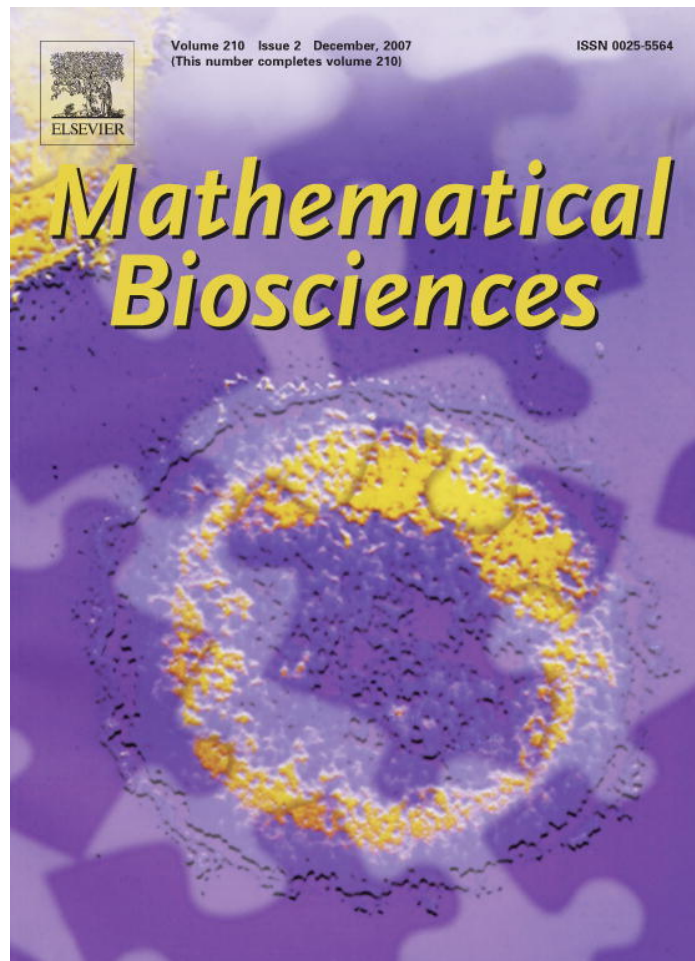


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com ScienceDirect

Mathematical Biosciences 210 (2007) 490–507

**Mathematical
Biosciences**

www.elsevier.com/locate/mbs

Heterogeneity in dynamic species abundance models: The selective effect of extinction processes

Steinar Engen *

Department of Mathematical Sciences, Norwegian University for Science and Technology, N-7491 Trondheim, Norway

Received 14 February 2007; received in revised form 27 May 2007; accepted 4 June 2007

Available online 10 June 2007

Abstract

There is often large variation in traits across the species of a community. In particular, variation in life history traits affecting population dynamics is likely to affect the species abundance distribution. Applying a dynamic and heterogeneous species abundance model we study how differences in extinction time for species in a community act as a force changing the distribution of dynamic parameters across species. This process may generate communities that are more heterogeneous than the heterogeneity measured as the species enter the community. Analytical results for some versions of the lognormal and gamma species abundance model are given as exemplifications of this process, together with stochastic simulations demonstrating the temporal changes in number of species and community heterogeneity through time.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Species abundance distribution; Lognormal distribution; Dynamic lognormal model; Gamma distribution; Gamma distribution of the second kind; Heterogeneity; Neutral model; Extinction; Extinction process; Colonization; Invasion; Speciation

1. Introduction

The concept of stochastic abundance models was first introduced by Fisher et al. [1] who assumed the abundance of the species in a community follows a gamma distribution. Assuming

* Fax: +47 73 591038.

E-mail address: steinaen@math.ntnu.no

Poisson sampling with mean number of individuals sampled being proportional to the abundances, they proposed Fisher's well known log-series distribution describing the sequence n_1, n_2, \dots of the number of species with 1, 2, \dots , individuals in the sample, by allowing the shape parameter of the gamma distribution approaching zero. Fisher used the gamma distribution purely as a tool for describing the variation in abundance in a community, considering these abundances as fixed in the calculation of the uncertainty of his diversity parameter α estimated from a sample [2]. Later, there has been a large number of papers discussing possible mechanisms for generating this distribution, Preston's lognormal distribution of abundances [3,4], and a number of other minor deviations from these models. Some of these explanatory models are based on niche theory, focusing on how the species in a community divide the available resources between them, reviewed by May [5], Pielou [6], Engen [2] and Magurran [7], while others focus on temporal variations of abundances including speciation and extinction [8–15].

A possible explanation of the fact that species abundance varies a lot within a community, is the variation of traits among species. In particular, life history parameters that determine the population dynamics are traits affecting species abundances. Distribution of such parameters across species are widely discussed in the literature [16]. It has also been shown that population dynamical parameters, estimated from time series data as well as demographic data from single individuals, vary considerably across species [17]. However, there has been little emphasis on relating such variation to dynamic models generating species abundance distributions. The main focus of the present paper is to establish this connection. Extending previous work on heterogeneous species abundance models by Engen and Lande [18], we analyze species abundance distributions generated by dynamic models with emphasis on how extinction processes affect the distribution of parameters across species.

This approach contrasts the leading theory in community dynamics based on neutral models [8,9] in which all species are interchangeable and their abundance fluctuations in time are driven by demographic stochasticity, that is, by the between individuals differences in survival and reproduction. Such models have explained some classical patterns of communities such as the shape of species abundance distributions and species-area relations [9–11,19]. The empirical basis for the neutral theory has mainly been studies of tropical trees with extreme competition for space in a region with small temporal environmental fluctuations. Hence, random substitution of dead individuals by others of the same or different species, giving mathematically a demographic type of noise in the process, may be a realistic description. On the other hand, studies of animal populations of closely related species in northern areas with fluctuating environments have shown that species are rarely interchangeable, as the parameters describing the dynamics tend to differ among species [20], a difference that here will be modeled as a distribution of parameters across species.

Statistical analysis of the stochasticity in the dynamics has shown that environmental stochasticity is usually the major stochastic mechanism unless populations are very small, say at the order of 50 individuals, in which case environmental stochasticity acts together with demographic stochasticity [12]. These findings also contradict the idea of neutrality, which is based solely on demographic stochasticity. Environmental stochasticity is generated by stochastic effects, physical or biological, affecting all individuals in the population in the same way [12,21,22]. In a population of size N these effects generate variances in the yearly change in population size that are of order N^2 , while demographic variance terms are proportional to N and can be neglected for large populations. Theoretical analysis has shown that probability of extinction or quasi-extinction may

be extremely wrong, not even of the right order of magnitude, if the environmental variance is neglected [12]. The demographic variance also affects the extinction process, but the effect of modeling the demographic stochasticity by environmental terms is not as dramatic as doing the opposite choice. Hence, management of animal communities in fluctuating environments, most likely to occur outside the tropical climate zone, is preferably based on theories in which the basic stochastic effects are environmental rather than demographic, although the best approach would be to include both.

Applying diffusion theory Engen and Lande [18] defined a class of dynamic abundance models generating different types of species abundance distributions. They also included the new concept of heterogeneous species abundance models by allowing the dynamics of species to be defined by different parameters generated by some distribution at invasion (or speciation). Here we will analyze in some detail how extinction processes affect the heterogeneity of the community. The differences between species in expected time to extinction change the distribution of population parameters across species by giving larger weight to parameter sets promoting long time to extinction. Surprisingly, this process, in which the environments select some species to go extinct quickly and others to stay in the community for a long period of time, does not necessarily lead to smaller heterogeneity in the community as one is tempted to believe from similar kinds of problems in population genetics. Nature's 'sampling of species' will generally generate large changes in parameter distributions so that the distribution in a stable community is quite different from the distribution at invasion. However, the variance or coefficient of variation of parameters across species will not necessarily be reduced. In general, the effects of variability in time to extinction on the heterogeneity depend on the type of heterogeneity at invasion. Parameter variability will change, but the heterogeneity will not be wiped out and may even increase. This supports the claim that animal communities in fluctuating environments should be described by heterogeneous species abundance models.

Variation of dynamical parameters across species may be viewed as an outcome of niche theory as for example outlined 50 years ago by MacArthur [23], although there were no stochastic dynamics in that model. For more recent niche models reflecting heterogeneity among species see [24,25]. A given distribution of parameters, defining in particular a distribution of carrying capacities, may partly reflect or be a result of species organization. The distribution of parameters does change through time, but at a much slower time scale than species abundances, since changes only occur when there is an invasion or extinction. Nevertheless, the dynamics describing the stochastic fluctuations of each species will have a large effect on the realized distribution of parameters across species.

In summary, the main goal of this paper is to study the relation between the distribution of dynamical parameters among invading species and that of resident species. This will through some light on how fluctuating environments affect community heterogeneity.

2. Some general properties of dynamic and heterogeneous species abundance models

Engen and Lande [9] assumed that species entered the community at points of time following a Poisson process with constant rate β so that the average number of species entering during a time interval of length t is βt . This rate may be thought of as the speciation rate if large areas are con-

sidered, or it may represent the colonization rate of sub-populations or islands. The dynamics of a species that enters the community is approximated by a diffusion process with infinitesimal mean and variance $\mu(x; \theta)$ and $v(x; \theta)$, where x is the population size and θ some vector of parameters describing the dynamics. This may be considered as an approximation to the discrete model where the mean and variance of the change in population size during a time step $\Delta t = 1$ conditioned on x is $\mu(x, \theta)$ and $v(x; \theta)$, respectively. The purpose of applying the diffusion approximation is that well known results for extinction in diffusion processes [26] can be utilized to study the balance between colonization and extinction.

The processes for each species considered separately may in general be dependent. Strong competition between species for food or space may create negative correlations between species abundances. For large communities however, that is, communities with many species, Engen and Lande [13] showed that such correlations only had a small effect on the distribution of relative abundance. Actually, for a gamma type of model with parameters corresponding to Fisher's log-series model [1] it was shown that Fisher's diversity parameter α , which is often as large as 30–40, in the case of independent species was only reduced by 2 when forcing the total size of the community to be constant. For the lognormal model with heterogeneity in growth rates, neither a common density regulation defined by a function of all abundances nor a common noise component, will affect the dynamics of relative abundances [12]. Hence, ignoring such terms and assuming independence, is not likely to have much effect except on the fluctuations of the total community size.

The dependence may also be of the opposite type. There may be a common noise term that creates positive correlations in the total noise of two species and fluctuations in the size of the community. Engen and Lande [13] also discussed this dependence. For their model generating the lognormal species abundance distribution it was shown that a common noise term had no effect on the fluctuations of relative abundances, but could have some small effect on the extinction processes.

As a compromise between the two types of dependence described above we here apply the simplest type of model, assuming independent dynamics for the species in the community. Then the total community do actually fluctuate in size, the fluctuations being smaller than what should be expected when there are large common noise terms, but larger than expected if there is strong inter-specific competition.

Engen and Lande [18] used the diffusion approximation assuming first that the process for each species is a diffusion with infinitesimal mean and variance $\mu(x; \theta)$ and $v(x; \theta)$. Assuming that species enter the community at abundance 2 and go extinct at abundance 1, Engen and Lande [18] then showed that species abundances when the process has reached stationarity (balance between colonizations and extinctions) follows an inhomogeneous Poisson process with rate

$$\lambda(x; \theta) = \frac{2\beta}{v(x; \theta)} e^{2 \int_1^x \mu(u; \theta)/v(u; \theta) du}, \quad (1)$$

which means that the number of species with abundances in the interval $[a, b]$ at a given time is Poisson distributed with mean $\int_a^b \lambda(u; \theta) du$. In particular, the number S of species in the community is Poisson distributed with mean $\int_1^\infty \lambda(x; \theta) dx$. Conditioning on S the species abundance distribution is then $f(x; \theta) = \lambda(x; \theta) / \int_1^\infty \lambda(u; \theta) du$.

Heterogeneity is modeled by assuming that species entering the community have parameters θ generated by some distribution $\pi(\theta)$, which simply changes the model to an inhomogeneous Poisson process with rate

$$\lambda(x) = \int \lambda(x; \theta)\pi(\theta)d\theta, \tag{2}$$

where the integration is performed over all possible values of θ [18].

When dealing with species abundance models it is common to work with log abundances rather than abundances. The results given here and below are also valid if we interpret x as log abundances provided that $\mu(x, \theta)$ and $v(x; \theta)$ then are the infinitesimal mean and variance for the log population size and the lower integration limit representing the extinction barrier is chosen at zero rather than 1.

3. Relations to more common presentations of abundance models

This inhomogeneous Poisson process can now alternatively be expressed in more familiar terms used in the theory of species abundance models. First, conditioning on the number of species in the community the species abundance distribution, that is, the distribution of abundances among the species of the community, is

$$f(x) = \lambda(x) / \int_1^\infty \lambda(u)du. \tag{3}$$

Models are, however, most frequently expressed the way this was originally done by Fisher et al. [1], by plotting the observed number of species n_j with j individuals in the sample against j . Fisher assumed that the number of individuals in the sample of a species with abundance x is Poisson distributed with mean vx , where v is the sampling intensity. In our terminology we then find that

$$E(n_j) = \int_1^\infty \frac{(vx)^j}{j!} e^{-vx} \lambda(x) dx$$

which is directly comparable to an observed sequence (n_1, n_2, \dots) . Writing $EN = E(\sum j n_j)$ for the expected number of individuals in the sample, we find the relation between v and EN by summing over the Poisson distribution, giving

$$EN = v \int_1^\infty x \lambda(x) dx.$$

4. The preference function of dynamic parameters

The set of parameters θ determining the species dynamics also determines the distribution of the time the species spends in the community before it goes extinct. In this sense, there will be favorable parameter combinations such as large growth rates, small environmental variances and weak

density regulation, that promote long species lifetime in the community. Extinctions therefore tend to act as a selection force on the species level, selecting parameter combinations that are favorable. Hence, in the stationary community, the observed distribution of θ across species, say $\pi^*(\theta)$, may be very different from the distribution $\pi(\theta)$ at colonization.

Considering only species with parameters in the interval $[\theta, \theta+d\theta]$, their colonization rate is $\beta\pi(\theta)d\theta$. Hence, the species abundance model for these species is the inhomogeneous Poisson process defined by Eq. (1) with β replaced by $\beta\pi(\theta)d\theta$, so that the total number of species in the community with parameters in $[\theta, \theta+d\theta]$ is Poisson distributed with mean $h(\theta)d\theta = \int_1^\infty \lambda(x; \theta)\pi(\theta)d\theta dx$. Consequently, again conditioning upon S , the distribution of parameters across the species present in the community is $\pi^*(\theta) = h(\theta)/\int h(u)du$ where the integral runs over all possible parameter values of colonizing species.

When the process has reached stationarity the rate of extinctions must be equal to the rate of new species arriving, which is β . On the other hand, not conditioning on θ , the mean extinction rate is also ES/ET , where T is the time to extinction. One can alternatively see this from the fact that $\lambda(x; \theta)$ is the product of β and the Green function for the process [18] and using the result that the integral of the Green function is the expected time to extinction [12,26]. From this observation we see that $ES = \iint \lambda(x; \theta)\pi(\theta)dx d\theta$ and $ET = \int w(\theta)\pi(\theta)d\theta$, where

$$w(\theta) = \int_1^\infty \beta^{-1} \lambda(x; \theta)dx \tag{4}$$

expressing the expected time a species stays in the community as function of the parameters defining the dynamics of the species. The function $w(\theta)$ can be viewed as a *preference function* expressing to which extent species with different parameters θ are preferred by nature's selection process.

Alternatively the relation between $\pi(\theta)$ and $\pi^*(\theta)$ can now be expressed as

$$\pi^*(\theta) = \frac{\pi(\theta)w(\theta)}{\int \pi(u)w(u)du} = \frac{\pi(\theta)w(\theta)}{\bar{w}}, \tag{5}$$

where \bar{w} is the mean preference of colonizing species. Notice that the invasion rate β does not occur in Eq. (5) and also that the preference function acts on species in the same as fitness in population genetics acts on genotypes.

Another interesting observation is the analogy between Eq. (5) and Bayes formula used in Bayesian statistics. We may consider the distribution of θ among invading species, $\pi(\theta)$, to be the prior distribution of θ subject to no other information than the species being one of the invaders. The preference function $w(\theta)$ then plays the role of the likelihood function, that is, the distribution of the parameters given the data. However, there are no data to be inserted here, but rather the observation that the species is resident, actually one of the species being present in the population at a given time. The distribution $\pi^*(\theta)$ is then the posterior distribution given that we have a resident species. So, the term preference function could have been named likelihood function, but since there is no statistical inference involved here I prefer using the term preference.

The process described by Eq. (5), showing how differences in time to extinction changes the distribution of parameters across species from $\pi(\theta)$ to $\pi^*(\theta)$ will from now on be called *parameter sampling*.

5. Simulating the parameter process

Although we will present several analytical results it may be useful for illustrations to simulate the parameter sampling process. A complete simulation of the process would involve simulations of the abundance of each species, as well as keeping track of parameter values defining their dynamics. Such simulations will be time consuming as well as difficult to express graphically. However, as we are basically most interested in studying the heterogeneity of the community and how heterogeneity may change by parameter sampling, we choose the simpler approach of simulating only the *parameter process* by recording what happens when there is a transition changing the parameter set of the community, that is, when either a new species enter the community or a species goes extinct. The time scale for this process in discrete time will then be the number such events, which is different from the running time. However, when the process has reached stationarity, the two measurements of time will be approximately proportional so the graphs will give the correct picture of temporal changes in heterogeneity.

Most species entering the community will fluctuate around their carrying capacity for a long time before they go extinct. Then, the distribution of time to extinction is approximately the exponential distribution with mean value equal to the expected time to extinction which is $w(\theta)$. The total rate of events is then $\delta = \beta + \sum w(\theta)^{-1}$, where the sum is taken over all species in the community. When an event occurs, this is the arrival of a new species with probability β/δ and the extinction of the species with parameters θ with probability $w(\theta)^{-1}/\delta$. Hence, keeping record of the values of θ for the species in the community, the next transition may easily be simulated.

6. Parameter distribution in mixtures of lognormal models

Engen and Lande [18] considered the linear model on the log scale (x denotes log abundance) with $\mu(x; s, \gamma) = s - \gamma x$ and $v(x; \sigma_e^2) = \sigma_e^2$, which is known as the Ornstein–Uhlenbeck process [26]. Here s is the expected growth rate on the log scale for small population sizes, in biological literature often called ‘the stochastic growth rate’, while the growth rate for the expected population size $r = s + \sigma_e^2/2$ is usually called ‘the deterministic growth rate’ [12]. The parameter γ describes the strength of density-regulation and σ_e^2 is the environmental variance. The vector of parameters in the general presentation is accordingly $\theta = (s, \gamma, \sigma_e^2)$. Since we are working on the log scale the extinction boundary and lower integration limit is chosen as $0 = \ln 1$ as indicated in Section 2.

The inhomogeneous Poisson process now has intensity given by Eq. (1) which turns out to be proportional to the normal distribution

$$\lambda(x; \theta) = \frac{2\beta}{\sigma_e^2} e^{s^2/(\gamma\sigma_e^2)} e^{-\frac{1(x-s/\gamma)^2}{2\sigma_e^2/(2\gamma)}} \tag{6}$$

so that the log abundance distribution is the normal distribution censored at zero. Hence this is the lognormal species abundance distribution censored at the lowest possible abundance which is one.

Combining Eqs. (6) and (4) we find the preference function

$$w(\theta) = \frac{2\sqrt{\pi}}{\sigma_e\sqrt{\gamma}} e^{s^2/(\gamma\sigma_e^2)} \Phi\left(\frac{2s}{\sigma_e\sqrt{2\gamma}}\right), \tag{7}$$

where Φ is the cumulative standard normal distribution. Notice that the last factor can be approximated by 1 when the total probability mass of the normal distribution is above zero. Actually, this factor is between 0.99 and 1 if $2s/(\sigma_e\sqrt{2\gamma})$ is greater than 2.33. Fig. 1 shows this preference function for different combinations of s and $\gamma\sigma_e^2$, while Fig. 2 shows the gradient, indicating the direction of the selection force using the same two parameters.

6.1. Heterogeneity only in the stochastic growth rate

Let us first illustrate the parameter sampling using the heterogeneous model defined by Engen and Lande [18] who assumed that the stochastic growth rate s varied between species, while the other parameters were constant. For species entering the community s is normally distributed with mean s_0 and variance τ^2 . The parameter τ^2 then expresses quantitatively the heterogeneity of invading species. Then, using Eq. (2), the abundances are described by an inhomogeneous Poisson process still defined by a Gaussian function [18]

$$\lambda(x) = \alpha\beta e^{-(x-\eta)^2/(2\rho^2)}. \tag{8}$$

The mean and variance in this Gaussian function are $\eta = \psi s_0/\gamma$ and $\rho^2 = \psi\sigma_e^2/(2\gamma)$, that is, the variability in the value of s for invading species has the effect that the mean and variance of the normal distribution are multiplied by the same factor

$$\psi = \left(1 - \frac{2\tau^2}{\gamma\sigma_e^2}\right)^{-1}. \tag{9}$$

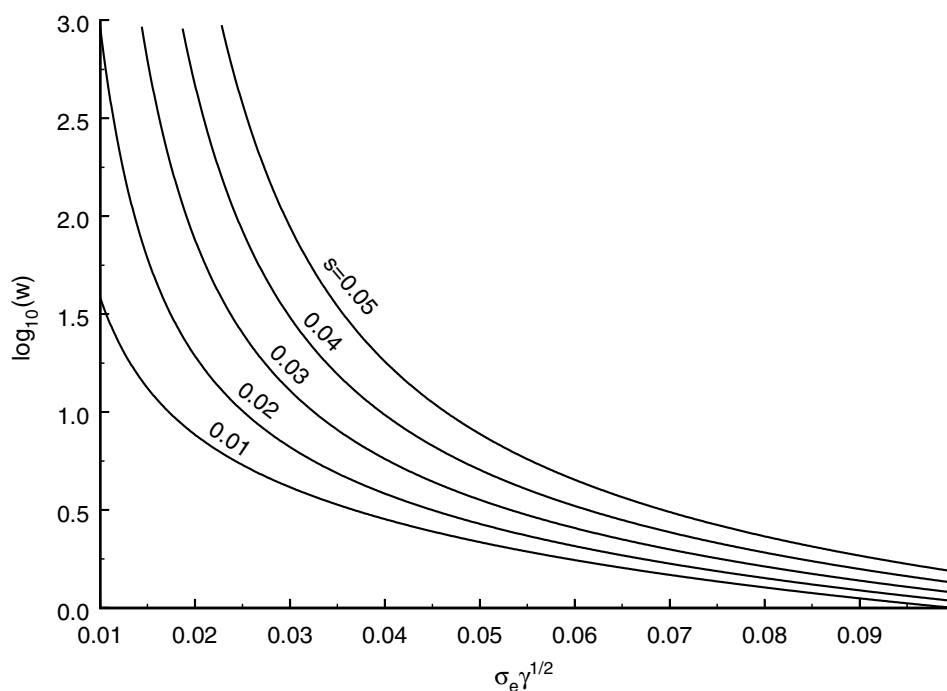


Fig. 1. The log with base 10 of preference in the lognormal abundance model as function of $\sigma_e\sqrt{\gamma}$ for different values of the stochastic growth rate $s = r - \sigma_e^2/2$.

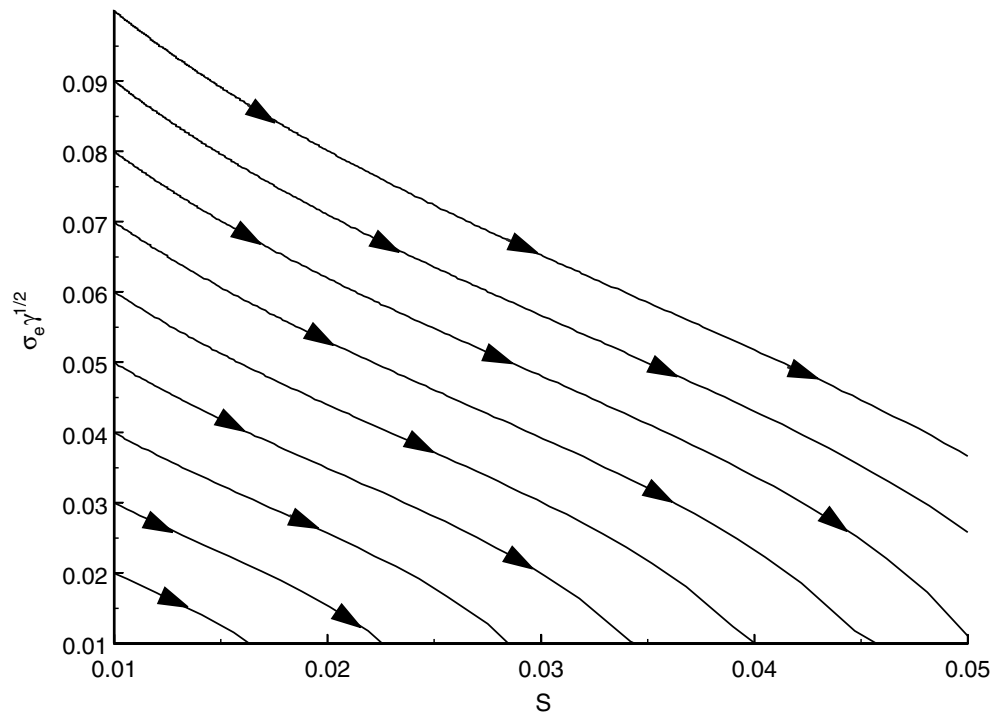


Fig. 2. The gradient of the preference function (the direction of the movement due to parameter sampling) in the lognormal model with variables s and $\sigma_e \sqrt{\gamma}$.

We must require that $\tau^2 \leq \frac{1}{2} \sigma_e^2 \gamma$, corresponding to a finite $\psi \geq 1$, in order to obtain stationarity and hence preventing the expected number of individuals and species from growing to infinity, which will occur if too many species with very large growth rates enter the community.

The expected number of species in the community is proportional to the invasion rate β as well as the constant factor $\alpha = (2/\sigma_e^2) \exp(s_0 \eta / \sigma_e^2)$ in Eq. (8).

We now assume that s is large enough for the last factor in Eq. (7) to be approximated by 1, which is the case if the carrying capacities s/γ on the log scale are sufficiently large compared to the standard deviation $\sigma_e / \sqrt{2\gamma}$. From the general equation Eq. (5) we further find the distribution of s in the community

$$\pi^*(s) = c e^{s^2 / (\gamma \sigma_e^2)} e^{-(s-s_0)^2 / (2\tau^2)}, \tag{10}$$

where c is a normalizing constant. Since this is the exponential function of a second degree polynomial in s it is a normal distribution. Rewriting it on standard form we find that the mean is s_0/ψ and the variance is $\tau^2 \psi$. Since $\psi > 1$ if $\tau^2 > 0$ we see that the selection process in this simple model always increases the between species variance in the stochastic growth rates s , that is, parameter sampling increases the heterogeneity of the community. Actually, a very small variance τ^2 in stochastic growth rates at colonization may produce quite large heterogeneity in the community. Notice that the carrying capacities $K = e^{s/\gamma}$, are lognormally distributed across species. The squared coefficient of variation in carrying capacity is accordingly $C^2 = e^{\tau^2 \psi / \gamma^2} - 1$. Hence, not only the variance of this parameter, but also its coefficient of variation, is increased by the parameter sampling defined by the extinction processes.

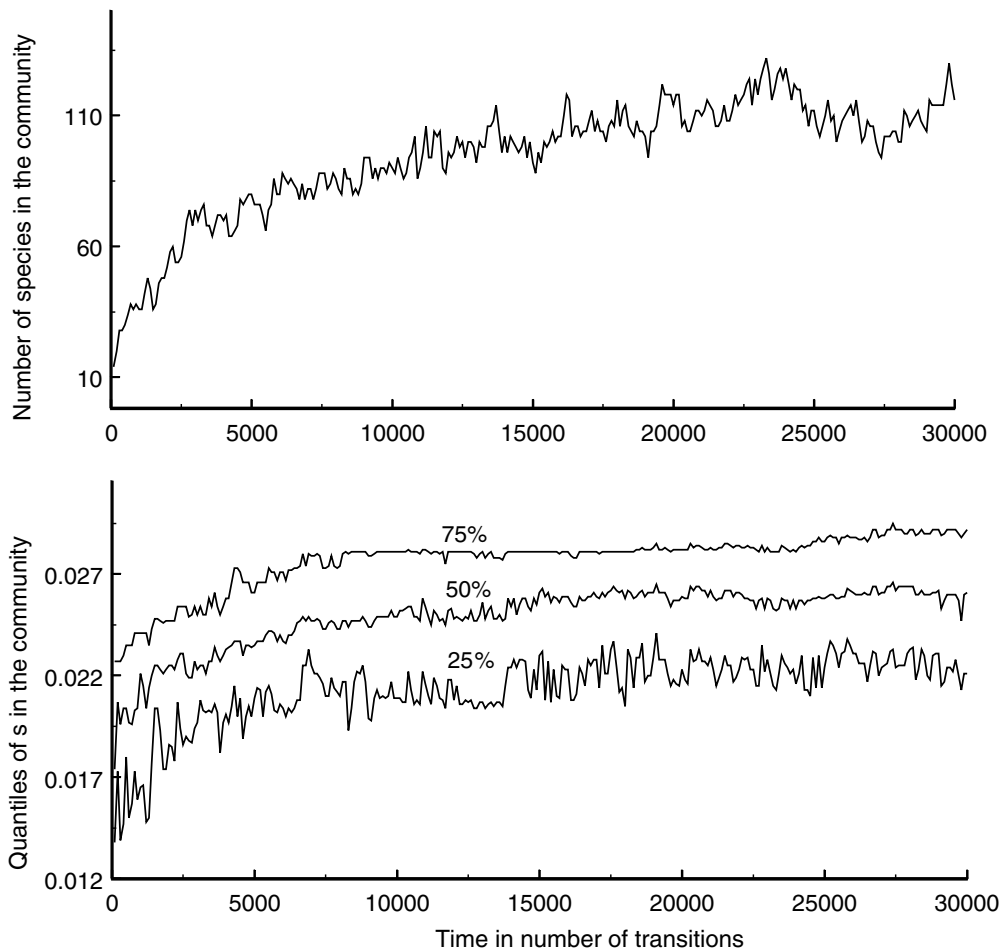


Fig. 3. Starting with an empty community the graph shows a simulation of the number of species in the community (upper panel) as well as the the median and the 25 and 75 percent quantiles of the distribution of the parameter s in the community. The model is the lognormal species abundance model. The parameter values are $\sigma_e^2 = 0.01$, $\gamma = 0.01$ while the stochastic growth rate at colonization is normal with mean $s_0 = 0.01$ and coefficient of variation $\tau/s_0 = 0.67$. The time scale is the number of transitions (see text). The parameter $\psi = 3.16$ giving that the mean of s at stationarity is 0.0316. After 100000 transition (not shown) the median of s has reached 0.030, so the convergence to stationarity is slow.

Fig. 3 shows a simulation of quantiles of s in the community as well as the number of species, starting from an empty community.

6.2. Heterogeneity in γ and σ_e^2

From equation Eq. (7) we see that the preference of the parameters γ and σ_e^2 is a function of their product. Strong density-regulation as well as large environmental stochasticity tend to reduce the preference through the product $\gamma\sigma_e^2$ as demonstrated in Fig. 2. Hence, in the same way as parameter sampling tends to increase the mean value of s , it also tends to decrease the value of $\gamma\sigma_e^2$. Writing $z = \gamma\sigma_e^2$ we see that the preference function is proportional to $z^{-1/2} \exp(s^2/z)$

provided that the last factor in Eq. (7) is close to 1. Hence, we can study this problem by assuming that z for invading species has the inverse gamma distribution

$$\pi(z) = \frac{\alpha^k}{\Gamma(k)} z^{-k-1} e^{-\alpha/z} \tag{11}$$

which has mean $\alpha/(k - 1)$ and squared coefficient of variation $C^2 = 1/(k - 2)$ (defined for $k > 2$). Using equation Eq. (7) then gives that $\pi^*(z)$ is also an inverse gamma distribution but with parameters $k' = k + 1/2$ and $\alpha' = \alpha - s^2$ so that α must be larger than s^2 in order that the solution is stationary. Hence, parameter sampling acting on $z = \gamma\sigma_c^2$ reduces the expected value of this parameter to $(\alpha - s^2)/(k - 1/2)$ and the squared coefficient of variation to $1/(k - 3/2)$. The important observation, however, is that parameter sampling does not dramatically reduce the heterogeneity.

Fig. 4 shows some examples of how the selection of species parameters act on the heterogeneity of the community in this mixture of lognormal models.

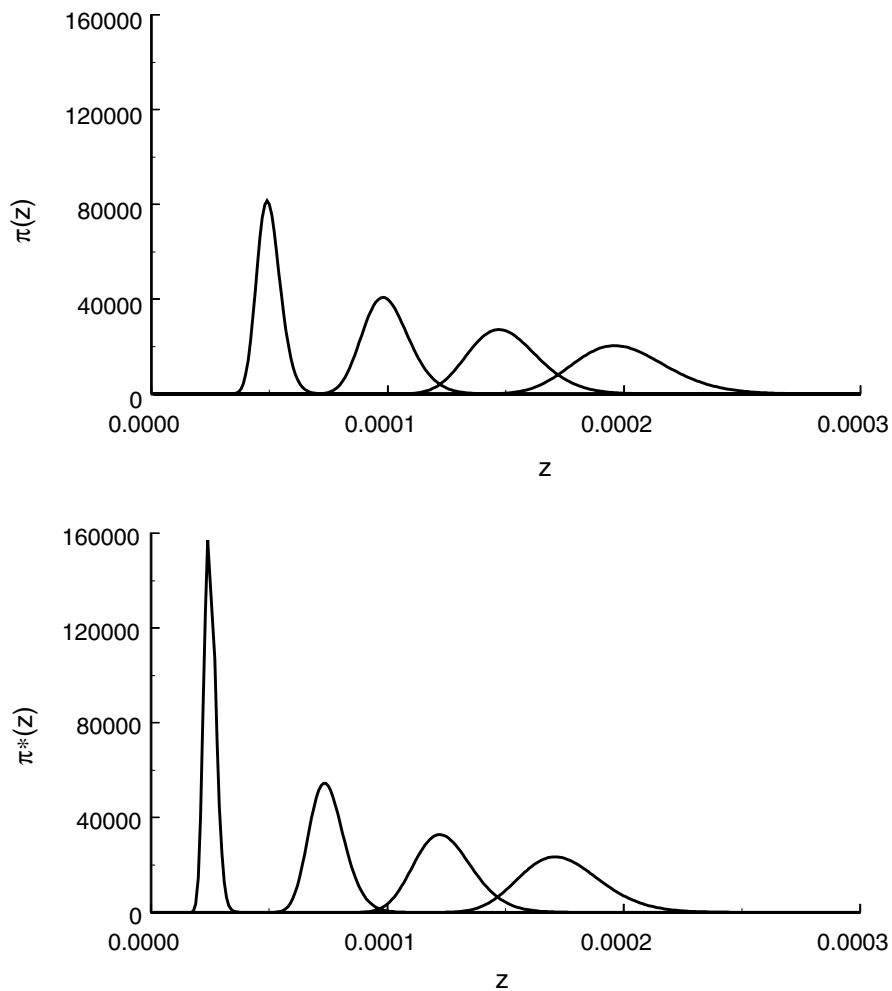


Fig. 4. Upper panel shows four different distributions of $z = \sigma_c^2 \gamma$ with mean values 0.00005, 0.00010, 0.00015 and 0.00020 in the lognormal model. The coefficient of variation of z is $C = 0.1$ and stochastic growth rate $s = 0.05$ in all these distributions. The lower panel shown the corresponding four distributions of z in the community which are all changed towards somewhat smaller values by parameter sampling.

Although this model determines uniquely the parameter sampling of $\gamma\sigma_e^2$ it will not uniquely determine the species abundance distribution. It appears from Eq. (6) that the integration in Eq. (2), which is required to find the abundance distribution, requires a specification of the joint distribution of γ and σ_e^2 and not just the distribution of $\gamma\sigma_e^2$. The simplest solution is found by keeping γ constant so that σ_e^2 then has the inverse gamma distribution with shape parameter k and scale parameter α/γ so that $z = \gamma\sigma_e^2$ still has the above inverse gamma distribution. Then, the species abundance distribution, which is derived in the Appendix, is similar to a Student t -distribution rather than the normal for the log abundances. This distribution has somewhat heavier tails than the lognormal species abundance distribution although the form is rather similar to a lognormal.

It is shown in Appendix and demonstrated in Fig. 5. that the variance of log abundance in this model decreases as the coefficient of variation in σ_e^2 across species increases, provided that the mean of σ_e^2 is kept constant. This may at first seem contra-intuitive. However, what happens is that increasing variance gives many species with large environmental variances that are likely to extinct soon after invasion, as well as many species with small variances which are likely to be resident species for a long time. The overall variation of σ_e^2 for resident species will therefore be small when the variation at invasion increases.

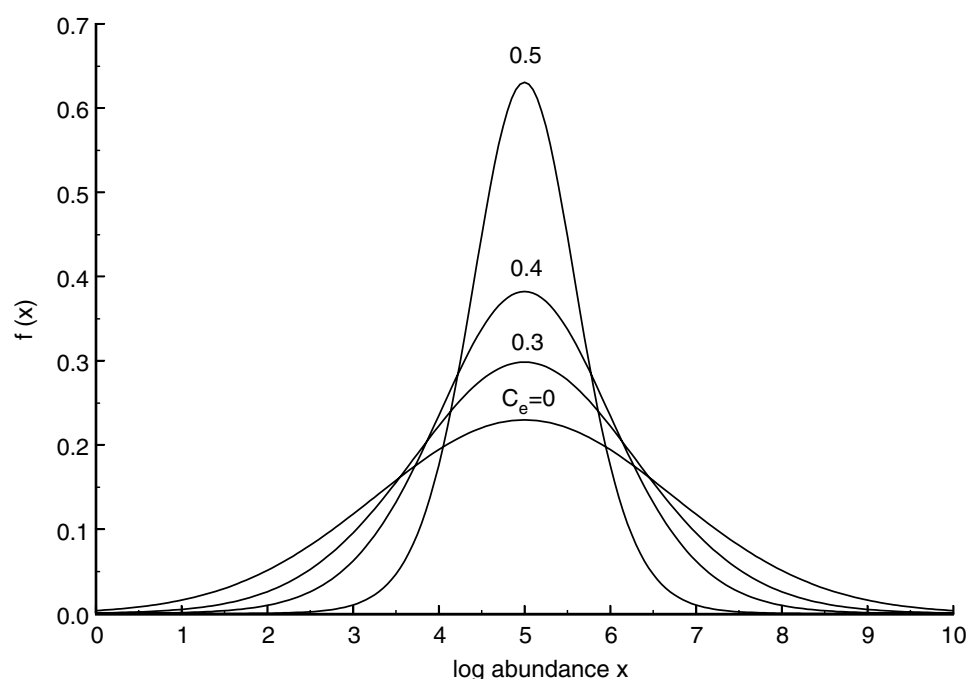


Fig. 5. The student type of log abundance distribution generated from the lognormal abundance distribution by adding stochasticity in the environmental variance σ_e^2 at invasion, assuming that σ_e^2 at invasion has the inverse gamma distribution across species. The mean environmental variance $E\sigma_e^2$ at invasion is kept constant equal to 0.01 while the coefficient of variation $C_e = SD(\sigma_e^2)/E\sigma_e^2$ varies between graphs. The mean carrying capacity $E(\ln x) = s/\gamma$ on the log scale is 5. The graph for $C_e = 0$ is the lognormal species abundance distribution.

7. Parameter sampling in mixtures of gamma models

The gamma distribution of abundances is obtained from the logistic type of density-regulation [13], that is $\mu(x) = rx - \gamma x^2$ and constant environmental variance so that $v(x) = \sigma_e^2 x^2$, where x is now the abundance. Using Eq. (1) the inhomogeneous Poisson process for the abundances in the community takes the form

$$\lambda(x; \theta) = \frac{2\beta}{\sigma_e^2} e^{2\gamma/\sigma_e^2} x^{2s/\sigma_e^2 - 1} e^{-(2\gamma/\sigma_e^2)x}, \tag{12}$$

where now $\theta = (r, \gamma, \sigma_e^2)$ and s is the stochastic growth rate $r - \sigma_e^2/2$. Scaling this function we see that the abundance distribution is the gamma distribution with shape parameter $2s/\sigma_e^2$ and scale parameter $2\gamma/\sigma_e^2$.

Notice that we obtain Fisher’s model exactly by choosing $s = 0$ and that this is a well defined model with $ES = \int \lambda(x) dx$. If we choose the smallest possible abundance to be zero rather than one, then ES is infinite in accordance with [1]. For this model the species abundance distribution $f(x)$ does not exist, which illustrates one advantage of describing abundance distributions by the inhomogeneous Poisson process rather than distribution of abundance since the Poisson process is still well defined. Actually, we have a well defined model with infinite number of species also for $s < 0$, corresponding to the extended negative binomial model defined by Engen [2,27]. However, since the expected number of individuals in the community is $\int x \lambda(x) dx$ we must require that this integral converge, which occurs if the shape parameter of the gamma $2s/\sigma_e^2 > -1$. This corresponds exactly to the lower limit of -1 for the shape parameter in the extended negative binomial model [2,27]. Interestingly, this requirement is the same as requiring that the deterministic growth rate $r = s + \sigma_e^2/2$ is positive. On the other hand, the problem of an infinite number of species does not occur when the lowest possible abundance is chosen as 1.

Integrating Eq. (12) we now obtain the preference function

$$w(\theta) = \frac{2\Gamma(2s/\sigma_e^2)}{\sigma_e^2} \left(\frac{\sigma_e^2}{2\gamma}\right)^{2s/\sigma_e^2} e^{2\gamma/\sigma_e^2} [1 - P(2s/\sigma_e^2, 2\gamma/\sigma_e^2)], \tag{13}$$

where P is the incomplete gamma function ([35], pp. 260) defined as

$$P(k, x) = \int_0^x \frac{1}{\Gamma(k)} u^{k-1} e^{-u} du. \tag{14}$$

7.1. Heterogeneity in γ

We can illustrate the effect of heterogeneity in γ in this model by some simple analytical results. Assuming that the heterogeneity is given by variability in γ only, and that this parameter is gamma-distributed with shape parameter k and scale parameter α we find the abundance model

$$\lambda(x) = \frac{2\beta}{\sigma_e^2} x^{2s/\sigma_e^2 - 1} \left(\frac{\alpha\sigma_e^2/2}{\alpha\sigma_e^2/2 - 1 + x}\right)^k \tag{15}$$

defined for $x \geq 1$. Since this function is proportional to a beta-distribution of the second kind ([28], p. 151), the species abundance distribution must be such a distribution censored at $x = 1$. It is interesting to notice that this is exactly the species abundance distribution used by Kempton [29] who treated this as a generalized form of Fisher's log series distribution [1] when adding Poisson sampling to the model. This model turned out to give a remarkably good fit to data of insect catches from traps at Rothamstead Experimental Station. Writing the beta-distribution of the second kind on the standard form

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{b^p x^{p-1}}{(1+bx)^{p+q}} \quad (16)$$

we see that $p = 2s/\sigma_e^2$, $q = k - 2s/\sigma_e^2$ and $b = 1/(\alpha\sigma_e^2/2 - 1)$, so this model is stationary only if all three parameters are positive, that is if $k > 2s/\sigma_e^2$ and $\alpha > 2/\sigma_e^2$.

Further, inserting Eq. (13) into Eq. (5) and assuming parameter values so that the last factor involving the incomplete gamma function can be approximated by 1 (which is realistic for large carrying capacities, that is large values of s/γ), we find that the distribution of γ in the community also is a gamma distribution, but with shape parameter $k - 2s/\sigma_e^2 = q + 1$ and scale parameter $\alpha - 2/\sigma_e^2 = 2/(b\sigma_e^2)$. Hence, the variance of γ in the community is $(k - 2s/\sigma_e^2)/(\alpha - 2/\sigma_e^2)^2$ while the variance for invading species is k/α^2 . From this we find that the variance of γ is increased by parameter sampling if $k > \alpha^2 s / (2\alpha - 2/\sigma_e^2)^2$ and otherwise reduced. The squared coefficient of variation of γ , however, which is $1/k$ for invading species and $1/(k - 2s/\sigma_e^2)$ in the community, is always increased by the sampling process.

8. Discussion

Although variation in abundance, as described by these models, is of fundamental interest in understanding community ecology, it is strange that relatively little effort has been made in relating possible variation between species characteristics other than abundance to the classical concept of species abundance models. Exceptions are the interesting non-neutral resource competition model by Tilman [30], who also give a number of references to empirical studies finding significant correlations among species traits and their abundances ([31] and others), followed up by empirical findings of strong correlations in grassland communities [32]. Although none of these traits are the kind of population dynamic parameters considered here, it is likely in general that differences in traits lead to differences in dynamics, which again implies that parameters defining the dynamics vary among species in the community. Distribution of life history parameters among species [16] are also widely discussed in the literature, and even distribution of population dynamical parameters has been estimated [17] but not related to the concept of species abundance distributions and diversity.

In addition to ignoring environmental stochasticity, neutral species abundance models ignores the possibility of variation in parameters describing the dynamics of species. The basic assumption of neutral models is the equality of dynamical properties of species in analogy with the equality of neutral genes in population genetics [33]. Engen and Lande [13,18], and Lande et al. [12] considered a very different class of species abundance models, that could

generate the gamma as well as the lognormal model, allowing variation in dynamic parameters across species. Fitting such models to community data on tropical butterflies sampled in space and time, Engen et al. [34] found that variation in parameters (growth rates and carrying capacities) explained more than half of the variance in log abundances of this community, which indicates that parameter variation may be the major part of the variation of species abundance in a community. Another indication of heterogeneity was given by Sæther et al. [20] who estimated the dynamical parameters of many closely related bird species and discussed how the variation in dynamic parameters across species would change the distribution of these parameters by extinction processes.

The main result of the present paper, except from showing that several types of models can be generated by different parameter distributions, is that nature's way of selecting species (and parameters) through extinctions does not in general reduce variation in parameters among species. This process that we have called parameter sampling, acting through a preference function $w(\theta)$, may change the mean values of parameters, promoting larger stochastic growth rates and smaller strength of density regulation, both leading to larger carrying capacities. The same process also tends to reduce the mean environmental stochasticity of a community because large stochasticity leads to short expected time to extinction. The variance or coefficient of variation in these parameters are, however, not in general reduced, and may actually increase a lot by nature's sampling of species. This is an important observation that supports the hypothesis of between species parameter variation.

It is difficult to reach a general conclusion saying when the effect of parameter sampling will be an increase or a decrease of the variance. As a simple illustration, however, it is informative to consider the case that invading species can only take two values θ_1 and θ_2 with probabilities p_1 and p_2 . Writing simply w_1 and w_2 for the preference of the two types, the distribution π^* will be defined for the same two points, but with probabilities proportional to $w_1 p_1$ and $w_2 p_2$. So if for example $p_1 = p_2 = 1/2$, giving variance 0.25 for invading species, then $w_1 \neq w_2$ will lead to a distribution with smaller variance. On the other hand, if $p_1 < p_2$ and $w_1 > w_2$, then the variance may increase. To exemplify further, if $p_1 = 0.1$ and $p_2 = 0.9$ while $w_1 = 10$ and $w_2 = 10/9$ then the variance at invasion is 0.09 while being 0.25 for resident species, which is a large increase. Another illustration is the case when a parameter θ is normally distributed while the preference is proportional to $e^{\alpha\theta}$. Then it is straightforward to check that the mean of θ is increased for a positive α while the variance is unchanged. To obtain an increase in the variance the preference must increase more strongly with θ supporting more heavily large values that are rare at invasion, which is actually the case for our mixture of lognormal models.

Although the selection of parameters may be quite strong, our simulation results show that the process of convergence toward the stationary solution, starting with an empty community, is rather slow with the parameters used in Fig. 3. Thousands of transitions (speciation or extinction) are required, so it remains an open question whether most communities have reached stationarity.

The only way of getting a more realistic picture of community dynamics is to collect time series community samples. Spatial samples may also give additional insight. There is definitely a need for improved statistical methods for performing analysis of such data with possible rejection of the hypothesis of homogeneity, that is, all species having exactly the same kind of dynamics.

Acknowledgement

This project has been supported by the Norwegian Research Council, Project 159571/V40, Stochastic dynamics of bird communities.

Appendix

We now pass from the lognormal abundance model defined by Eq. (6) to the mixture found by assuming that the environmental variance σ_e^2 have the inverse gamma distribution with shape parameter k and scale parameter α/γ so that the distribution of $z = \gamma\sigma_e^2$ is the inverse gamma with parameters k and α as in the main text. Performing the integration in equation Eq. (2) writing $y = \sigma_e^2$ we find

$$\lambda(x) = c \int_0^\infty y^{-(k+1)-1} e^{-[\gamma(x-s/\gamma)^2 + (\alpha-s^2)/\gamma]/y} dy,$$

where c is a constant, giving

$$\lambda(x) = c' [\gamma(x - s/\gamma)^2 + (\alpha - s^2)/\gamma]^{-(k+1)},$$

where c' is another constant. Using the well known Student's t -distribution with ν degrees of freedom ([28], p. 374)

$$g(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\nu^{1/2} \Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})} (1 + t^2/\nu)^{-(\nu+1)/2}$$

and performing the transformation

$$t = (x - s/\gamma)\gamma \sqrt{\frac{2k+1}{\alpha - s^2}}$$

we see that the distribution of t is the t -distribution with $\nu = 2k + 1$ degrees of freedom (which in general does not have to be an integer). Performing the backwards transformation we then find the distribution of log abundance x ignoring censoring at $x = 0$

$$f(x) = \frac{\Gamma(k+1)}{\Gamma(1/2)\Gamma(k+1/2)} \eta^{-1/2} (1 + (x - s/\gamma)^2/\eta)^{-(k+1)},$$

where $\eta = (\alpha - s^2)/\gamma^2$. Generally, the distribution of log abundance is this distribution censored at $x = 0$ as for the lognormal model.

As this distribution has been constructed by heterogeneity in the environmental variances, it may be easier to interpret the resulting distribution using the mean value across species of the variances of log abundances (ignoring extinctions) which is

$$\sigma^2 = E\left(\frac{\sigma_e^2}{2\gamma}\right) = \frac{\alpha}{2(k-1)\gamma^2}.$$

Further, writing $\mu = s/\gamma$ for the mean values of log abundances we see that $\eta = 2(k - 1)\sigma^2 - \mu^2$. Then we also see that we obtain the normal distribution with mean μ and variance σ^2 in the limit as k approaches infinity, that is, as the heterogeneity approaches zero, in according with the initial homogeneous model. Using the fact that the variance of the t-distribution is $v/(v - 2)$ for $v > 2$ and writing $C_e^2 = 1/(k - 2)$ for the squared coefficient of variation of σ_e^2 across species and $C = \sigma/\mu$, we find that the variance of the log abundance relative to this variance for the homogeneous model is

$$\frac{\text{var}(x)}{\sigma^2} = \frac{\eta}{2k - 1} = \frac{2(1 + C_e^{-2}) - C^{-2}}{2C_e^{-2} + 3}$$

which is a decreasing function of C_e .

References

- [1] R.A. Fisher, A.S. Corbet, C.B. Williams, The relation between the number of species and the number of individuals in a random sample of an animal population, *J. Anim. Ecol.* 12 (1943) 42.
- [2] S. Engen, *Stochastic Abundance Models with Emphasis on Biological Communities and Species Diversity*, Chapman & Hall, London, 1978.
- [3] F.W. Preston, The commonness and rarity of species, *Ecology* 29 (1948) 254.
- [4] F.W. Preston, The canonical distribution of commonness and rarity, *Ecology* 43 (1962) 185.
- [5] R.M. May, Patterns of species abundance and diversity, in: M.L. Cody, J.L. Diamond (Eds.), *Ecology and Evolution of Communities*, Harvard University, Cambridge, Massachusetts, 1975, p. 81.
- [6] E.C. Pielou, *Ecological Diversity*, Wiley, New York, 1975.
- [7] A.E. Magurran, *Ecological Diversity and its Measurement*, Princeton University, Princeton, 1988.
- [8] H. Caswell, Community structure: a neutral model analysis, *Ecol. Monogr.* 46 (1976) 327.
- [9] S.P. Hubbell, *The Unified Theory of Biodiversity and Biogeography*, Princeton University, Princeton, 2001.
- [10] G. Bell, Distribution of abundance in neutral communities, *Am. Nat.* 155 (2000) 606.
- [11] G. Bell, Neutral macroecology, *Science* 293 (2001) 2413.
- [12] R. Lande, S. Engen, B.-E. Sæther, *Stochastic Population Models in Ecology and Conservation*, Oxford University, Oxford, 2003.
- [13] S. Engen, R. Lande, Population dynamic models generating species abundance distributions of the gamma type, *J. Theor. Biol.* 178 (1996) 325.
- [14] S.P. Hubbell, Towards a theory of biodiversity and biogeography on continuous landscapes, in: G.R. Carmichael, G.E. Folk, J.L. Schnoor (Eds.), *Preparing for Global Change: A Midwestern Perspective*, SPB Academic, Amsterdam, 1995, p. 171.
- [15] S.P. Hubbell, A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs, *Coral Reefs* 16 (suppl.) (1997) 9.
- [16] P.H. Harvey, M.D. Pagel, *The Comparative Method in Evolutionary Biology*, Oxford University, Oxford, 1991.
- [17] B.-E. Sæther, R. Lande, S. Engen, H. Weimerskirch, M. Lillegard, R. Altwegg, P.H. Becker, T. Bregnballe, J.E. Brommer, R.H. McCleery, J. Merila, E. Nyholm, W. Rendell, R.R. Robertson, P. Tryjanowski, M.E. Visser, Generation time and temporal scaling of bird population dynamics, *Nature* 436 (2005) 99.
- [18] S. Engen, R. Lande, Population dynamic models generating the lognormal species abundance distribution, *Math. Biosci.* 132 (1996) 169.
- [19] C.K. Kelly, M.G. Bowler, Coexistence and relative abundance in forest trees, *Nature* 417 (2002) 437.
- [20] B.-E. Sæther, S. Engen, A.P. Møller, H. Weimerskirch, M.E. Visser, W. Fiedler, E. Matthysen, M.M. Lambrechts, A. Badyaev, P.H. Becker, J.E. Brommer, D. Bukacinsky, M. Bukacinsky, H. Christensen, J. Dickinson, C. Du Feu, F.R. Gehlbach, D. Heg, H. Hötker, J. Merilä, J.T. Nielsen, W. Rendell, R.J. Robertson, D.L. Thomsen, J. Török,

- P. Van Hecke, Life history variation predicts the effects of demographic stochasticity on avian population dynamics, *Am. Nat.* 164 (2005) 793.
- [21] R.M. May, Stability in randomly fluctuating versus deterministic environments, *Am. Nat.* 107 (1973) 621.
- [22] S. Engen, Ø. Bakke, A. Islam, Demographic and environmental stochasticity – concepts and definitions, *Biometrics* 54 (1998) 39.
- [23] R.H. MacArthur, On the relative abundance of bird species, *Proc. Nat. Sci. Washington* 43 (1957) 293.
- [24] G. Sugihara, Minimal community structure: an explanation of species abundance patterns, *Am. Nat.* 116 (1980) 770.
- [25] G. Sugihara, L.-F. Bersier, T.R.E. Southwood, S.L. Pimm, R.M. May, Predicted correspondence between species abundance and dendrograms of niche similarities, *Proc. Natl. Acad. Sci. USA* 100 (2003) 5246.
- [26] S. Karlin, H.M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.
- [27] S. Engen, On species frequency models, *Biometrika* 61 (1974) 263.
- [28] M.G. Kendall, A. Stuart, *The advanced theory of statistics*, 3rd ed., vol. 1, Griffin, London, 1969.
- [29] R.A. Kempton, A generalized form of Fisher's logarithmic series, *Biometrika* 62 (1975) 29.
- [30] D. Tilman, Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly, *Proc. Natl. Acad. Sci. USA* 101 (2004) 10854.
- [31] B.J. McGill, A test of the unified neutral theory of biodiversity, *Nature* 422 (2003) 881.
- [32] W.S. Harpole, D. Tilman, Non-neutral patterns of species abundance in grassland communities, *Ecol. Lett.* 9 (2006) 15.
- [33] W.J. Ewens, The sampling theory of selectively neutral alleles, *Theor. Popul. Biol.* 3 (1972) 87.
- [34] S. Engen, R. Lande, T. Walla, P.J. DeVries, Analyzing spatial structure of communities by the two-dimensional Poisson lognormal species abundance model, *Am. Nat.* 160 (2002) 60.
- [35] M. Abramowitz, I.A. Stegun (Eds.), *Handbook of Mathematical Functions*, Dover, New York, 1972.