

Supporting Information for *Local intraspecific aggregation in phytoplankton model communities: spatial scales of occurrence and implications for coexistence* by C. Picoche, W.R. Young & F. Barraquand.

S1 Derivation of the turbulent map

We show here how to derive a discrete-time map for turbulence from the continuous-time formula. We consider that the velocity field $\mathbf{u}^T = (u_x, u_y, u_z)$ at position $\mathbf{x}^T = (x, y, z)$ alternates between the three dimensions during a period τ , so that

$$\mathbf{u}^T(\mathbf{x}, t) = \begin{cases} (U \cos(ky + \phi), 0, 0) & \text{for } n\tau \leq t < (n + \frac{1}{3})\tau \\ (0, U \cos(kz + \theta), 0) & \text{for } (n + \frac{1}{3})\tau \leq t < (n + \frac{2}{3})\tau \\ (0, 0, U \cos(kx + \psi)) & \text{for } (n + \frac{2}{3})\tau \leq t < (n + 1)\tau. \end{cases} \quad (\text{S1})$$

The discrete-time map can be obtained by computing the displacement over a period, between $t = n\tau$ and $t + 1 = (n + 1)\tau$, with $\mathbf{x}(t + 1) = \mathbf{x}(t) + \int_{n\tau}^{(n+1)\tau} \mathbf{u}(\mathbf{x}, t) dt$, and knowing the initial position $\mathbf{x}(t)$. This can be solved in three steps (eqs. S2, S3 and S4). We start with

$$\begin{aligned} x(t + \tau/3) &= x(t) + \frac{U\tau}{3} \cos(ky(t) + \phi) \\ y(t + \tau/3) &= y(t) \\ z(t + \tau/3) &= z(t). \end{aligned} \quad (\text{S2})$$

Then,

$$\begin{aligned} x(t + 2\tau/3) &= x(t + \tau/3) \\ y(t + 2\tau/3) &= y(t) + \frac{U\tau}{3} \cos(kz(t) + \theta) \\ z(t + 2\tau/3) &= z(t). \end{aligned} \quad (\text{S3})$$

And finally,

$$\begin{aligned} x(t + \tau) &= x(t + \tau/3) \\ y(t + \tau) &= y(t + 2\tau/3) \\ z(t + \tau) &= z(t) + \frac{U\tau}{3} \cos(kx(t + \tau) + \psi). \end{aligned} \quad (\text{S4})$$

In the third step, we need z to be a function of $x(t + \tau)$, not $x(t)$, so that the volume is conserved (the determinant of the Jacobian matrix is equal to 1).

S2 Characteristics of standard spatial point processes

In order to get the reader acquainted with the spatial point process metrics that we use in the main text, we present here the analytical formulas and corresponding figures (Fig. S1 and S2)

for the pair correlation function, Ripley's K -function and dominance index for well-known point processes. We focus on the uniform distribution, i.e., the Poisson point process, and a clustered distribution, the Thomas point process. The Thomas point process is the result of a two-stage mechanism: a Poisson point process generates “parent points” around which “daughter points” are scattered, their locations following a Gaussian distribution centered on the parent location, with standard deviation σ . The numbers of parents and daughters per parent follow two Poisson distributions with mean N_p and N_d respectively. All solutions are given for three-dimensional spatial distributions.

S2.1 Pair correlation function

In the case of a Poisson point process,

$$\forall r \geq 0, g_{ii}(r) = 1. \quad (\text{S5})$$

For a Thomas point process, the expected value of the pcf is

$$g_{ii}(r) = 1 + \frac{1}{C_p} \frac{1}{(4\pi\sigma^2)^{3/2}} e^{-\left(\frac{r^2}{4\sigma^2}\right)} \quad (\text{S6})$$

where $C_p = N_p/V$ is the concentration/intensity of the parent process in the volume V .

For a random superposition of stationary point processes with marks (species) i and j , $\forall i \neq j, \forall r \geq 0, g_{ij}(r) = 1$ (Illian *et al.*, 2008, p. 326, eq. 5.3.13).

S2.2 Ripley's K -function

In the case of a Poisson point process,

$$\forall r \geq 0, K_{ii}(r) = \frac{4}{3}\pi r^3. \quad (\text{S7})$$

For a Thomas point process,

$$K_{ii}(r) = \frac{4}{3}\pi r^3 + \frac{1}{C_p \sigma \sqrt{\pi}} \left(\sigma \sqrt{\pi} \operatorname{erf} \left(\frac{r}{2\sigma} \right) - r e^{-\left(\frac{r}{2\sigma}\right)^2} \right). \quad (\text{S8})$$

For a random superposition of stationary point processes, $K_{ij}(r) = \frac{4}{3}\pi r^3$ (Illian *et al.*, 2008, p. 324, eq. 5.3.5).

S2.3 Dominance index

In the Poisson point process, $K_{ii}(r) = K_{ij}(r)$, which means that the dominance index can be reduced to ratios of concentrations:

$$\mathcal{D}_i(r) = \frac{C_i}{\sum_{j=1}^S C_j}. \quad (\text{S9})$$

In the Thomas process, using eq. S8,

$$\mathcal{D}_i(r) = \frac{C_i \left(\frac{4}{3} \pi r^3 + \frac{F(r)}{C_{p,i}} \right)}{C_i \frac{F(r)}{C_{p,i}} + \sum_j C_j \frac{4}{3} \pi r^3} \quad (\text{S10})$$

with $F(r) = \frac{1}{\sigma\sqrt{\pi}} \left(\sigma\sqrt{\pi} \operatorname{erf} \left(\frac{r}{2\sigma} \right) - r e^{-\left(\frac{r}{2\sigma} \right)^2} \right)$.

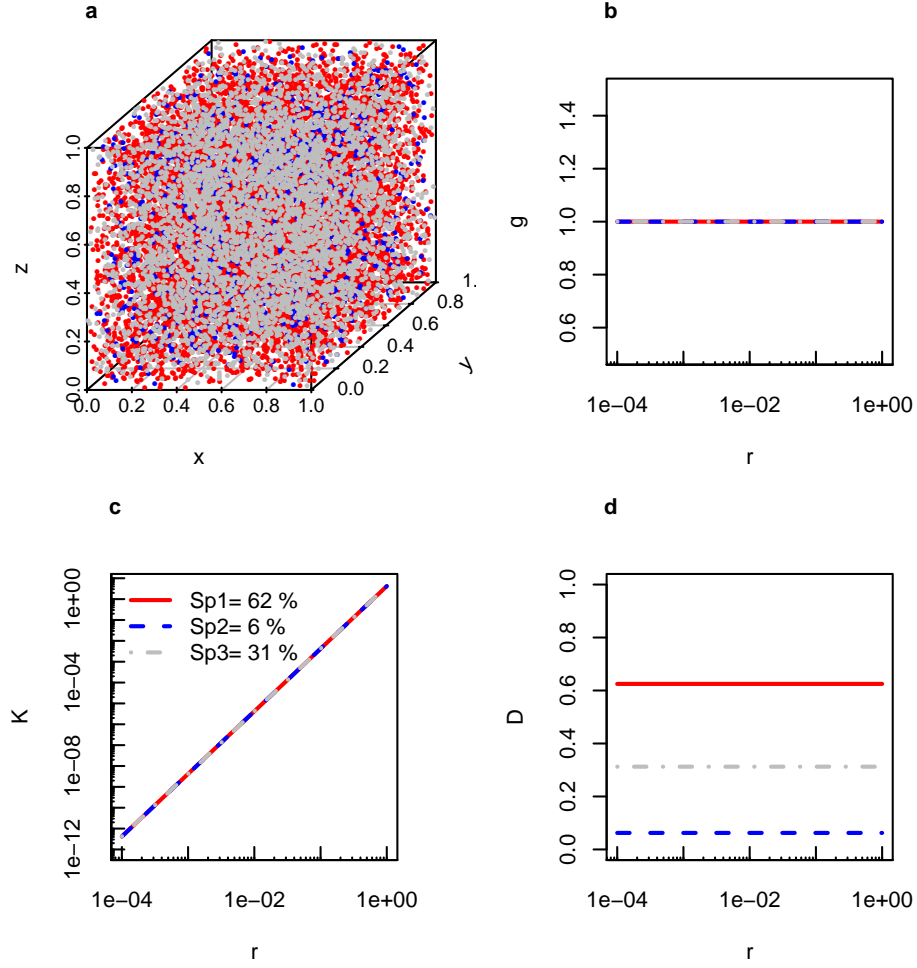


Figure S1: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K -function (c) and dominance index (d) for a Poisson point process in 3-species communities with different intensities (10000 cm^{-3} , 1000 cm^{-3} , 5000 cm^{-3} ; proportions in the community are given in the figure).

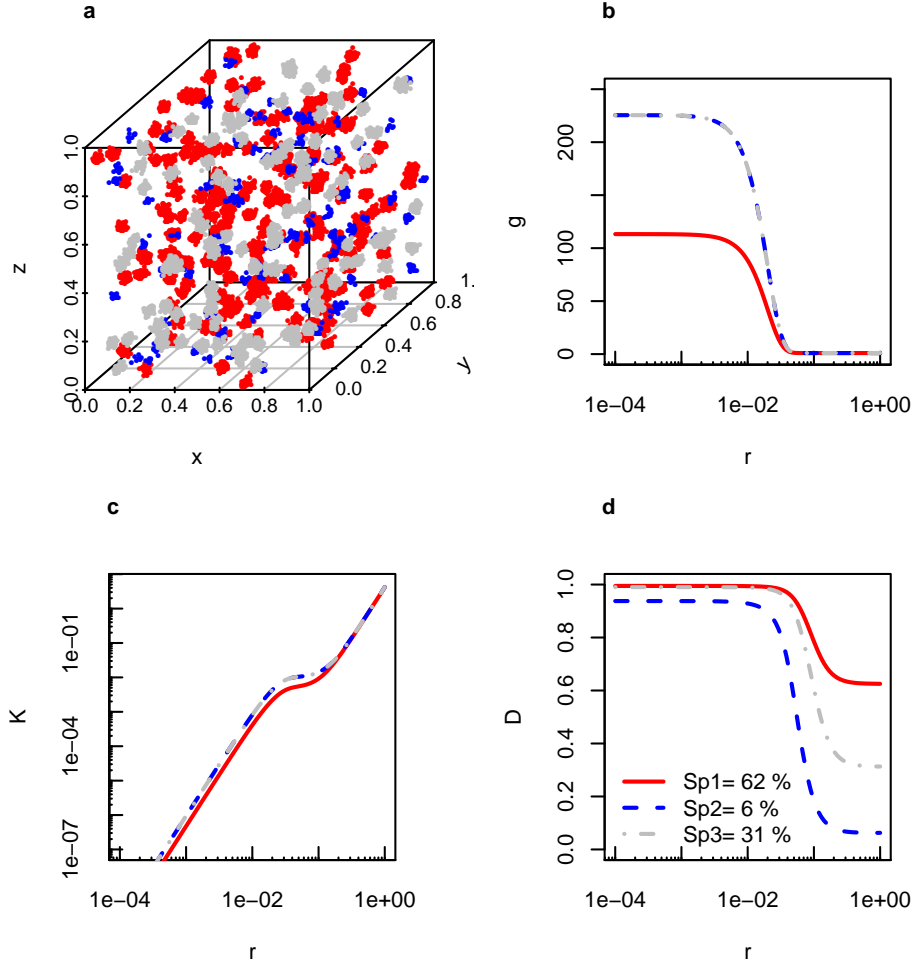


Figure S2: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K -function (c) and dominance index (d) for a Thomas point process in 3-species communities with different parent intensities (200 cm^{-3} , 100 cm^{-3} , 100 cm^{-3}), and different children per parent intensities (50, 10, 50; final proportions in the communities are given in the figure), with $\sigma = 0.01$.

S3 Convergence in time of the spatial characteristics of the BBM

The theoretical formulas of g , K and \mathcal{D} can be used to study the behaviour of the BBM. In the absence of advection, convergence cannot be reached in a reasonable timeframe: even a week is not long enough for the steady-state solution to be reached (see blue line in Fig. S3). However, the population-at-equilibrium hypothesis that we use cannot hold for such a long amount of time, which led us to use the time-dependent formulas shown in Eqs. 5, 9 and 12 in the main text.

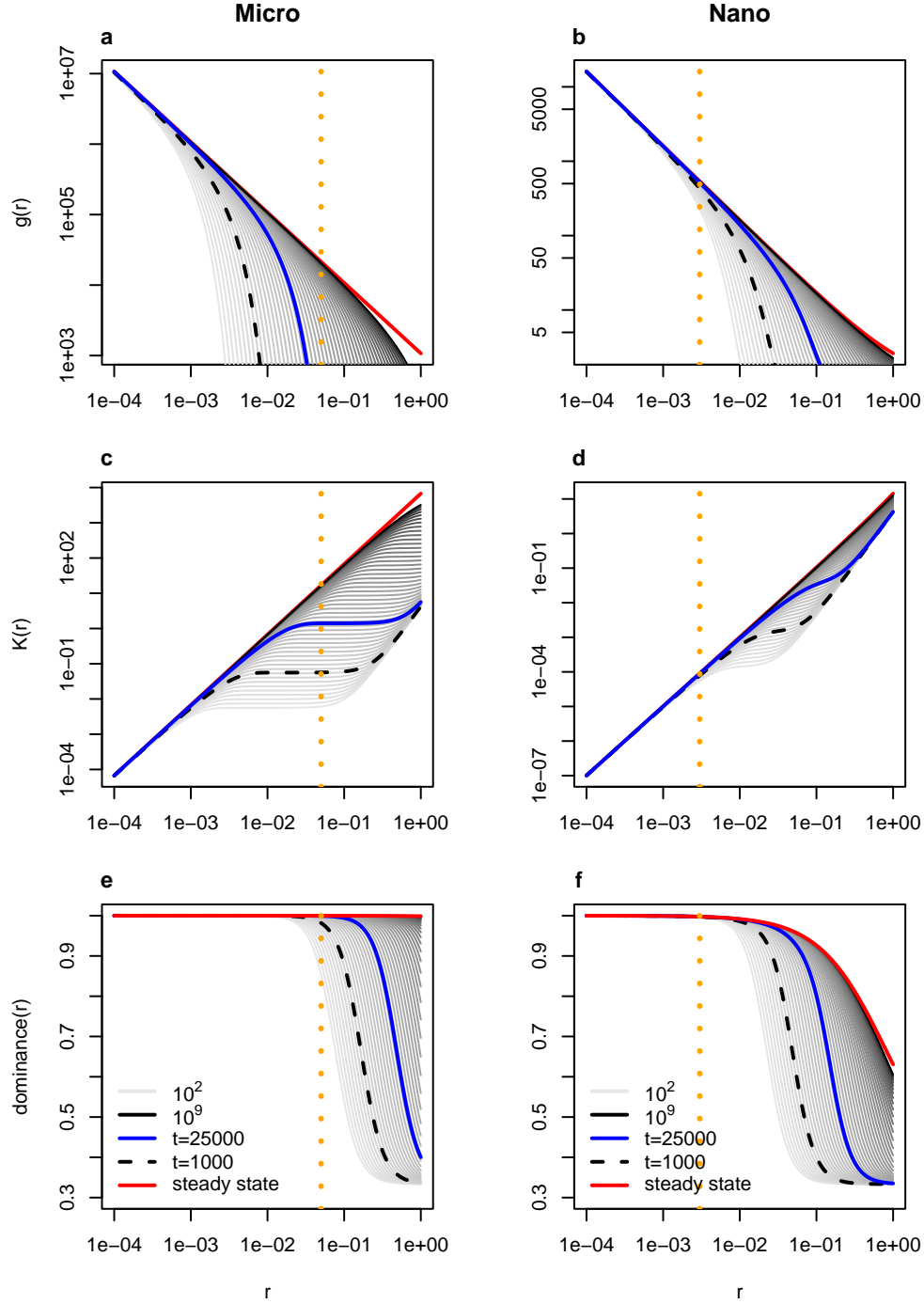


Figure S3: Intraspecific pair correlation function (a, b), Ripley's K -function (c,d) and dominance index (e,f) as a function of distance (in cm) for microphytoplankton and nanophytoplankton in the absence of advection, for a single species in a 3-species community with an even abundance distribution. Shorter timeframes are shown with light grey lines while longer ones are shown with darker shades. The theoretical value at steady state is shown in red. The duration currently used in the simulations ($t = 1000\tau$) is shown with dashed black lines. A duration corresponding to a week is shown with solid blue lines. Dotted orange lines correspond to the distance threshold for interaction.

In a similar fashion, we can show with the dominance index (Fig. S4) the progressive clustering of individuals with time when advection is absent, and compare it to the steady state this time *with* advection. We see that even after a short period of time ($t = 100\tau$), the dominance index without advection is larger than with advection, and this spatial aggregation only grows with time in absence of turbulent advection.

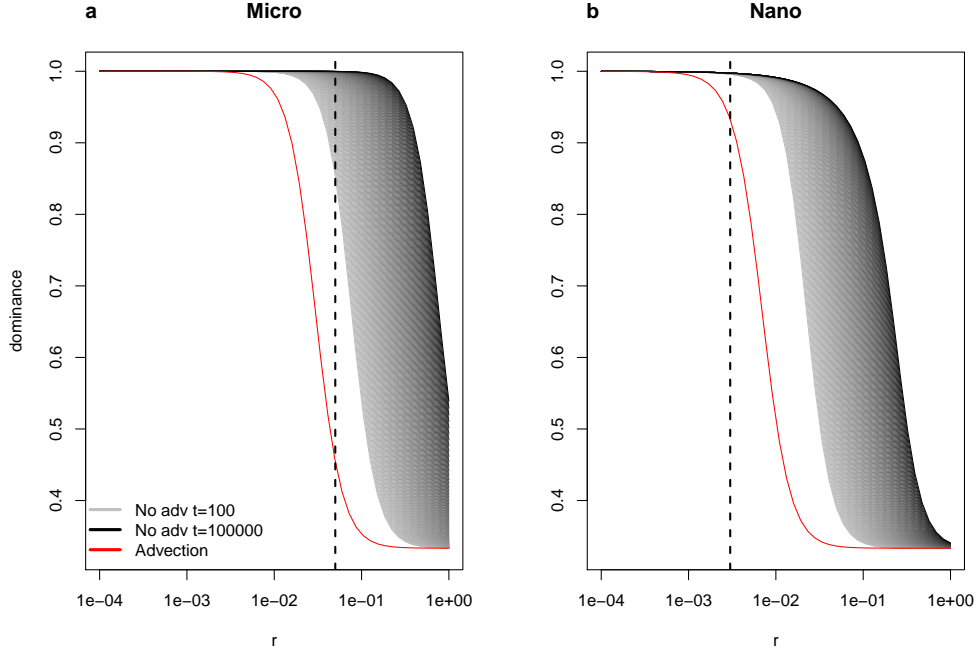


Figure S4: Theoretical dominance indices as a function of the distance (in cm) from a particle of a given species, for a microphytoplankton (a) and nanophytoplankton (b) 3-species community with an even abundance distribution, with (red line) and without (grey to black lines, with darker lines for longer simulations) advection. The vertical dashed line corresponds to the distance threshold for interaction.

S4 Computation of the pair correlation function and Ripley's K -function

The algorithm for pcf computation was adapted from the function `pcf3est` in `spatstat` 2.2-0 (Baddeley *et al.*, 2015) and slightly modified to compute the interspecific pcf (i.e., the pcf for marked point processes).

The pcf estimate $\hat{g}_{ij}(r)$ is computed via the use of the Epanechnikov kernel κ_E with bandwidth δ , i.e.

$$\hat{g}_{ij}(r) = \frac{1}{\hat{C}_i} \frac{1}{\hat{C}_j} \frac{1}{4\pi r^2} \sum_{k \in i} \sum_{l \in j} \kappa_E(r - \|\mathbf{x}_k - \mathbf{x}_l\|) w(\mathbf{x}_k, \mathbf{x}_l) \quad (\text{S11})$$

where $w(\mathbf{x}_k, \mathbf{x}_l)$ is the Ohser translation correction estimator (Ohser, 1983) and the kernel is defined as follow.

$$\kappa_E(x) = \begin{cases} \frac{3}{4\delta} \left(1 - \frac{x^2}{\delta^2}\right) & \text{for } -\delta \leq x \leq \delta \\ 0 & \text{otherwise.} \end{cases} \quad (\text{S12})$$

The estimate $\hat{g}_{ij}(r)$ is therefore very sensitive to the bandwidth: if it is too small, the estimate is noisy and may even be missing several pairs of points; if it is too large, the smoothing might be so important that values are strongly underestimated. In spatstat 2.2-0 (Baddeley *et al.*, 2015), the bandwidth default value is $\delta = 0.26C^{-1/3}$. The pcf computation function was first tested on standard distributions (with the default bandwidth), then on the Brownian Bug Model (with different bandwidths, see Fig. S7).

Estimates of the Ripley's K -function were also computed with the Ohser translation correction estimator but did not require any kernel smoothing. The same computation could be done using wrapped-around boundary conditions (for pcf estimation; for simulation we always consider periodic boundary conditions).

S4.1 Standard point processes

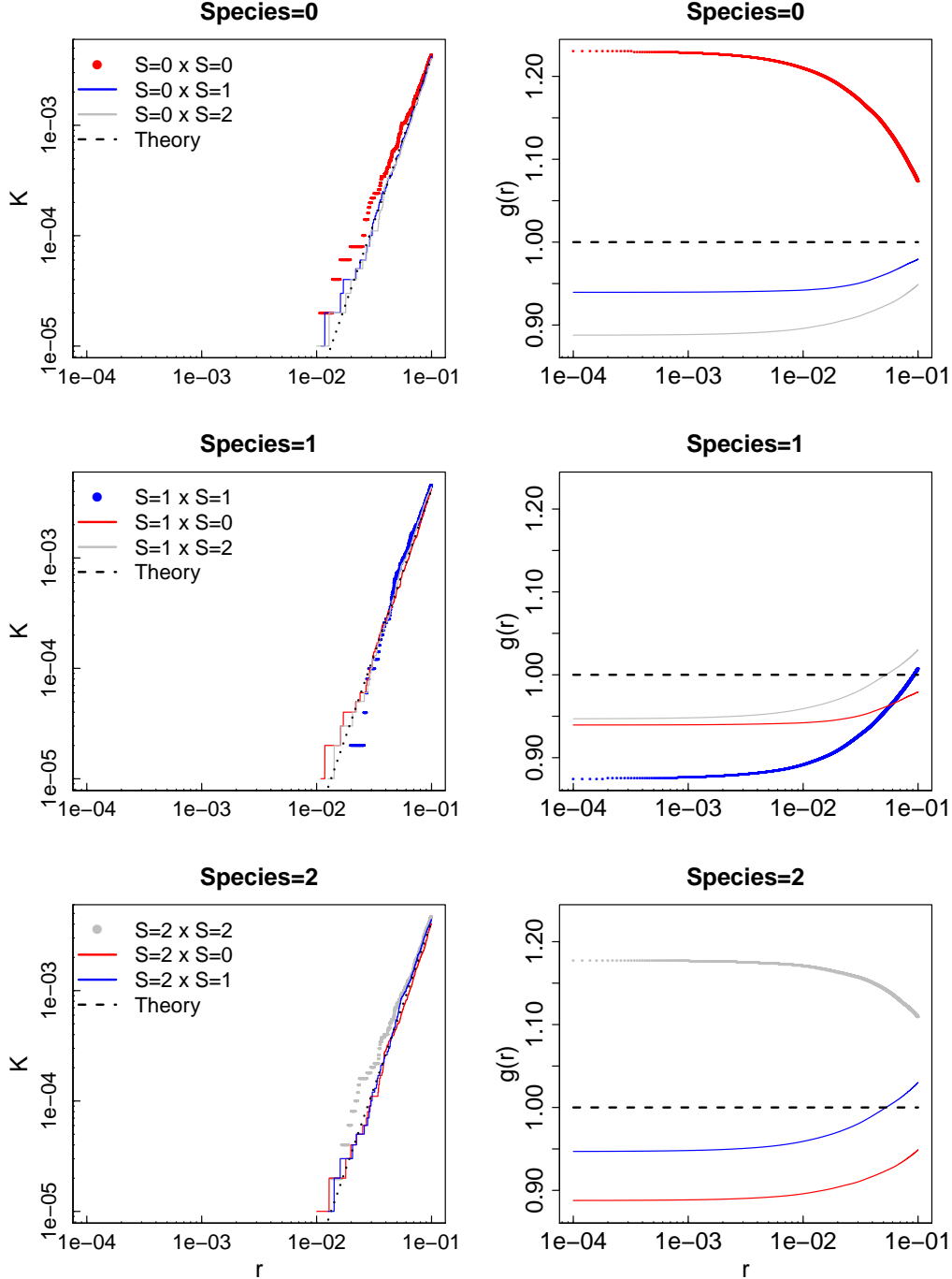


Figure S5: Intra- and inter-specific Ripley's K -function and pair correlation function values as a function of distance (in cm) for 3 species following a Poisson process with intensity 10 cm^{-3} , in a volume of 1000 cm^3 . Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formulas (dotted lines). Note that theoretical values are the same for intra and interspecific indices for the Poisson distribution. Colors correspond to the different species (red for species 0, blue for species 1, grey for species 2).

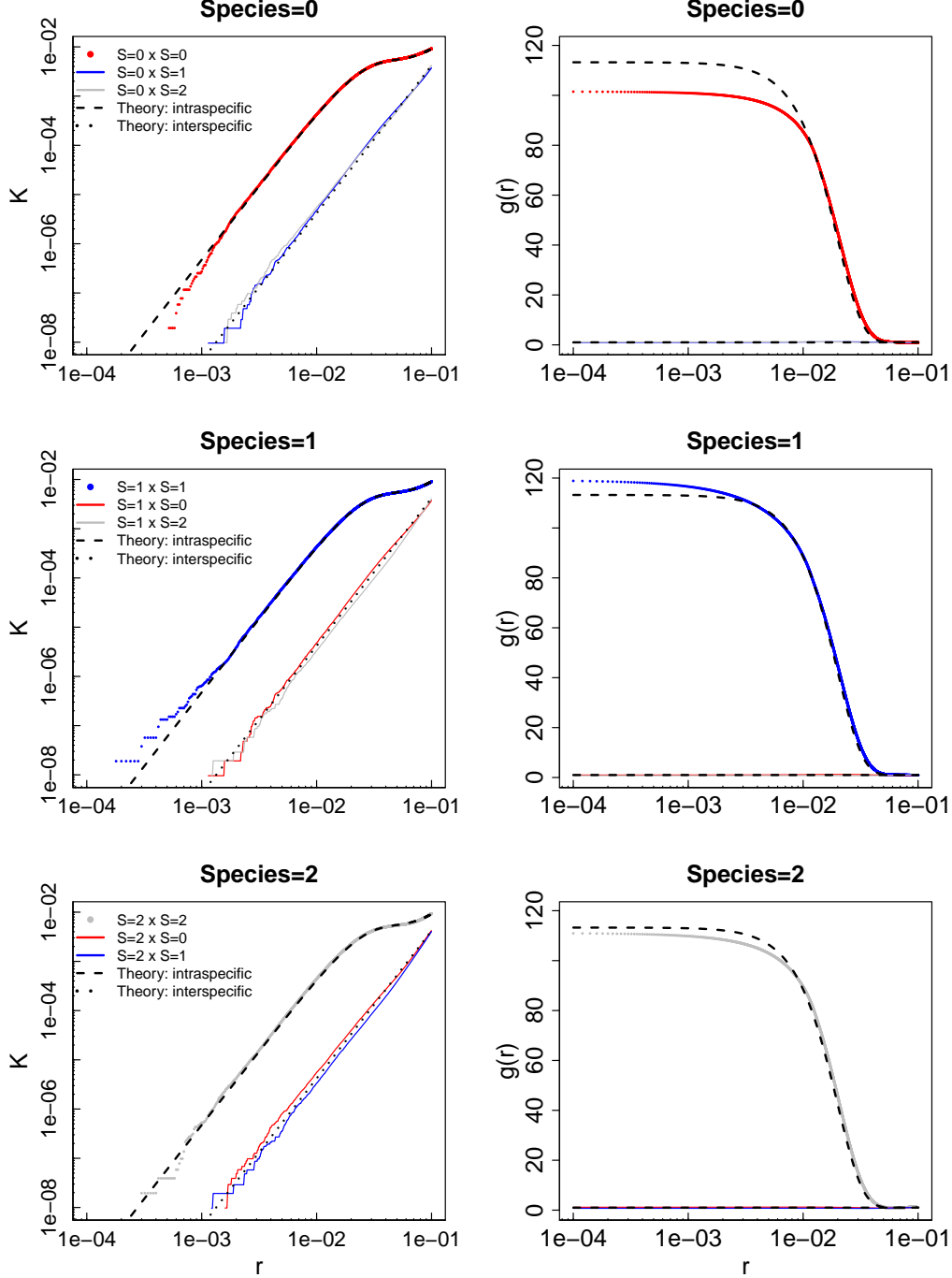


Figure S6: Intra- and inter-specific Ripley's K -function and pair correlation function values as a function of distance (in cm) for 3 species following a Thomas process with parent intensity $C_p = 200 \text{ cm}^{-3}$, number of children per parent $N_c = 50$, in a volume of 1 cm^3 , $\sigma = 0.01$ and $\delta \approx 0.012$. Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formulas (dashed and dotted lines for intra- and interspecific values, respectively). Colors correspond to the different species (red for species 0, blue for species 1, grey for species 2).

S4.2 Brownian Bug Model

While the pcf was one of the first indices that we intended to use, we quickly realized that the combination of the large range of distances we wanted to explore (from 10^{-4} to 1 cm) and the low density of individuals, at least for microphytoplankton, made the estimation difficult as the choice of the bandwidth was critical. We give an example of the sensitivity of the pcf computation to the bandwidth below (Fig. S7).

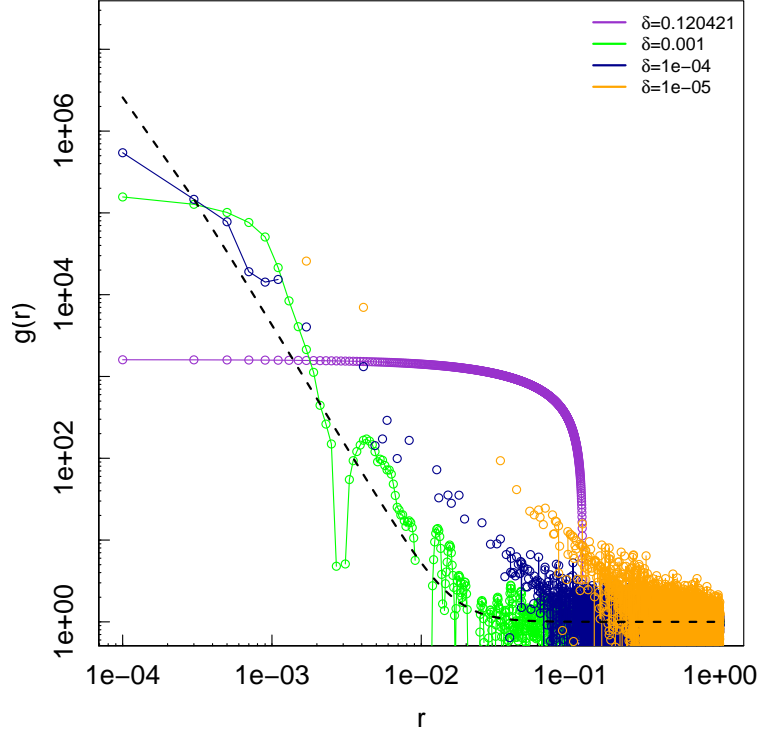


Figure S7: Intraspecific pair correlation function as a function of distance (in cm) computed for the Brownian bug model with microphytoplankton individuals, after 1000 time steps, with different values of the bandwidth δ . The dashed line indicates the theoretical pcf.

We decided, realizing that it would be very challenging to obtain a non-noisy pcf curve matching the theoretical expectation, to focus on Ripley's K -function whose cumulative nature helps the estimation process, which enabled us to compute the dominance index without having to calibrate a bandwidth beforehand.

S5 Spatial distributions

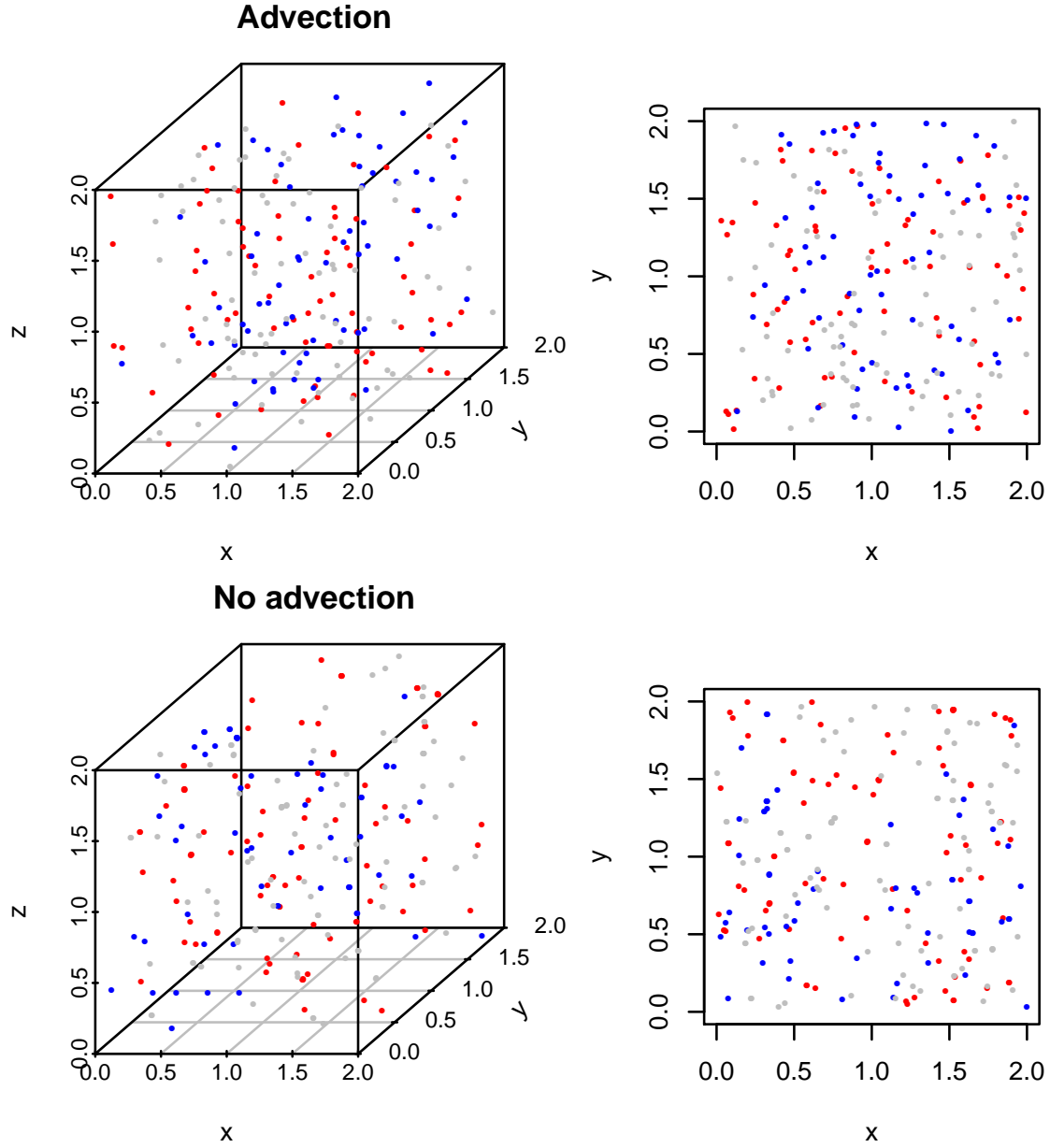


Figure S8: Spatial distributions of a 3-species community of microphytoplankton with and without advection with density $C = 10$ cells cm^{-3} after 1000 time steps. Each color corresponds to a different species. On the left-hand side, only a zoom on a $2 \times 2 \times 2$ cm^3 cube is shown, and its projection on the x-y plane is shown on the right-hand side.

S6 Sensitivity to the computation of the advection parameter

To compute the value of the maximum velocity of an organism in our model at the Kolmogorov scale, we used the formula $\text{Re} = U/k\nu \approx 1$ where k is the smallest wavenumber associated with turbulence. However, we could compute the Reynolds number with another, slightly different formula, using the equivalent sphere diameter (L_v) of our system (eq. S13). In this case, $\text{Re} = UL_v/\nu$ and

$$\begin{aligned} \frac{4}{3}\pi \left(\frac{L_v}{2}\right)^3 &= L_c^3 \\ \Leftrightarrow L_v &= 2L_c \left(\frac{3}{4\pi}\right)^{1/3} \\ \Leftrightarrow L_v &= 1.24 \text{ cm.} \end{aligned} \quad (\text{S13})$$

If we use $U \approx \nu/L_v$, $U \approx 8.1 \times 10^{-5} \text{ m s}^{-1}$. Using $U\tau/3 = 0.5 \text{ cm}$, we have $\tau = 185 \text{ s} = 2.1 \times 10^{-3} \text{ d}$. This means that $\gamma = 164 \text{ d}^{-1}$. As could be expected, when the flow velocity decreases, mixing decreases (Fig. S9).

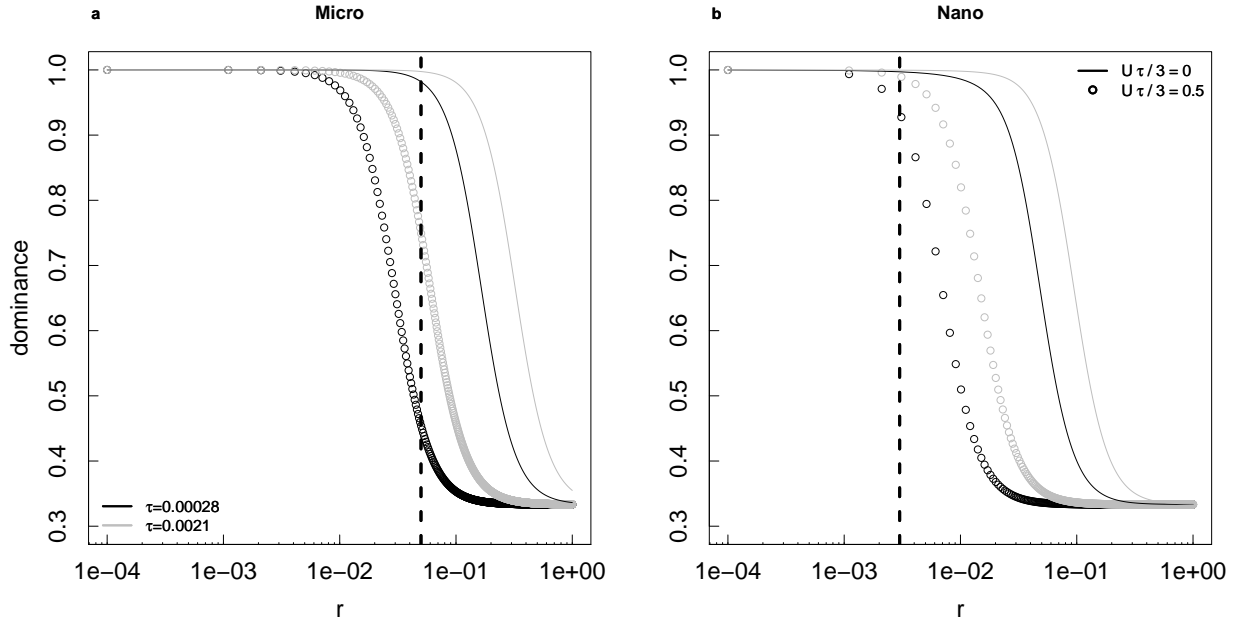


Figure S9: Dominance indices as a function of distance (in cm) for one species in a microphytoplankton (a) and nanophytoplankton (b) 3-species community with even distributions after 1000 timesteps with (circles) and without (lines) advection for different durations of the timesteps, with reference parameters (black) and lower flow velocity (grey).

S7 Minimum distances between individuals

Theory

One of the reasons why estimating $K(r)$, and even more so $g(r)$, is difficult is that for small distances (below 10^{-2}), we can find very few observations of pairs of points. In order to better understand at which distance ranges we should expect some estimation difficulties, we wanted to compute the minimum expected distance between points (distance to the nearest neighbour, DNN) when they are uniformly distributed.

In d dimensions, the probability distribution of the distance r to the nearest-neighbour follows $f(r) = db_d C r^{d-1} \exp(-b_d r^d C)$ where C is the intensity of the process. If we want to find the distribution of the minimum DNN between n realized points of a Poisson process with intensity C , we can write

$$\begin{aligned} \mathbb{P}(\min(R_1, \dots, R_n) > r) &= \mathbb{P}(R_1 > r, \dots, R_n > r) \\ &= \prod_i^n \mathbb{P}(R_i > r) \\ &= \prod_i^n \exp(-b_d r^d C) \\ &= \exp(-b_d r^d \sum_i^n C). \end{aligned} \tag{S14}$$

We can then conclude that the distribution of the minimum distance follows the same distribution as the DNN, but with intensity nC .

Clark & Evans (1979) show that a variable with probability distribution (with notations changed to fit our own) $f(r) = \frac{dC\pi^{d/2}r^{d-1}}{\Gamma(\frac{d}{2}+1)} \exp(-\frac{C\pi^{d/2}r^d}{\Gamma(\frac{d}{2}+1)}) = dCb_d r^{d-1} \exp(-Cb_d r^d)$ has an expected value of $\mu_d = \frac{(\Gamma(\frac{d}{2}+1))^{1/d} \Gamma(\frac{1}{d}+1)}{C^{1/d} \pi^{1/2}}$.

With intensity nC , we can write $\frac{(\Gamma(\frac{d}{2}+1))^{1/d} \Gamma(\frac{1}{d}+1)}{(nC)^{1/d} \pi^{1/2}}$.

In three dimensions,

$$\begin{aligned} \mu_d &= \frac{(nC)^{-1/3} (\Gamma(\frac{3}{2}+1))^{1/3} \Gamma(\frac{1}{3}+1)}{\pi^{1/2}} \\ &= (nC)^{-1/3} \left(\frac{3}{2}\Gamma(3/2)\right)^{1/3} \frac{1}{3}\Gamma(1/3) \frac{1}{\pi^{1/2}} \\ &\approx 0.554 \frac{1}{(nC)^{1/3}}. \end{aligned} \tag{S15}$$

This needs to be taken into account when defining C . For microphytoplankton, using $C = 10 \text{ cells cm}^{-3}$ and $n \approx 10^4$, the expected smallest NN distance for a uniform distribution is $1.2 \times 10^{-2} \text{ cm}$. For nanophytoplankton, using $C = 10^3 \text{ cells cm}^{-3}$ and $n \approx 10^4$, it is reduced to $2.6 \times 10^{-3} \text{ cm}$.

Simulations

We can compute the simulated distance to the nearest neighbour in the BBM and compare it to what we should obtain with a uniform spatial distribution: the simulated mean distance to the nearest organism, regardless of its species, is close to the expected value under a uniform distribution, but the minimum distance to a conspecific is much lower than expected (Fig. S10 for microphytoplankton, results are similar for nanophytoplankton).

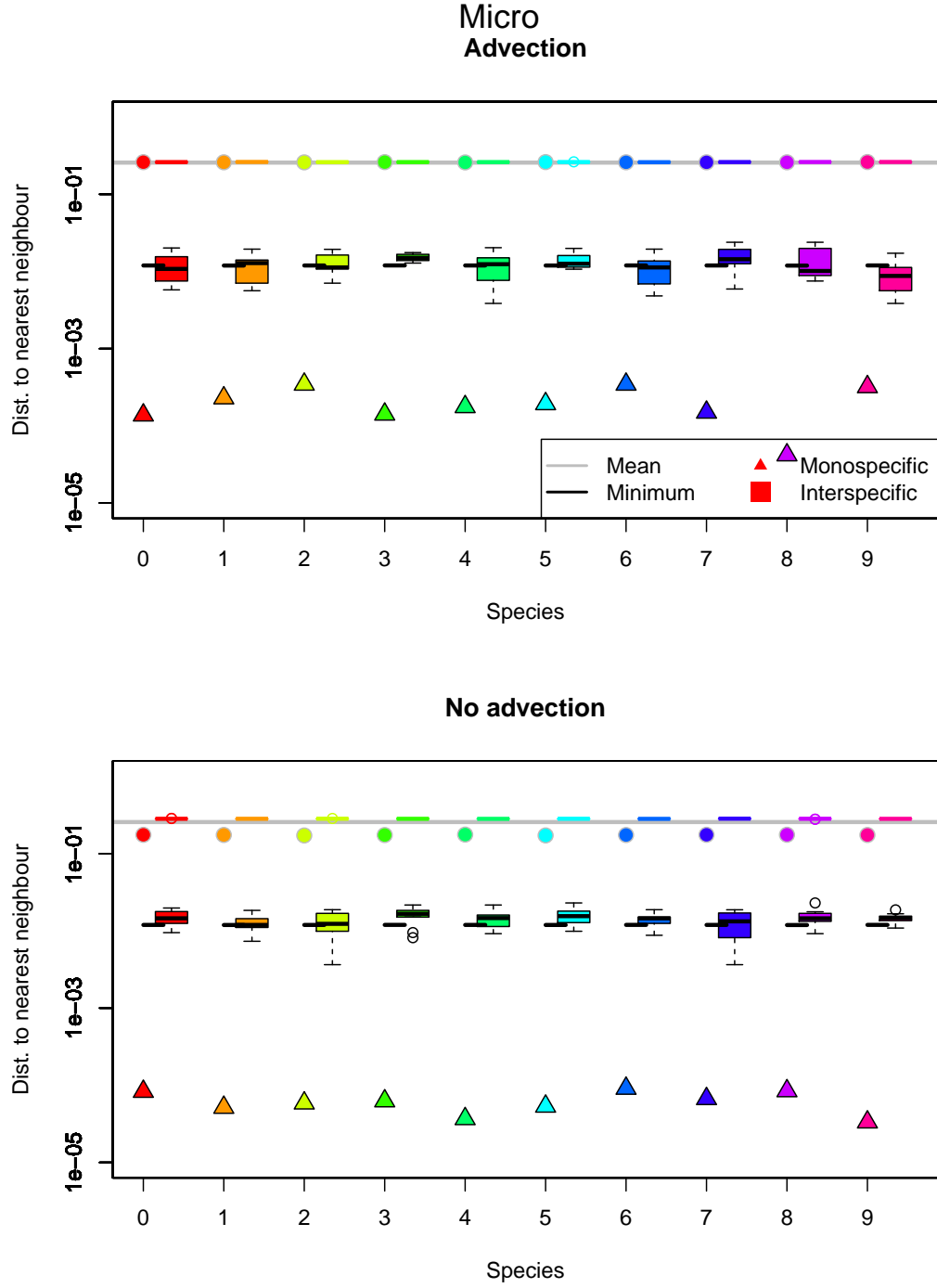


Figure S10: Mean and minimum distance (in cm) to the nearest neighbour for 10 microphytoplankton species with density $C = 10 \text{ cells cm}^{-3}$, with and without advection, after 1000 time steps, compared to predictions for a uniform distribution. Horizontal lines show the average distance to the nearest neighbour (grey line) and the expected minimum distance to the nearest neighbour with the actual number of realizations (black line). Circles and triangles represent mean and minimum distance to a conspecific, respectively. Boxplot corresponds to the distribution of mean (grey outlines) and minimum (black outlines) distances to a heterospecific. Colors correspond to different species.

In the case of a uniform distribution, an increase in density leads to a decrease in distance to the nearest neighbour (eq. S15). Mechanically, we can indeed expect that if the number of particles increases within the same volume, they likely get closer to each other. We confirm that this is also the case in the Brownian Bug Model.

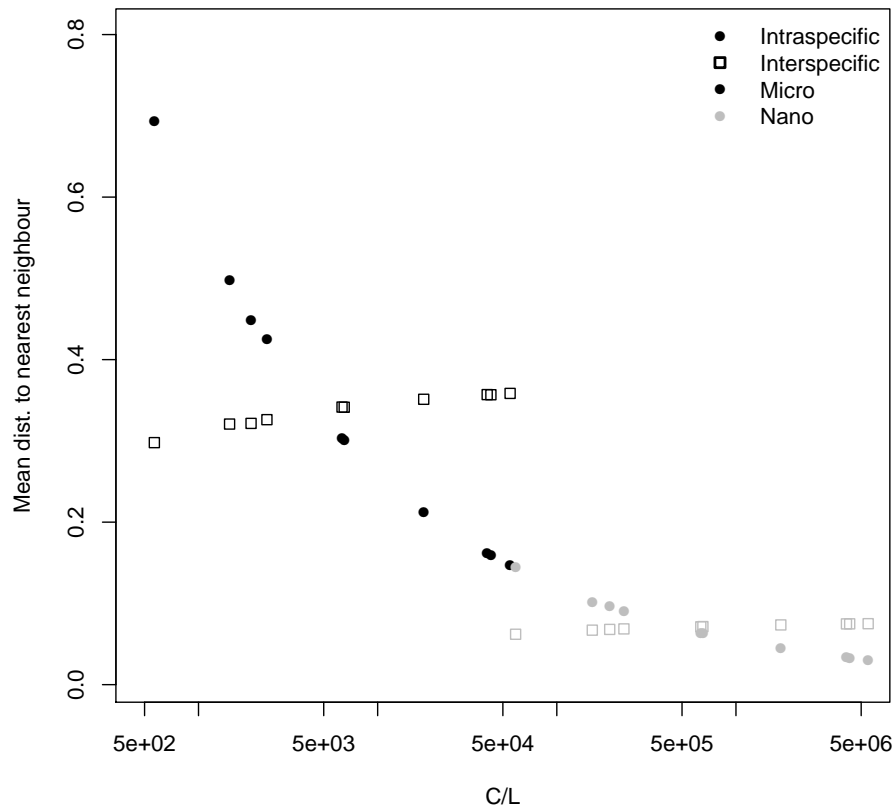


Figure S11: Mean distance (in cm) to the nearest conspecific (filled circle) or heterospecific (empty square) as a function of density in the environment for both microphytoplankton (black) and nanophytoplankton (grey) communities with a skewed abundance distribution, in the presence of advection.

S8 Relationship between the dominance index, relative strengths of interactions and coexistence in Lotka-Volterra models

In this section, we evaluate the potential relationship between local dominance, ratios of intra-to-interspecific interaction strengths observed at the population level, and their consequences in a spatial, dynamic point process Lotka-Volterra framework. Let us define μ_i the average growth rate of a typical individual of species i . Assuming it is linearly dependent on the abundances of the individual's conspecifics and heterospecifics within a neighbourhood of radius r ,

$$\begin{aligned}\mu_i(r, t) &= b_i + \beta_{ii}K_{ii}(r, t)C_i(t) + \beta_{io}\Sigma_{j \neq i}K_{ij}(r, t)C_j(t) \\ &= b_i + \beta_{ii}M_{ii}(r, t) + \beta_{io}M_{io}(r, t)\end{aligned}\tag{S16}$$

where b_i is the intrinsic individual growth rate, and β_{ii}/β_{io} are individual-level interaction coefficients with a conspecific / heterospecific, respectively. $C_j(t)K_{ij}(r, t)$ is the expected number of individuals of species j around a typical individual of species i within a sphere of radius r centered on the focal individual at time t . $M_{ii}(r, t) = K_{ii}(r, t)C_i(t)$ and $M_{io}(r, t) = \Sigma_{j \neq i}K_{ij}(r, t)C_j(t)$.

If we are close to an equilibrium at the local scale, and intra- and interspecific interaction strengths are equal *at the individual level* ($\beta_{ii} = \beta_{io} = \beta$), on average,

$$b_i + \beta M_{ii}(r, t) + \beta M_{io}(r, t) \approx 0.\tag{S17}$$

We can now focus on the dynamics at the community level. We denote α_{ij} the interactions at population level (by contrast to β_{ij} at individual level, as in Wiegand *et al.*, 2021). Assuming that all *interspecific* population-level interactions are similar to one another so that $\alpha_{ij} = \alpha_{io}$ if $j \neq i$, the per capita growth rate at population level can be written as

$$\frac{1}{C_i} \frac{dC_i(t)}{dt} = b_i + \alpha_{ii}C_i(t) + \alpha_{io}C_o(t) \approx 0.\tag{S18}$$

We can then write the approximate equalities $\alpha_{ii} \approx \beta \frac{M_{ii}(r, t)}{C_i(t)}$ and $\alpha_{io} \approx \beta \frac{M_{io}(r, t)}{C_o(t)}$ by matching Eqs. S16 and S18, and obtain the population-level interaction strength ratio

$$\frac{\alpha_{io}}{\alpha_{ii}} \approx \frac{M_{io}(r, t)}{M_{ii}(r, t)} \frac{C_i(t)}{C_o(t)}.\tag{S19}$$

Using the formula for the dominance index

$$\begin{aligned}\mathcal{D}_i(r, t) &= \frac{M_{ii}(r, t)}{M_{ii}(r, t) + M_{io}(r, t)} \\ \Leftrightarrow \frac{M_{io}(r, t)}{M_{ii}(r, t)} &= \frac{(1 - \mathcal{D}_i(r, t))}{\mathcal{D}_i(r, t)}.\end{aligned}\tag{S20}$$

Thus the population-level interaction strength ratio can be written out as a function of the dominance index and of the ratio of conspecific to heterospecific density:

$$\frac{\alpha_{io}}{\alpha_{ii}} \approx \frac{(1 - \mathcal{D}_i(r, t))}{\mathcal{D}_i(r, t)} \frac{C_i(t)}{C_o(t)}.\tag{S21}$$

Let us first focus on microphytoplankton in a 3-species community with an even distribution of abundances. We know that $\mathcal{D}(d_{\text{threshold}}) \approx 0.4$ at equilibrium. In this case, $\frac{(1 - \mathcal{D}_i(r, t))}{\mathcal{D}_i(r, t)} \frac{C_i(t)}{C_o(t)} = 0.75$. For nanophytoplankton, $\mathcal{D}(d_{\text{threshold}}) \approx 0.9$, thus $\frac{\alpha_{io}}{\alpha_{ii}} \approx 0.06$. Both ratios of population-level interaction strength are below 1, in spite of the $\beta_{ii} = \beta_{io} = \beta$ assumption, and therefore meet a necessary condition for diversity maintenance in a Lotka-Volterra model. Similar calculations for the 10-species communities, combining small dominance indices to low average concentrations, lead to $\frac{\alpha_{io}}{\alpha_{ii}} \ll 1$, compatible with coexistence (Barabás *et al.*, 2016).

References

- Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London.
- Barabás, G., Michalska-Smith, M.J. & Allesina, S. (2016). The effect of intra- and interspecific competition on coexistence in multispecies communities. *The American Naturalist*, 188, E1–E12.
- Clark, P.J. & Evans, F.C. (1979). Generalization of a nearest neighbor measure of dispersion for use in k dimensions. *Ecology*, 60, 316–317.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. vol. 70. John Wiley & Sons.
- Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14, 63–71.
- Wiegand, T., Wang, X., Anderson-Teixeira, K.J., Bourg, N.A., Cao, M., Ci, X., Davies, S.J., Hao, Z., Howe, R.W., Kress, W.J., Lian, J., Li, J., Lin, L., Lin, Y., Ma, K., McShea, W., Mi, X., Su, S.H., Sun, I.F., Wolf, A., Ye, W. & Huth, A. (2021). Consequences of spatial patterns for coexistence in species-rich plant communities. *Nature Ecology & Evolution*, 5, 965–973.