Ecography

ECOG-00175
Bellier, E., Grøtan, V., Engen, S., Schartau, A.-K., Herfindal, I. and Finstad, A. G. 2013. Distance decay of similarity, eects of environmental noise and heterogeneity among species in the spatio-temporal dynamics of a dispersal-limited community. – Ecography 36: xxx–xxx.

Supplementary material

APPENDIX A

Computation of the likelihood function for the two-stage sampling procedure

The probability that the number of individuals of a species in two samples from different locations with sampling intensities ν_1 and ν_2 is (x, y) is given by the two-dimensional Poisson lognormal distribution with parameters $(\mu_1 + \ln \nu_1, \mu_2 + \ln \nu_2, \sigma_2^2, \sigma_2^2, \rho)$ which we for simplicity of notation write as $p(x, y; \nu_1, \nu_2)$. Similarly, for the marginal distributions, which are both univariate Poisson lognormals, we write generally $p(x; \nu)$ for a sampling intensity ν. We assume that sampling is performed in two steps as described in the main text. Let q_1 and q_2 be the sampling intensities of the two small samples where complete counts are performed, and let the sampling intensities for the large samples where only presence absence is recorded be 1. There is no loss of generality in defining these sampling intensities as 1 because sampling intensity is confounded with the mean values and have not effect on the correlation. The q_1 and q_2 may be interpreted as the fractions of the large samples where all individuals are counted. For further details about the bivariate Poisson lognormal see Engen et al. (2002, 2011). In order to find the likelihood function for this procedure it is preferable to define new variables based on the sampling method. We define new adjusted counts as -1 if the species is not observed at all, 0 if it is not observed in the small sample but seen in the large one, and the actual count in the small one if it is seen and counted in the small sample. The distribution of these redefined 'counts' we denote $s(x,y)$, where x and y here can take integer values $-1, 0, 1, 2, ...$. Then there is a unique relation between $s(x,y)$ and the two-dimensional Poisson lognormal distribution. For positive values x and y we see immediately that

$$s(x,y) = p(x,y;q_1,q_2)$$

$$s(x,-1) = p(x,0;q_1,1)$$

$$s(-1,y) = p(0,y;1,q_2)$$

$$s(-1,-1) = p(0,0;1,1).$$

To find the values involving zeros we introduce the notation $s(x,+) = \sum_{y=1}^{\infty} s(x,y)$, and similarly for $s(+,y)$ and $s(+,+)$. Then the event that the counts in the first small sample is some positive x is the union of three disjoint events giving $p(x,q_1) = s(x,+) + s(x,-1) + s(x,0)$. Using the fact that $s(x,+) = p(x;q_1) - p(x,0,q_1,q_2)$ this yields

$$s(x,0) = p(x,0;q_1,q_2) - p(x,0;q_1,1),$$

and by symmetry

$$s(0,y) = p(0,y;q_1,q_2) - p(0,y;1,q_2).$$

Further we observe that the event with probability $p(0,0;q_1,1)$ is a union of two disjoint events giving $p(0,0;q_1,q_2) = s(0,-1) + s(-1,-1)$ giving

$$s(0,-1) = p(0,0;q_1,1) - p(0,0;1,1),$$

and by symmetry

$$s(-1,0) = p(0,0;1,q_2) - p(0,0;1,1).$$

Finally, the event with probability $p(0,0;q_1,1)$ is also a union of four disjoint events, giving $p(0,0;q_1,q_2) = s(0,0) + s(-1,0) + s(0,-1) + s(-1,-1)$ giving the last expression

$$s(0,0) = p(0,0;q_1,q_2) - p(0,0;1,q_2) - p(0,0;q_1,1) + p(0,0;1,1).$$

To find the likelihood function we must use the distribution conditioned on the event that the species is observed in at least one of the samples which has probability $1 - s(-1,-1)$. Hence, for n observations (x_i, y_i) with at least one of the components larger than -1, the log likelihood function is

$$\ln L = \sum_{i=1}^{n} \ln s(x_i, y_i) - n \ln[1 - s(-1,-1)].$$

APPENDIX B

Stochastic model

We apply the model of Engen and Lande (1996) assuming that the log abundance $X_i(z,t)$ of species i at time t and location z, follows the Ornstein-Uhlenbeck process (Karlin and Taylor 1981),

$$\begin{aligned} dX_i(z,t) &= [r_i - \delta X_i(z,t)]dt + \sigma_e dB_e(z,t) \\ &+\sigma_d dB_d(z,t)/\sqrt{(}K_i), \text{ for i=1,2,\ldots,S.} \end{aligned} \tag{1}$$

Here $B_e(z,t)$ and $B_d(z,t)$ are independent standard Brownian motions, and the parameters σ_e^2 and σ_d^2 are commonly known as the environmental and demographic variances. The stochastic growth rate of species i is r_i in the absence of density regulation (small population sizes) and is normally distributed among species with variance σ_r^2. The parameter δ is the strength of density-regulation, giving carrying capacity $K_i = e^{r_i/\delta}$ also varying among species. The environmental noise component may be split into two components $\sigma_e^2 = \sigma_c^2 + \sigma_s^2$ in agreement with the relationship $\sigma_e dB_e(z,t) = \sigma_c dB_c(z,t) + \sigma_s dB_i(z,t)$, for species i, $i = 1,2,\ldots,S$, where the different Brownian motions are all independent. Here, $B_c(z,t)$ is a noise process that is common for all species, while $B_i(z,t)$ is a species specific noise for species i. The species specific noise processes $B_i(z,t)$ have the same spatial autocorrelation function $\rho_s(v)$ defined by the relation $EdB_i(z,t)dB_i(z+v,t+u) = \rho_s(v)dt$. Due to properties of the Ornstein-Uhlenbeck process, the noise term that is common will give the same additive stochastic contribution to the log abundance of all species and will therefore be confounded with the expectation. Then the bivariate stationary distribution of $[X_i(z,t), X_i(z+v,t+u)]$ in the conditional model conditioned on common noise is the bivariate normal. The variance of each component is $[\sigma_s^2 + (\sigma_d^2 e^{-r/\delta})]/2\delta$, while the covariance is $\sigma_s^2 \rho_s(v)/2\delta$. Using general properties of the Ornstein-Ulhenbeck process, Engen et al. (2002) found for a time difference u between the observations that

$$\text{cov}[X_i(z,t), X_i(z+v,t+u)] = \text{cov}[X_i(z,t), X_i(z+v,t)]e^{-\delta u}. \tag{2}$$

For this model with between species variation in r, Engen et al. (2002) showed that

$$\text{cov}[X_i(z,t), X_i(z+\nu,t+u)] = \frac{\sigma_s^2 \rho_s e^{-\delta u}}{2\delta} + \frac{\sigma_r^2}{\delta^2}, \tag{3}$$

and the stationary variance σ^2 of any X_i can be decomposed into three additive components

$$\sigma^2 = \frac{\sigma_s^2}{2\delta} + \frac{\sigma_d^2}{2\bar{K}\delta} + \frac{\sigma_r^2}{\delta^2}, \tag{4}$$

where $1/\bar{K} = E_r e^{-r/\delta}$. The first term is the environmental component due to independent environmental noise terms with variance σ_s^2. The second term is the demographic component, and the last term is the interspecies heterogeneity component due to the heterogeneity between the species determined by the variance σ_r^2 of r among species. Engen et al. (2002) showed by combining the expression of the variance and covariance that the joint spatial and temporal correlation of relative log abundance of species in the community then takes the form,

$$\rho(u, v) = \frac{\sigma_s^2[\rho_s(v)]e^{-\delta u} + 2\sigma_r^2/\delta}{\sigma_s^2 + \sigma_d^2/\bar{K} + 2\sigma_r^2/\delta}. \quad (5)$$

We now write $\mathbf{X}(\mathbf{z}, t)$ for the vector of the log abundances at locations z and t. In the model conditioned on the common noise, the pairs $[X_i(z, t), X_i(z+ v, t + u)]$ for $i = 1, 2, \ldots, S$ are then independent bivariate normally distributed variables with variance $[\sigma_s^2 + (\sigma_d^2/\bar{K} + (2\sigma_r^2/\delta))]/2\delta$ for each component and correlation $\rho(u, v)$ between them. Then log abundances composing the vectors $\mathbf{X}(\mathbf{z}, t)$ and $\mathbf{X}(\mathbf{z} + \mathbf{v}, t + u)]$ are each independent and multinormally distributed with the same parameters. The components with the same subscript represents the log abundances of the same species at the two locations and have correlation $\rho(u, v)$. An estimate of the community correlation in relative log abundances for a pair of samples can be obtained by fitting the observed joint species abundance in the two samples to a bivariate Poisson-lognormal distribution (Engen et al. 2002; Lande et al. 2003; Engen et al. 2011a). The function $\rho_s(v)$ is the correlation between species specific noise terms at distance v. Here, we assume that the shape of the intraspecific spatio-temporal autocorrelation function $\rho_s(v)$ due to spatially autocorrelated specific environmental stochasticity has the form $\rho_\infty + (1 - \rho_\infty)h(v)$ where ρ_∞ represents the common environmental noise at infinite distance and $h(v)$ decreases from 1 to 0 as distance increases from 0 to infinity. We assume an isotropic decline of correlation with increasing distance and time lag and choose a form of autocorrelation $\rho_s(v)$ that gives rise to a positive semi-definite variance-covariance matrix. One likely positive definite autocorrelation function is a Gaussian type of function $h(v) = e^{-1/2(v^2/l^2)}$ where l is a measure of the spatial scale of autocorrelation. Inserting this in Eq. 2, we see that the spatio-temporal autocorrelation $\rho(u, v)$ of relative abundance takes the form

$$\rho(u, v) = \frac{\sigma_s^2[\rho_\infty + (1 - \rho_\infty)e^{-1/2(v^2/l^2)}]e^{-\delta u} + 2\sigma_r^2/\delta}{\sigma_s^2 + \sigma_d^2/\bar{K} + 2\sigma_r^2/\delta}. \qquad (6)$$

We included environmental dissimilarities W in the spatio-temporal correlation function in Eq. 6 to estimate the contribution of different environmental variables to spatio-temporal fluctuation in species relative abundances. The estimation of environmental dissimilarity was based on observed environmental variables. The time varying environmental variables were standardized within each lake. The dissimilarity W within and among lakes at different time lags u was calculated as the absolute value of the difference in the standardized variable. The spatio-temporal correlation of relative log abundance of species in the community then takes the form

$$\rho(u, v) = A[e^{-\delta u} - 1] + B[e^{-1/2((v+\beta W)^2/l^2)-\delta u} - 1] + \rho_0, \qquad (7)$$

where $v + \beta W$ is a generalized distance measure, β is an unknown parameter associated with the environmental dissimilarity and ρ_0 is the correlation at time lag 0 and a null distance.