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:

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

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 (eq. S2, S3 and S4).

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

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

$$\begin{cases} x(t+\tau) = & x(t+\tau/3) \\ y(t+\tau) = & y(t+\tau/3) \\ z(t+\tau) = & y(t) + \frac{U\tau}{3} cos(kx(t+\tau)+\psi) \end{cases} \tag{S4}$$

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 familiarize the reader with spatial point process metrics, 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 standard 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 location 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.

S2.1 Pair correlation function

In the case of a Poisson point process,

$$\forall r \ge 0,\, g_{ii}(r) = 1 \tag{S5}$$

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

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

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

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

S2.2 Ripley's K-function

In the case of a Poisson point process,

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

For a Thomas point process,

$$K_{ii}(r) = \frac{4}{3}\pi r^3 + \frac{1}{C_p \sigma \sqrt{\pi}} \left(\sigma\sqrt{\pi}\,\mathrm{erf}\left(\frac{r}{2\sigma}\right) - r e^{-\left(\frac{r}{2\sigma}\right)^2} \right) \tag{S8}$$

More generally, 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:

$$D_i(r) = \frac{C_i}{\sum_{j=1}^{S} C_j} \tag{S9}$$

In the Thomas process, using eq. S8,

$$D_i(r) = \frac{C_i \left(\frac{4}{3}\pi r^3 + \frac{F(r)}{C_{p,i}} \right)}{C_i \frac{F(r)}{C_{p,i}} + \sum_j C_j \frac{4}{3}\pi r^3} \tag{S10}$$

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

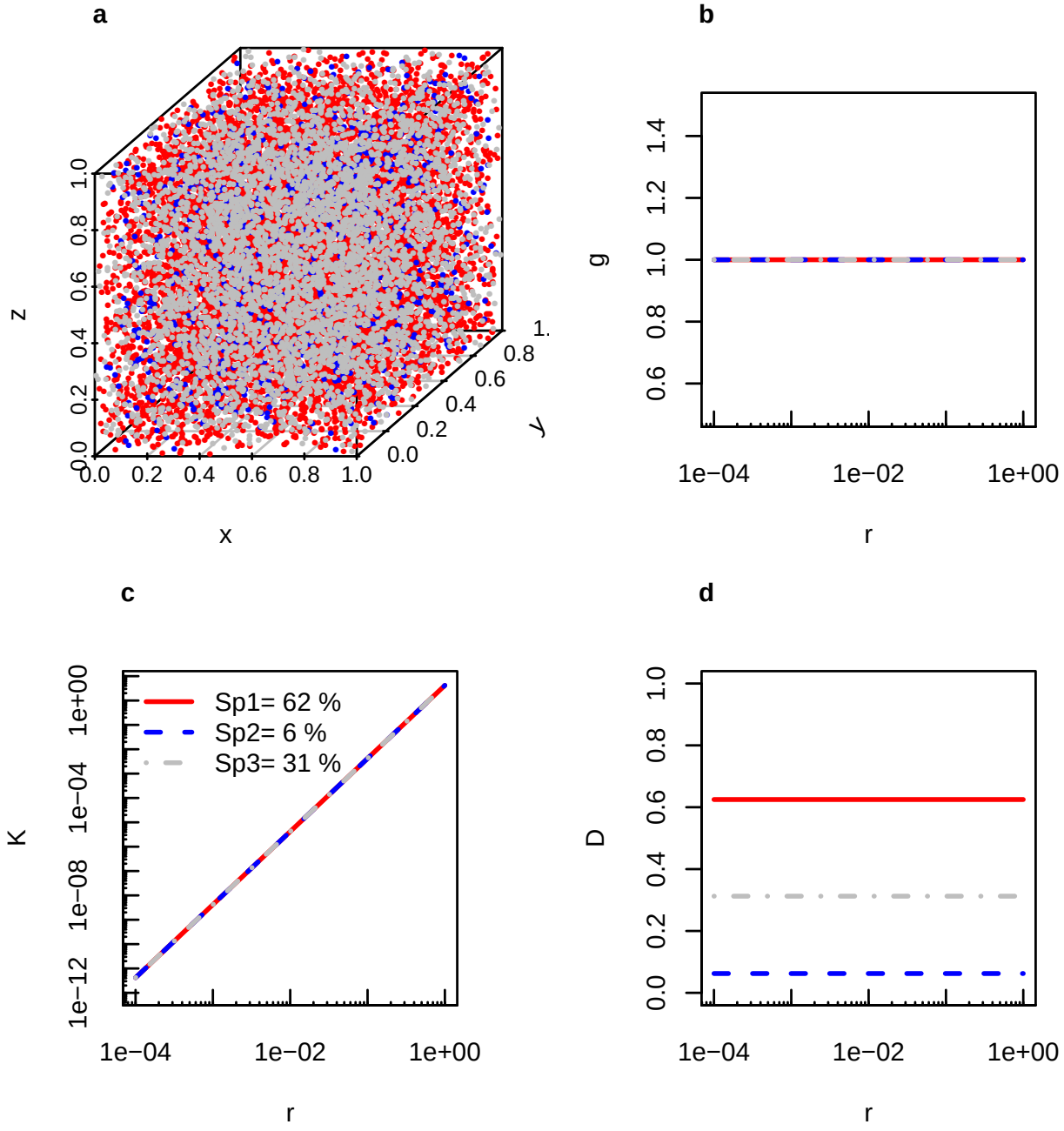

Figure S1: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K-function (c) and dominance index (d) for a Poisson point process in a 3-species community with different intensities (10000 cm^{-3}, 1000 cm^{-3}, 5000 cm^{-3}; proportions in the community are given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

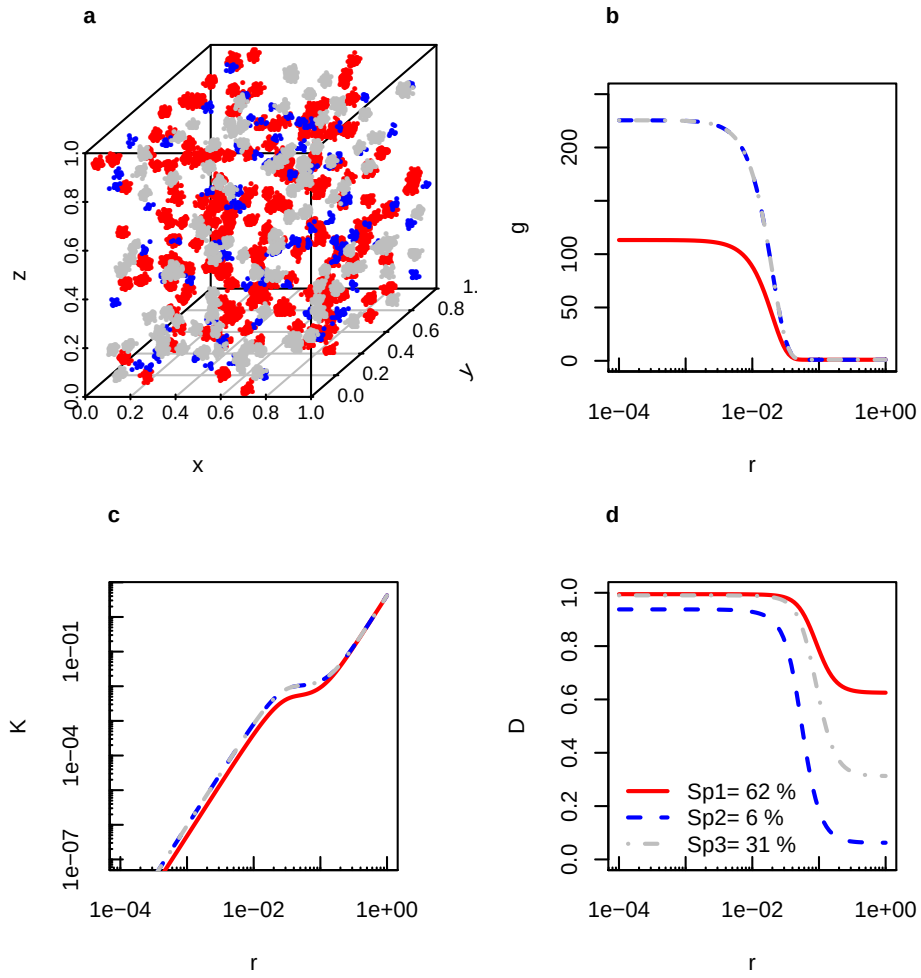

Figure S2: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K-function (c) and dominance index (d) for a Thomas point process in 3-species community with different parent intensities (200 cm^{-3}, 100 cm^{-3}, 100 cm^{-3}), and different children per parent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

S3 Theoretical behaviour of the dominance index in the BBM

With the theoretical formula for the dominance index in the Brownian Bug Model, we can show the progressive clustering of particles with time when advection is absent. Even after a short period of time, the dominance index without advection is larger than with advection.

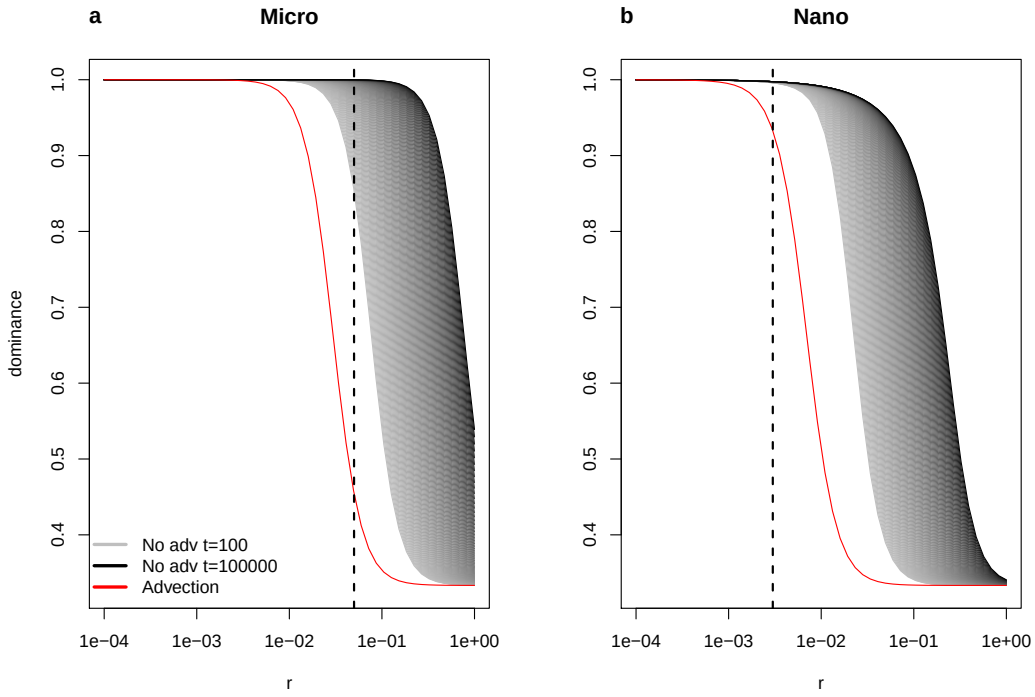

Figure S3: Theoretical dominance indices as a function of the distance (in cm) from a particle of a given species, for a microphytoplankton (a) and nanophytoplankton (b) 3-species community with an even abundance distribution, with (red line) and without (grey to black lines, with darker lines for longer simulations) advection. The vertical dashed line corresponds to the distance threshold for interaction.

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

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

Schematically, $g_{ij}(r)$ is estimated via the use of the Epanechnikov kernel κ_E with bandwidth δ.

$$\hat{g}_{ij}(r) \quad = \quad \frac{1}{\hat{C}_i}\frac{1}{\hat{C}_j}\frac{1}{4\pi r^2}\sum_i\sum_j \kappa_E(r-||x_i-x_j||)w(x_i,x_j) \tag{S11}$$

where $w(x_i, x_j)$ is the Ohser translation correction estimator (Ohser, 1983) and the kernel is defined as follow.

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

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 below).

Estimates of the Ripley's K-function were also performed with the Ohser translation correction estimator but did not require any kernel smoothing.

S4.1 Standard point processes

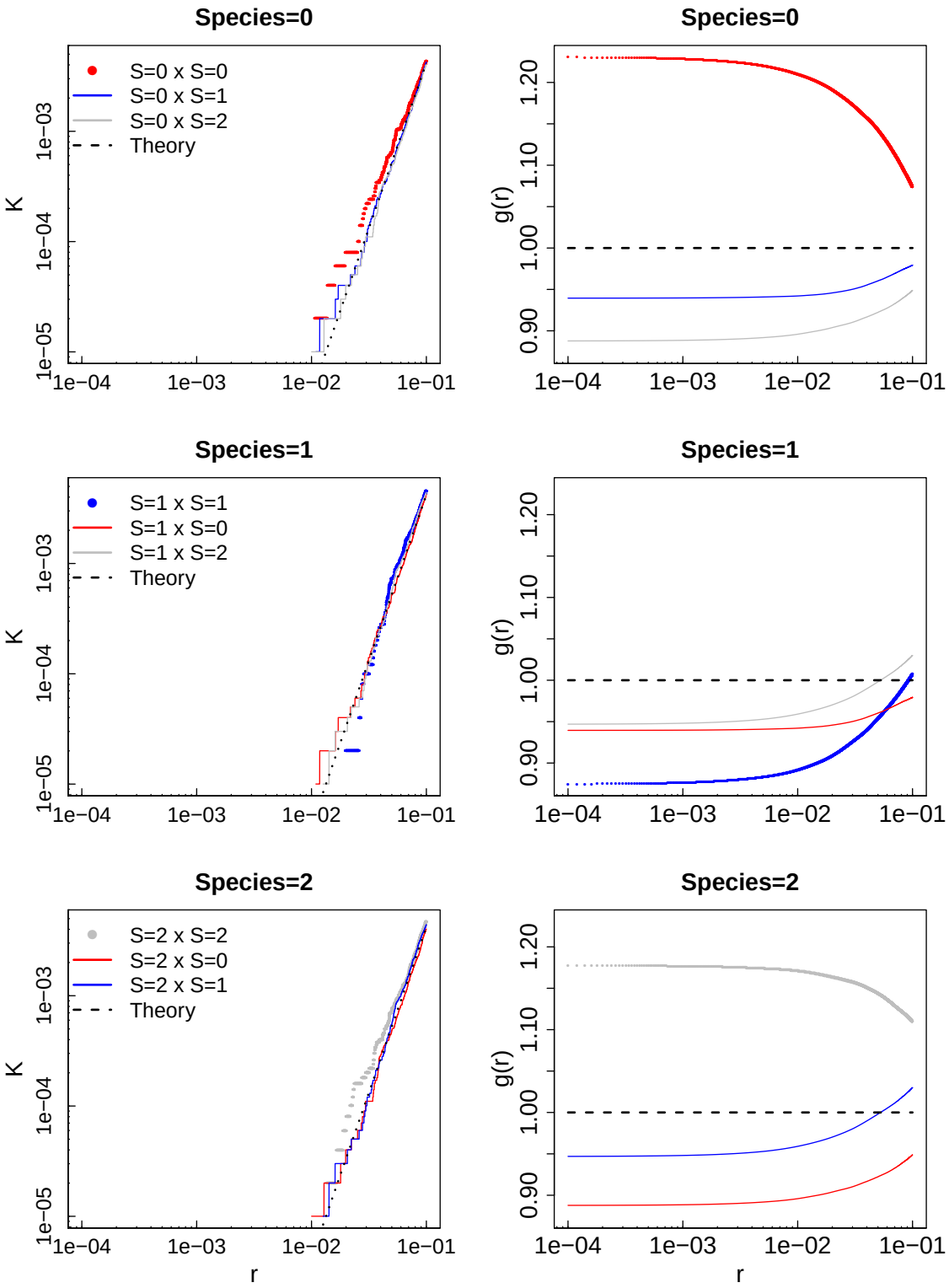

Figure S4: Intra- and inter-specific Ripley's K-function and pair correlation function values as a function of distance (in cm) for 3 species following a Poisson process with intensity 10 cm^{-3}, in a volume of 1000 cm^3. Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formula (dotted lines). Note that theoretical values are the same for intra and interspecific moments for the Poisson distribution. Colors correspond to the different species (red for species 0, blue for species 1, grey for species 2).

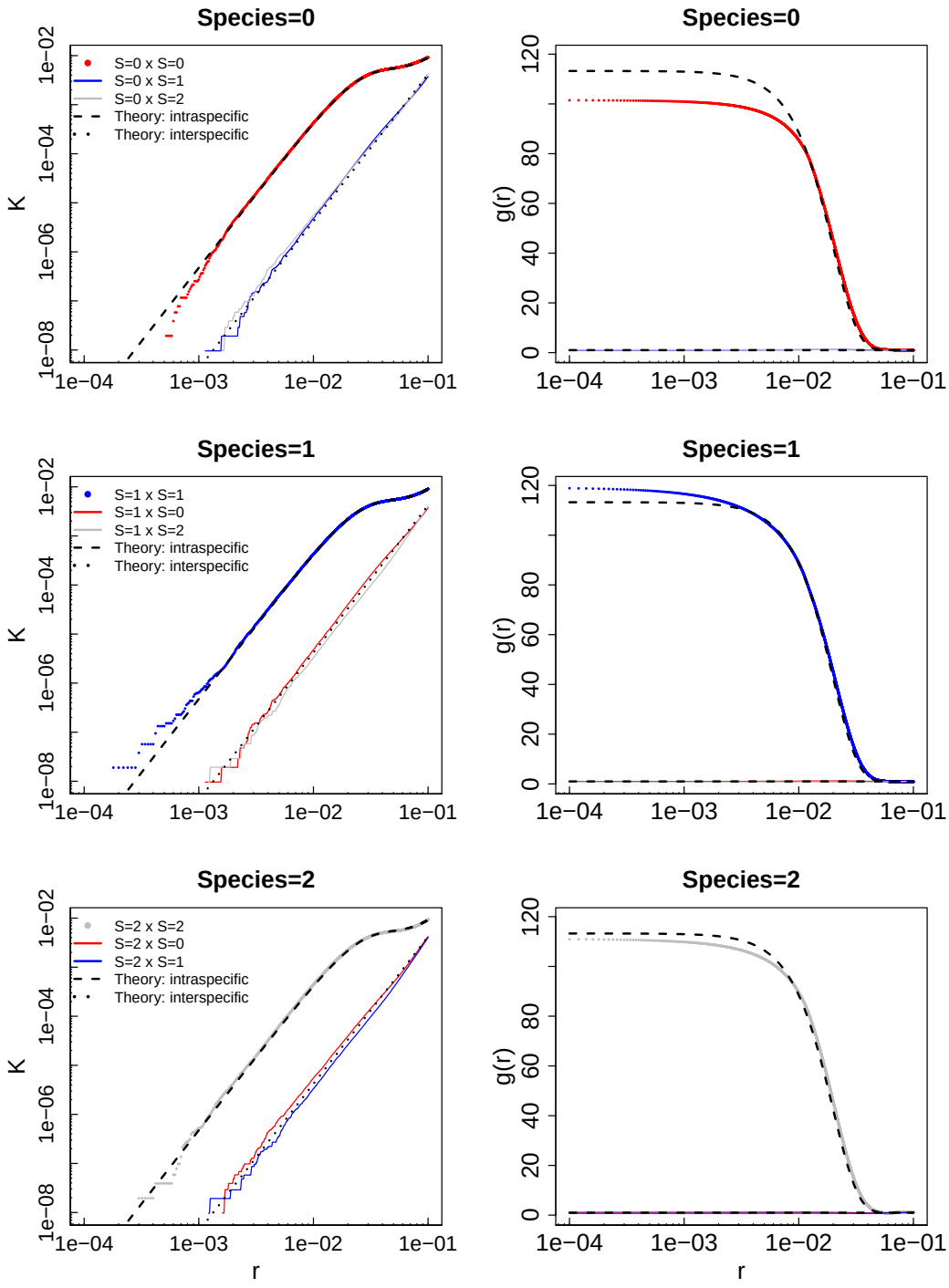

Figure S5: Intra- and inter-specific Ripley's K-function and pair correlation function values as a function of distance (in cm) for 3 species following a Thomas process with parent intensity $C_p = 200$ cm^{-3}, number of children per parent $N_c = 50$, in a volume of 1 cm^3, $\sigma = 0.01$ and $\delta \approx 0.012$. Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formula (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 particles, 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. S6).

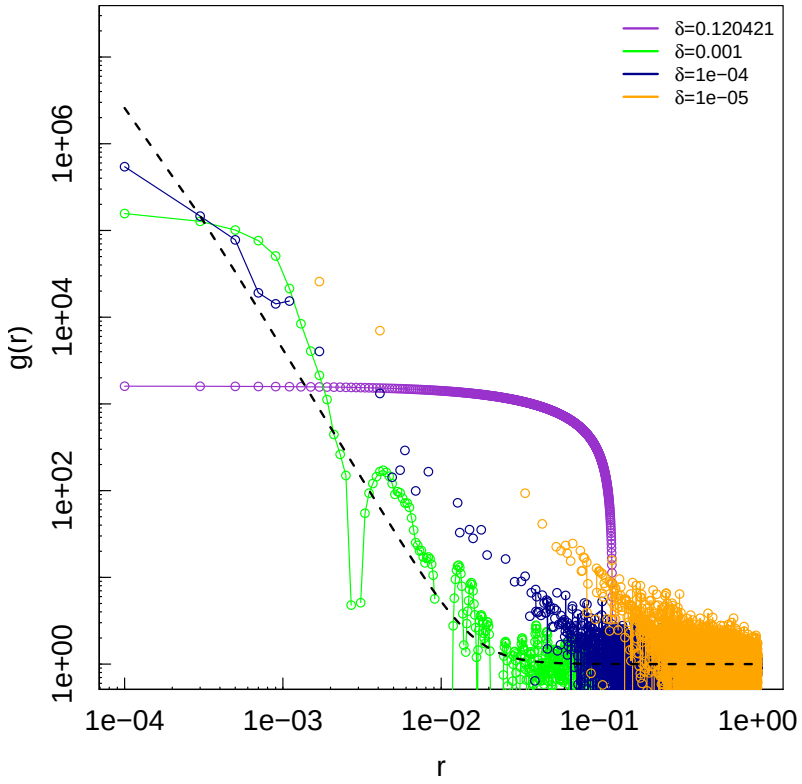

Figure S6: Intraspecific pair correlation function as a function of distance (in cm) computed for the Brownian bug model with microphytoplankton particles, after 1000 time steps, with different values of the bandwidth δ. The dashed line indicates the theoretical pcf.

We decided, from these results, to focus on Ripley's K-function, which enabled us to compute the dominance index without having to calibrate a bandwidth beforehand.

S5 Spatial distributions

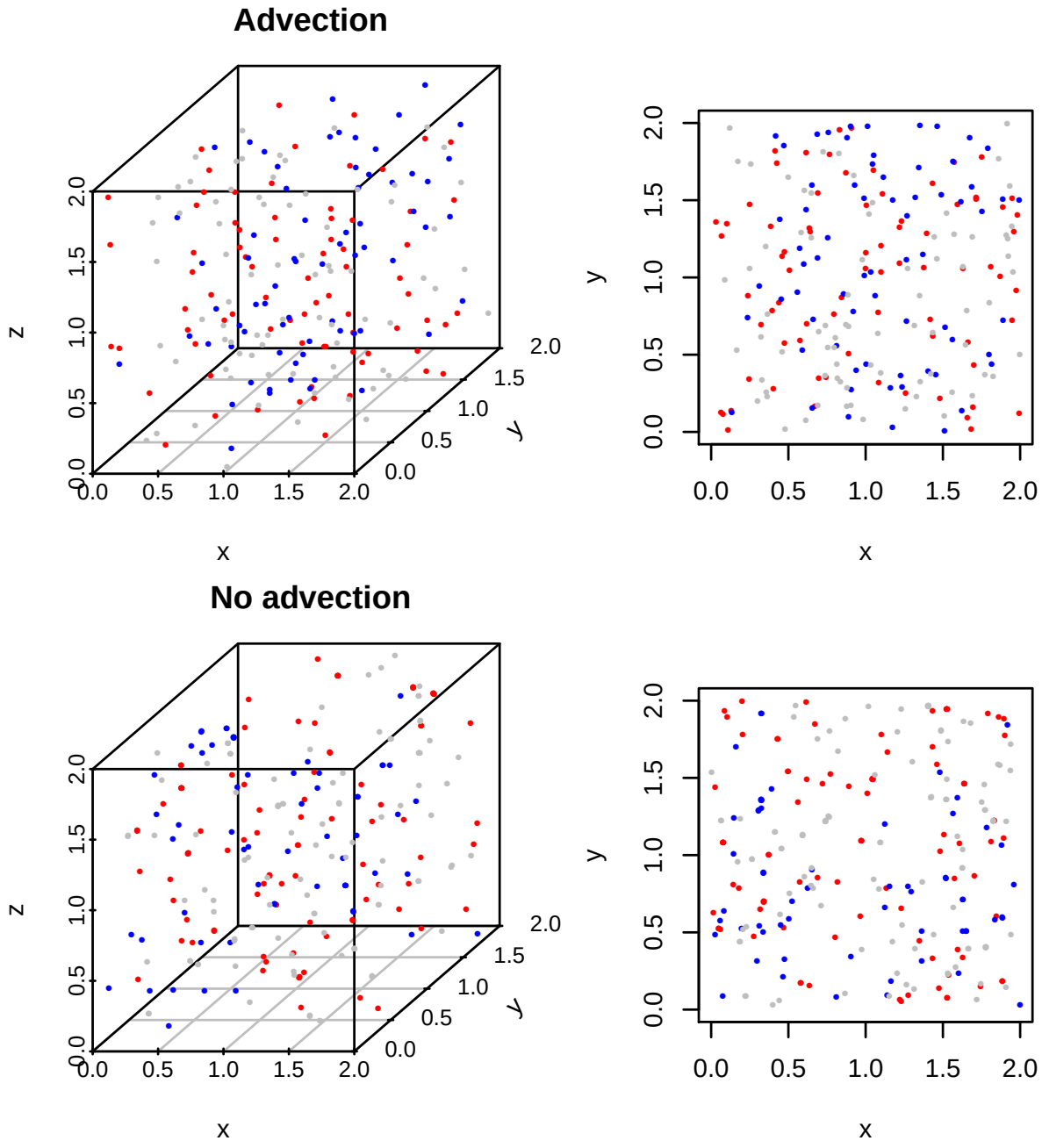

Figure S7: Spatial distributions of a 3-species community of microphytoplankton with and without advection with density $C = 10$ cm^{-3} after 1000 time steps. Each color corresponds to a different species. On the left-hand side, only a zoom on a $2 \times 2 \times 2$ cm^3 cube is shown, and its projection on the x-y plane is shown on the right hand-side.

S6 Minimum distances between points

Theory

One of the reasons why estimating K and $g(r)$ is difficult is that for small distances (below 10^{-2}), we can find very few observations of pairs of points. As a first proxy, we want to estimate 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:

$$\begin{aligned} \mathbb{P}(\min(R_1, ..., R_n) > r) &= \mathbb{P}(R_1 > r, ..., R_n > r) \\ &= \Pi_i^n \mathbb{P}(R_i > r) \\ &= \Pi_i^n \exp(-b_d r^d C) \\ &= \exp(-b_d r^d \Sigma_i^n C) \end{aligned} \tag{S13}$$

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 $\mu_d = \frac{\left(\Gamma(\frac{d}{2}+1)\right)^{1/d}\Gamma(\frac{1}{d}+1)}{C^{1/d}\pi^{1/2}}$.

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

In 3D,

$$\begin{aligned} \mu_d &= (nC)^{-1/3} \frac{\left(\Gamma(\frac{3}{2}+1)\right)^{1/3}\Gamma(\frac{1}{3}+1)}{\pi^{1/2}} \\ &= (nC)^{-1/3} \left(\tfrac{3}{2}\Gamma(3/2)\right)^{1/3} \tfrac{1}{3}\Gamma(1/3)\frac{1}{\pi^{1/2}} \\ &\approx 0.554 \frac{1}{(nC)^{1/3}} \end{aligned} \tag{S14}$$

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

Simulations

We can compute the simulated distance to the nearest neighbour and compare it to what we should obtain with a uniform distribution (Fig. S8 and S9): the simulated average DNN are close to the expected value for a uniform distribution, but the minimum distance to a conspecific is much lower than expected.

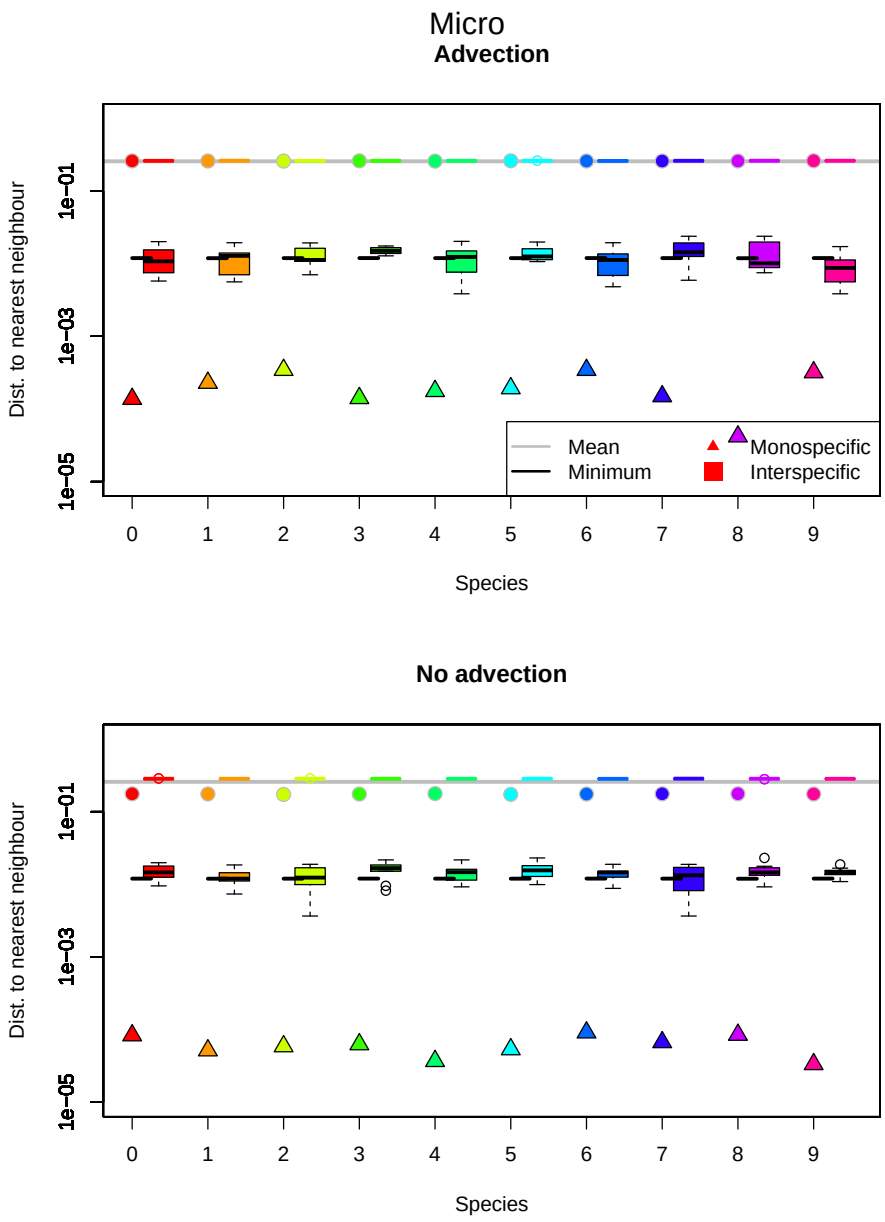

Figure S8: Mean and minimum distance (in cm) to the nearest neighbour for 10 microphytoplankton species with density $C = 10$ cm^{-3}, with and without advection, after 1000 time steps, compared to predictions for a uniform distribution. Horizontal lines show the average distance to the nearest neighbour (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.

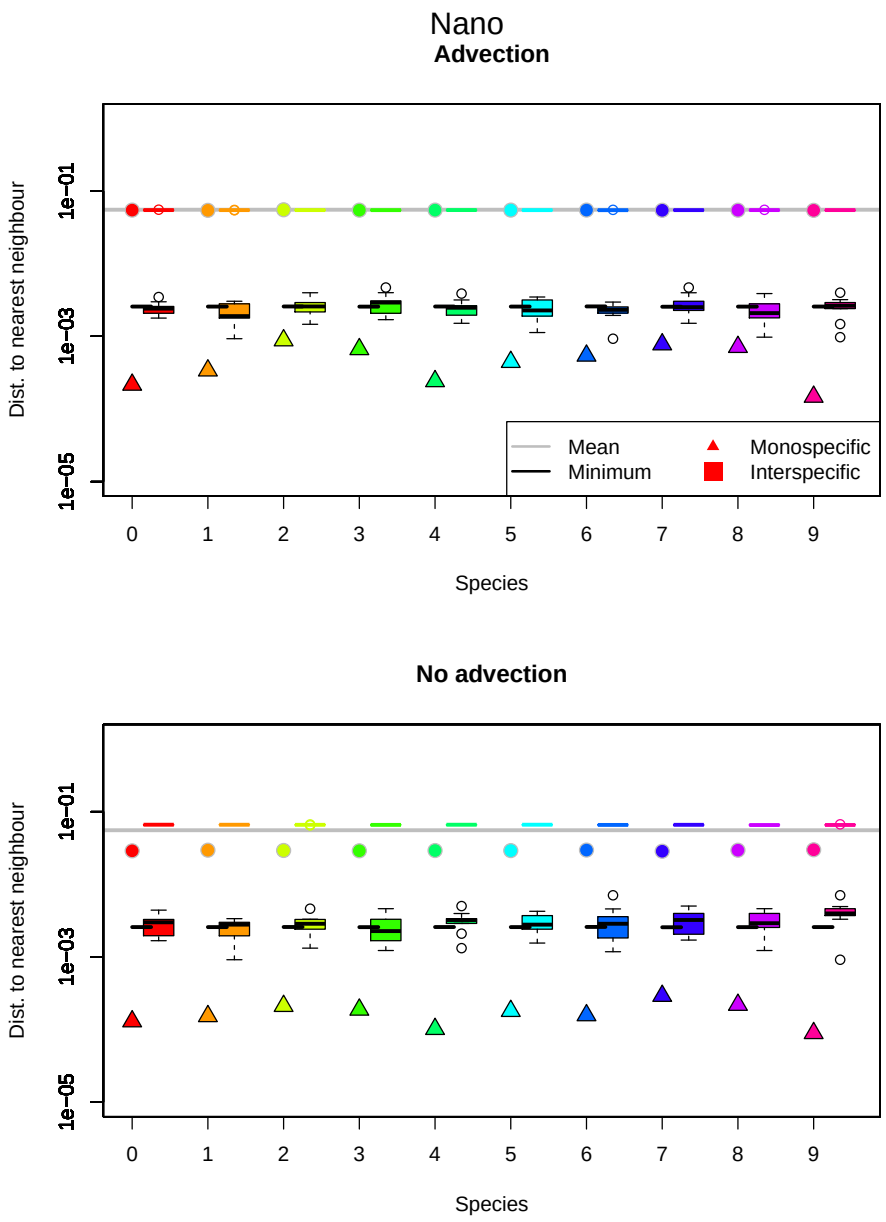

Figure S9: Mean and minimum distance (in cm) to the nearest neighbour for 10 nanophytoplankton species with density $C = 10^3$ cm^{-3}, with and without advection, after 1000 time steps, compared to predictions for a uniform distribution. Horizontal lines show the average distance to the nearest neighbour (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.

Relationship with densities

In the case of a uniform distribution, an increase in density leads to a decrease in distance to the nearest neighbour (eq. S14). 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.

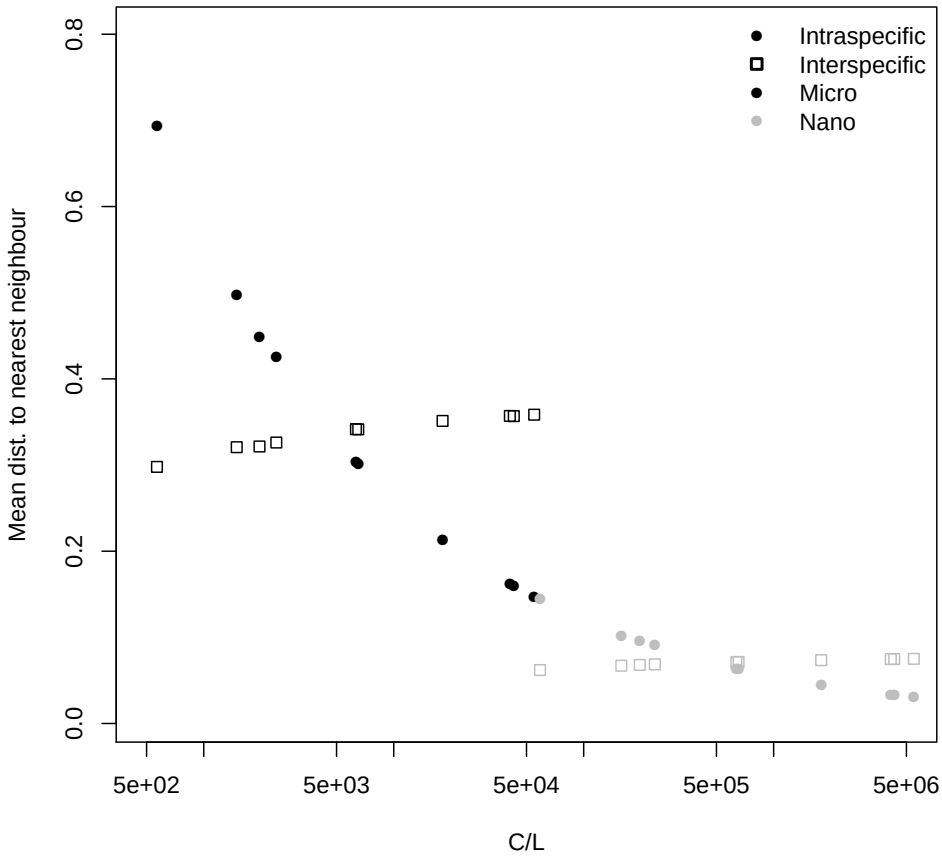

Figure S10: Mean distance (in cm) to the nearest conspecific (filled circle) or heterospecific (empty square) as a function of density in the environment for both microphytoplankton (black) and nanophytoplankton (grey) communities with a skewed abundance distribution, in the presence of advection.

S7 Effect of the initial distribution

Particles are uniformly distributed in the cube at the start of all simulations shown in the manuscript. However, we have no reason to believe that such spatial distribution is more appropriate than a more clustered one to begin with. In Fig. S11, we show the final dominances obtained with and without advection, starting from a superposition of Thomas processes, i.e. each species was distributed with its own Thomas process.

We need to modify the pair density function for the case without advection, as the initial distribution changes from $G(r,0) = C^2$ to $G(r,0) = C^2 + \frac{C^2}{C_p}\frac{1}{(4\pi\sigma^2)^{3/2}}e^{-\left(\frac{r^2}{4\sigma^2}\right)}$, leading to eq. S15.

$$K(r,t) = \frac{\lambda}{CD}\left(\frac{\rho^2}{2} - \frac{1}{2}\,\mathrm{erf}(\frac{\rho}{\sqrt{8Dt}})(\rho^2 - 4Dt) - \frac{\sqrt{2Dt}\rho}{\sqrt{\pi}}e^{-\rho^2/8Dt}\right) + \frac{4}{3}\pi r^3 + \frac{1}{C_p\sigma\sqrt{\pi}}\left(\sigma\sqrt{\pi}\,\mathrm{erf}\left(\frac{r}{2\sigma}\right) - re^{-\left(\frac{r}{2\sigma}\right)^2}\right) \tag{S15}$$

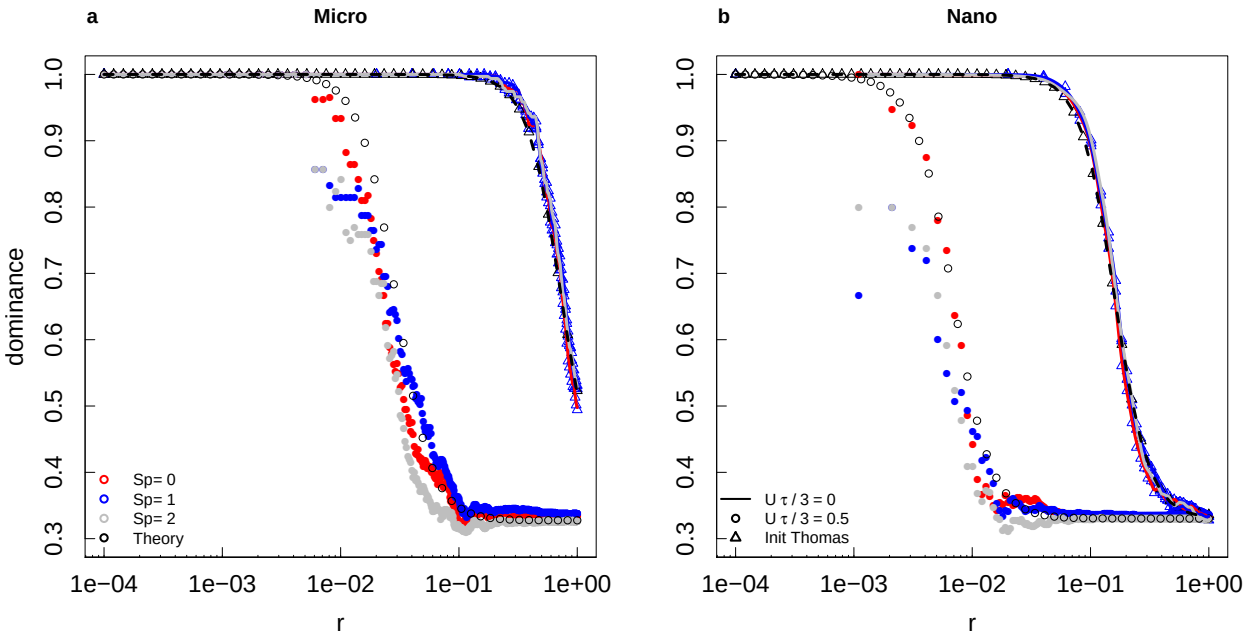

Figure S11: Dominance indices as a function of distance (in cm) for microphytoplankton (a) and nanophytoplankton (b) in a 3-species community with even distributions after 1000 timesteps starting from a superposition of species-specific Thomas point processes, with (circles) and without (lines) advection. The triangles correspond to the dominance index at the start of the simulation, i.e. when points are only distributed according to a Thomas point process with parent intensity $C_p = 0.2$ cm^{-3} and $C_p = 20$ cm^{-3} for microphytoplankton and nanophytoplankton respectively, number of children per parent $N_c = 50$, and $\sigma = 0.001$. Each color represents a different species. The black points and line corresponds to the theoretical values of the dominance index.

Simulated and theoretical values match for a 1000-time steps duration (Fig. S12).

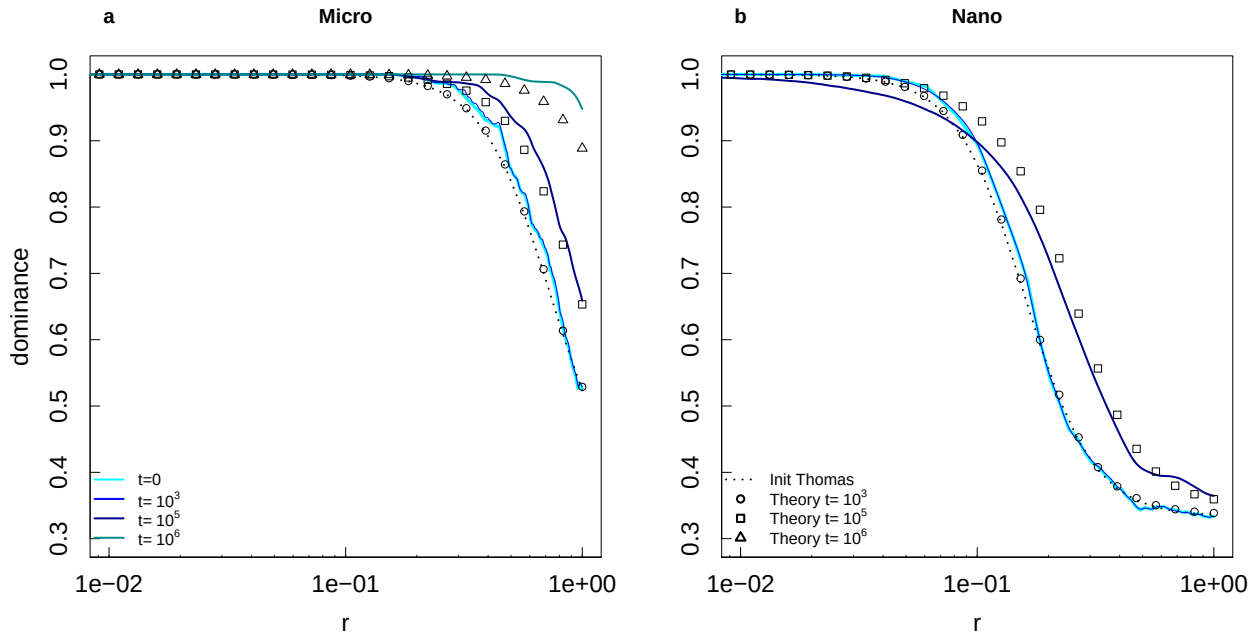

Figure S12: Dominance indices for large distances (in cm) for one species in a microphytoplankton (a) and nanophytoplankton (b) 3-species community with even distributions after different simulation durations starting from a superposition of species-specific Thomas point processes in the absence of advection. The dotted line corresponds to the dominance index at the start of the simulation, i.e. when points are only distributed according to a Thomas point process with parent intensity $C_p = 0.2$ cm^{-3} and $C_p = 20$ cm^{-3} for microphytoplankton and nanophytoplankton respectively, number of children per parent $N_c = 50$, and $\sigma = 0.001$. Each colored line represents a different simulation duration, and black symbols correspond to the theoretical values of the dominance index.

S8 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 $Re = U/k\nu \approx 1$ 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. S16). In this case, $Re = UL_v/\nu$.

$$\begin{array}{rcl} \frac{4}{3}\pi\left(\frac{L_v}{2}\right)^3 & = & L_c^3 \\ \Leftrightarrow \quad L_v & = & 2L_c\left(\frac{3}{4\pi}\right)^{1/3} \\ \Leftrightarrow \quad L_v & = & 1.24 \text{ cm} \end{array} \tag{S16}$$

If we use $U \approx \nu/L_v$, $U \approx 8.1 \times 10^{-5}$ m.s^{-1}. Using $U\tau/3 = 0.5$ cm, we have $\tau = 185$ s $= 2.1 \times 10^{-3}$ d. This means that $\gamma = 164$. As could be expected, when the flow velocity decreases, mixing decreases (Fig. S13).

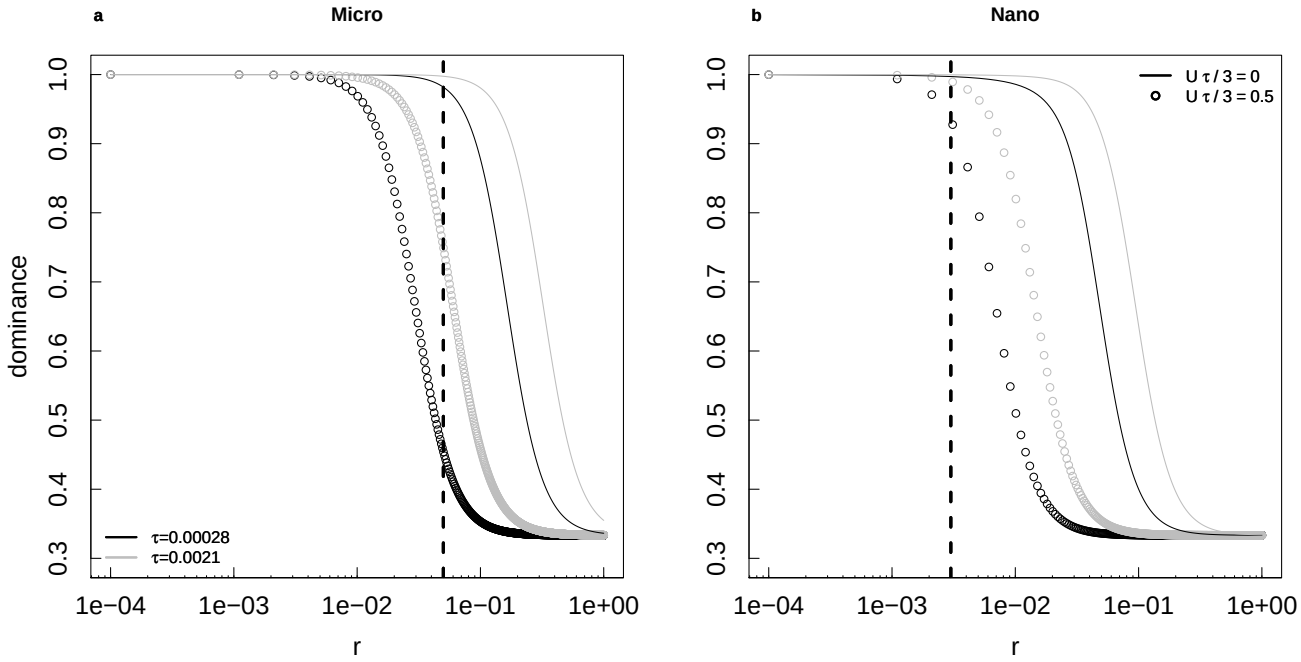

Figure S13: Dominance indices as a function of distance (in cm) for one species in a microphytoplankton (a) and nanophytoplankton (b) 3-species community with even distributions after 1000 timesteps with (circles) and without (lines) advection for different duration of the timesteps, with reference parameters (black) and lower flow velocity (grey).

References

Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London.

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.