermatt. 2015. Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity. Methods in Ecology and Evolution 6:895–906.

Chesson, P. 1990. MacArthur's consumer-resource model. Theoretical Population Biology 37:26–38.

Chesson, P. 2000. Mechanisms of Maintenance of Species Diversity. Annual Review of Ecology and Systematics 31:343–366.

Clark, A. T., C. Lehman, and D. Tilman. 2018. Identifying mechanisms that structure ecological communities by snapping model parameters to empirically observed tradeoffs. Ecology Letters 21:494–505.

Detto, M., and H. C. Muller-Landau. 2013. Fitting Ecological Process Models to Spatial Patterns Using Scalewise Variances and Moment Equations. The American Naturalist 181.

Dormann, C. F. 2008. On community matrix theory in experimental plant ecology. Web Ecology 8:108–115.

Dormann, C. F., and S. H. Roxburgh. 2005. Experimental evidence rejects pairwise modelling approach to coexistence in plant communities. Proceedings of the Royal Society B: Biological Sciences 272:1279–1285.

Gause, G. F. 1934. Experimental analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence. Science 79:16–17.

Grilli, J., G. Barabás, M. J. Michalska-Smith, and S. Allesina. 2017. Higher-order interactions stabilize dynamics in competitive network models. Nature.

Grimm, V., D. Ayllón, and S. F. Railsback. 2016. Next-Generation Individual-Based Models Integrate Biodiversity and Ecosystems: Yes We Can, and Yes We Must. Ecosystems 20:229–236.

Grimmett, G., and D. Stirzaker. 2001. Probability and random processes. 3rd ed. Oxford University Press, Oxford ; New York.

Haygood, R. 2002. Coexistence in MacArthur-Style Consumer–Resource Models. Theoretical Population Biology 61:215–223.

Lehman, C. L., and D. Tilman. 2000. Biodiversity, Stability, and Productivity in Competitive Communities. The American Naturalist 156:534–552.

Levin, S. A. 1970. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. The American Naturalist 104:413–423.

Levine, J. M., J. Bascompte, P. B. Adler, and S. Allesina. 2017. Beyond pairwise mechanisms of species coexistence in complex communities. Nature 546:56–64.

Litchman, E., and C. A. Klausmeier. 2008. Trait-Based Community Ecology of Phytoplankton. Annual Review of Ecology, Evolution, and Systematics 39:615–639.

Loreau, M. 2004. Does functional redundancy exist? Oikos 104:606–611.

Lotka, A. 1932. Growth of mixed populations. Journal of the Washington Academy of Science 22:461–469.

MacArthur, R. 1970. Species packing and competitive equilibrium for many species. Theoretical Population Biology 1:1–11.

MacArthur, R. H. 1958. Population Ecology of Some Warblers of Northeastern Coniferous Forests. Ecology 39:599–619.

MacArthur, R. H., and R. Levins. 1967. Limiting Similarity Convergence and Divergence of Coexisting Species. Am. Nat. 101:377–385.

Marsaglia, G. 2006. Ratios of Normal Variables. Journal of Statistical Software 16.

May, R. M. 1973. Stability and complexity in model ecosystems. 1st Princeton landmarks in biology ed. Princeton University Press, Princeton.

May, R. M., and R. M. Anderson. 1987. Transmission dynamics of HIV infection. Nature 326:137–142.

Mayfield, M. M., and D. B. Stouffer. 2017. Higher-order interactions capture unexplained complexity in diverse communities. Nature Ecology & Evolution 1:0062.

Meszéna, G., M. Gyllenberg, L. Pásztor, and J. A. J. Metz. 2006. Competitive exclusion and limiting similarity: A unified theory. Theoretical Population Biology 69:68–87.

Michalet, R., S. Chen, L. An, X. Wang, Y. Wang, P. Guo, C. Ding, and S. Xiao. 2015. Communities: are they groups of hidden interactions? Journal of Vegetation Science 26:207–218.

Miller, T. E. 1994. Direct and Indirect Species Interactions in an Early Old-Field Plant Community. The American Naturalist 143:1007–1025.

Palamara, G. M., F. Carrara, M. J. Smith, and O. L. Petchey. 2016. The effects of demographic stochasticity and parameter uncertainty on predicting the establishment of introduced species. Ecology and Evolution 6:8440–8451.

Park, T. 1936. The effect of differentially conditioned flour upon the fecundity and fertility of Tribolium confusum duval. Journal of Experimental Zoology 73:393–404.

Reich, P. B. 2014. The world-wide 'fast-slow' plant economics spectrum: a traits manifesto. Journal of Ecology 102:275–301.

Roff, D. A., and D. J. Fairbairn. 2007. The evolution of trade-offs: where are we? Journal of Evolutionary Biology 20:433–447.

Roxburgh, S. H., and J. B. Wilson. 2000. Stability and coexistence in a lawn community: mathematical prediction of stability using a community matrix with parameters derived from competition experiments. Oikos 88:395–408.

Schaffer, W. M. 1981. Ecological Abstraction: The Consequences of Reduced Dimensionality in Ecological Models. Ecological Monographs 51:383–401.

Schoener, T. W. 1974. Some Methods for Calculating Competition Coefficients from Resource-Utilization Spectra. The American Naturalist 108:332–340.

Tilman, D. 1982. A Comparison with Classical Theory. Pages 190–204 Resource Competition and Community Structure. Princeton University Press, Princeton, NJ.

Tilman, D. 1990. Constraints and Tradeoffs - toward a Predictive Theory of Competition and Succession. Oikos 58:3–15.

Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. Ecology 75:2–16.

Tilman, D. 2011. Diversification, Biotic Interchange, and the Universal Trade-Off Hypothesis. American Naturalist 178:355–371.

Vandermeer, J. H. 1969. The Competitive Structure of Communities: An Experimental Approach with Protozoa. Ecology 50:362–371.

Wangersky, P. J. 1978. Lotka-Volterra Population Models. Annual Review of Ecology and Systematics 9:189–218.

Weigelt, A., J. Schumacher, T. Walther, M. Bartelheimer, T. Steinlein, and W. Beyschlag. 2007. Identifying mechanisms of competition in multi-species communities. Journal of Ecology 95:53–64.

Wilbur, H. M. 1972. Competition, Predation, and the Structure of the Ambystoma-Rana Sylvatica Community. Ecology 53:3–21.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-Bares, T. Chapin, J. H. C. Cornelissen, M. Diemer, J. Flexas, E. Garnier, P. K. Groom, J. Gulias, K. Hikosaka, B. B. Lamont, T. Lee, W. Lee, C. Lusk, J. J. Midgley, M.-L. Navas, Ü. Niinemets, J. Oleksyn, N. Osada, H. Poorter, P. Poot, L. Prior, V. I. Pyankov, C. Roumet, S. C. Thomas, M. G. Tjoelker, E. J. Veneklaas, and R. Villar. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

Figures:

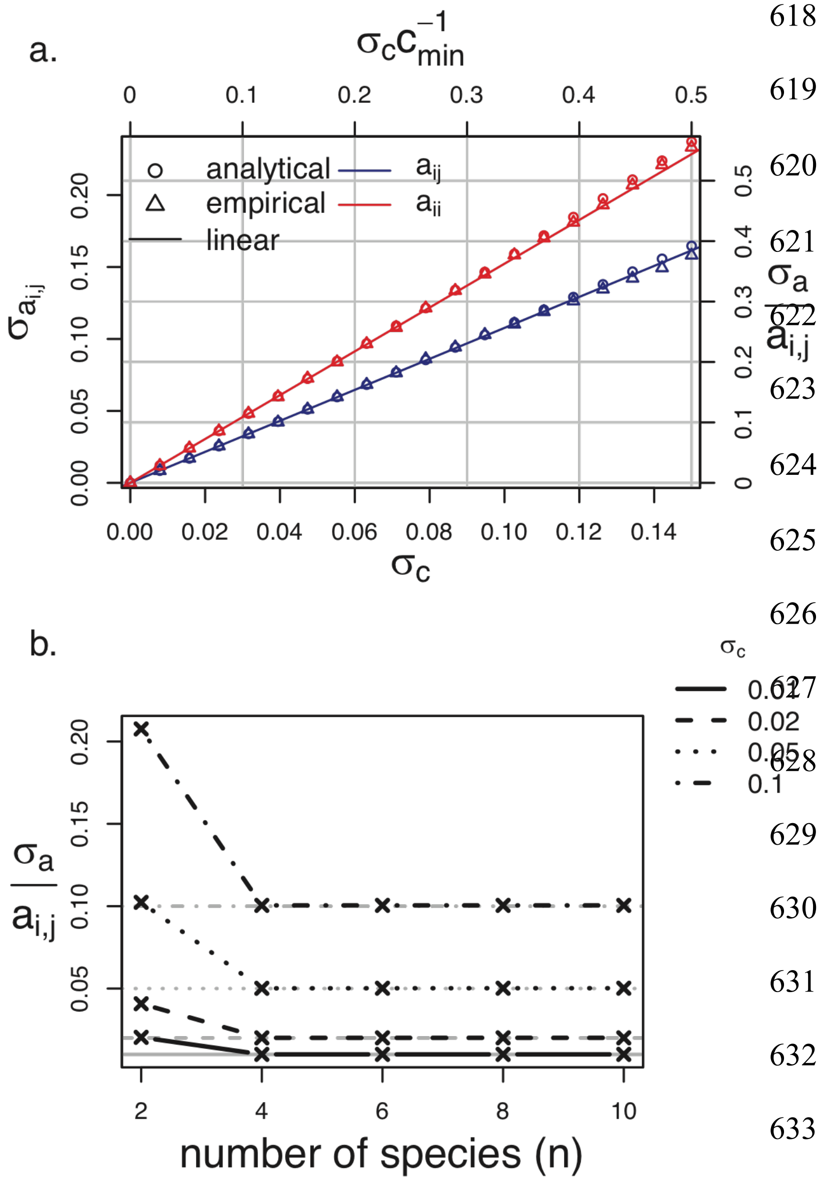

Figure 1: Propagation of process noise in model parameters. **(a)** σ_{aij} shows standard deviation of the competition coefficients a_{ij} (blue) and a_{ii} (red) as a function of σ_c, the standard deviation of noise added to the mechanistic consumption parameters, ***c***, in MacArthur's resource model. Circles show analytical expectations for the relationship, triangles show average results from 20,000 simulations of the mechanistic model, and lines show linear approximations, as described in the Appendix B.I in the supplement. Additional axis labels $\sigma_c(c_{min})^{-1}$ and $\sigma_a(a_{ij})^{-1}$ show the magnitude of these variabilities relative to the minimum ***c*** value and mean a_{ij} value, respectively. **(b)** Mean observed value of σ_{aij} as a function of community size, *n* ($l_{max} = n/2$ in all cases). With the exception of n=2, σ_{aij} scales with a_{ij} by a factor of σ_c (corresponding values of σ_c shown in grey). See online version of article for color rendering of figures.

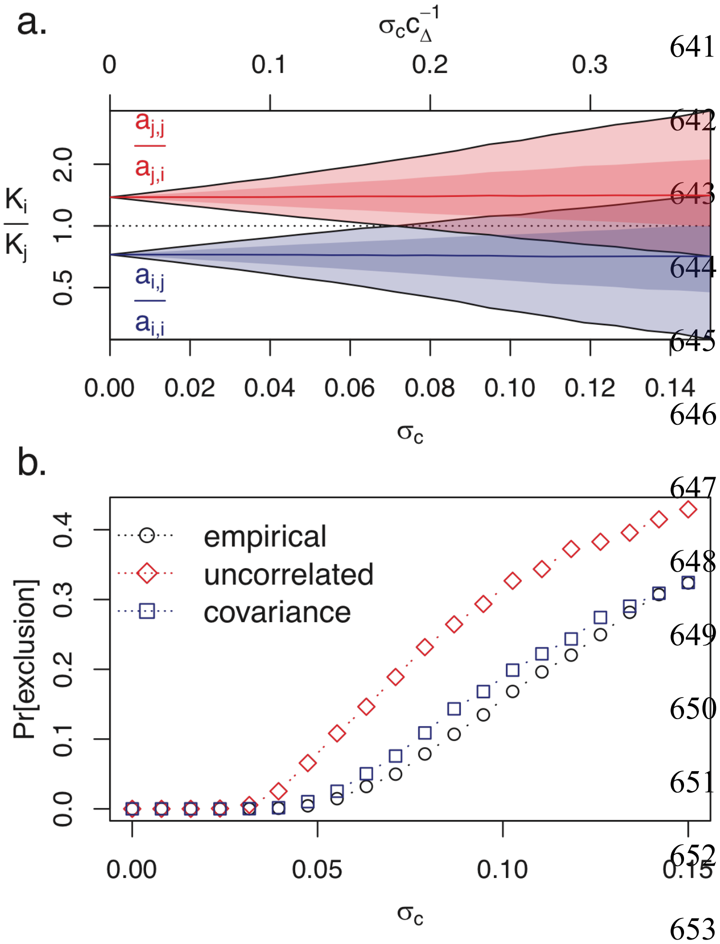

Figure 2: Effects of process noise on coexistence for a system with two species and two limiting resources. **(a)** Species *i* and *j* coexist when the ratio of their carrying capacities K_i/K_j falls between a_{ij}/a_{ii} and a_{jj}/a_{ji} (i.e. between the dark red and dark blue lines – see Fig. 1 in Chesson (1990) for a similar approach). Lines, dark shaded intervals, and light shaded intervals show the mean, standard deviation, and 95% confidence interval, respectively, for these ratios as a function of process noise, σ_c, while dotted line shows fixed value for K_i/K_j. Top axis shows magnitude of noise relative to the mean difference between species consumption rates for the two limiting resources. **(b)** Effect of process noise. Pr[exclusion] shows the probability that one of two competing species will drive its competitor extinct. Empirical results show the average of 20,000 simulations of the mechanistic model, while uncorrelated and covariance predictions are based on the analytical expectation of variance and covariance in $\boldsymbol{a}$, respectively, as described in the main text. Distance from the empirical estimates (black points) demonstrates prediction error. See online version of article for color rendering of figures.

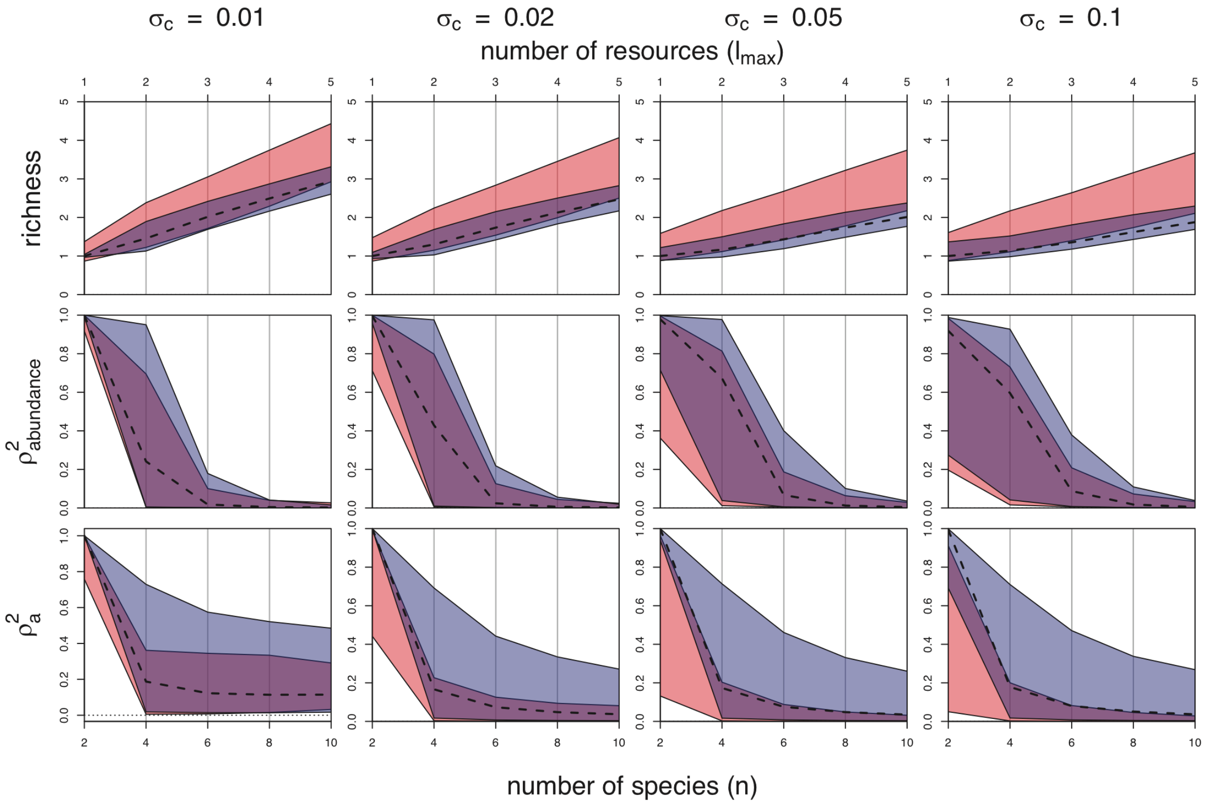

Figure 3: Correspondence between mechanistic model and approximations from the Lotka-Volterra competitive equations as a function of process noise, σ_c, for communities of 2 to 10 species (note that $l_{max} = n/2$ in all cases). Intervals show mean ± one standard deviation based on 20,000 simulations. Red intervals show results for uncorrelated model, blue shows covariance model, and purple regions show overlap between the two. Black dashed line shows mean result from the mechanistic model. $\rho^2_{abundance}$ and ρ^2_a show the square of Pearson's correlation coefficient comparing predictions of species abundances or predictions of the interaction matrix $\boldsymbol{a}$, respectively. See online version of article for color rendering of figures.

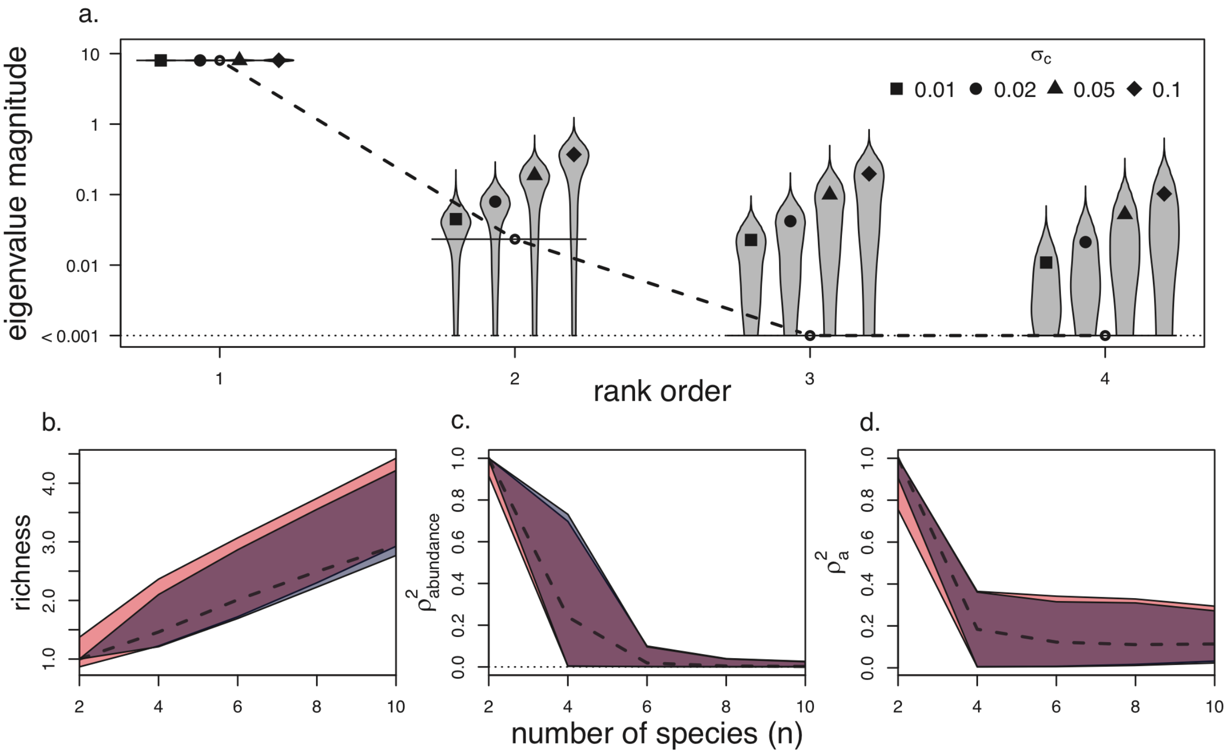

Figure 4: Influences of uncorrelated error on predictions of community stability. **(a)** Real components of the eigenvalues for the interaction matrix ***a*** in a system with four species and two limiting resources. Black horizontal lines and dashed segments show mean values for the mechanistic model (n.b. fourth and fifth eigenvalues are always zero). Width of shaded intervals shows frequency distribution of results for Lotka-Volterra systems with uncorrelated error (i.e. the uncorrelated model). **(b-d)** Correspondence between mechanistic model and Lotka-Volterra approximations for 2 to 10 species with $\sigma_c = 0.01$ and $l_{max} = n/2$. Red intervals show results for uncorrelated model as described in the legend to Fig. 3, blue intervals show results for model with eigenvalues $n>l_{max}$ coerced to zero, as described in Appendix D.I in the supplement, and purple regions show overlap between the two. See online version of article for color rendering of figures.

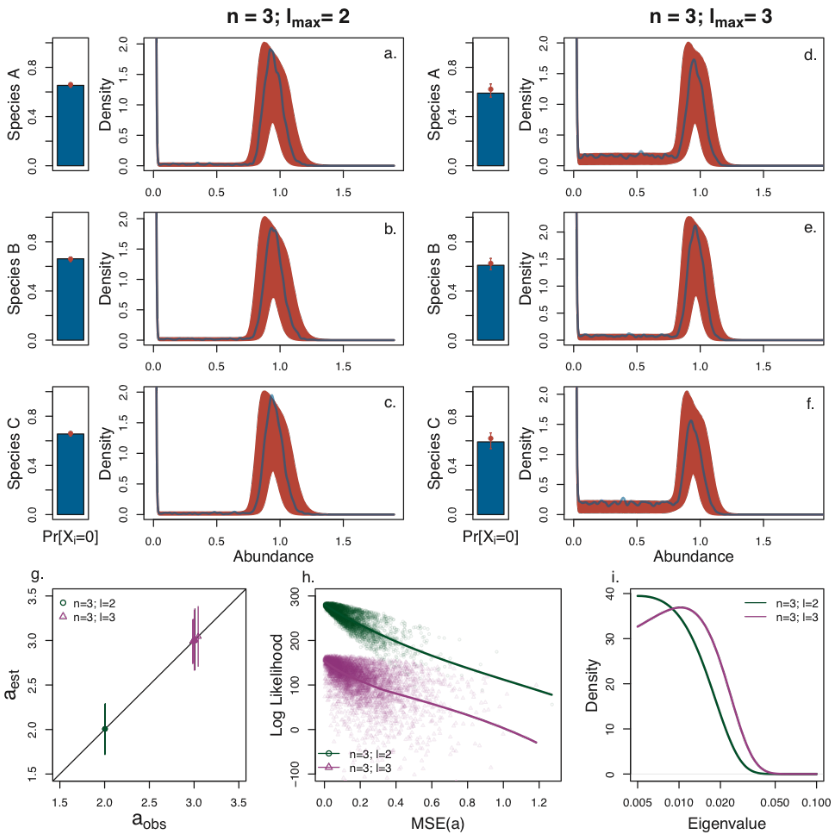

Figure 5: Fitting the covariance model to empirical data. Figures show variability in predictions from the covariance model based on observations from simulations from the mechanistic model of 100 "plots". **(a-f)** Frequency distributions of species abundance for a system with *n=3* and $l_{max}=2$ **(a-c)**, and for a system with *n=3* and $l_{max}=3$ **(d-f)**. In all simulations, $\sigma_c = 0.05$. Bar plots show frequency of zero abundances for each species, and density plots show distribution of positive abundances. Blue shows expected results for the mechanistic model, while red shows mean ± one standard deviation based on results from 5,000 iterations of the covariance model. **(g)** Comparison of observed and estimated elements of ***a***. Vertical lines show 95% confidence intervals for parameter estimates from the covariance model. (**h)** Model likelihood as a function of the mean square error (MSE) comparing estimated and observed parameters for the interaction matrix, ***a***. **(i)** Distribution of the third (smallest) eigenvalue of ***a*** for the two systems. See online version of article for color rendering of figures.