

# Local intraspecific aggregation in phytoplankton model communities: spatial scales of occurrence and implications for coexistence

Coralie Picoche<sup>1,\*</sup>, William R. Young<sup>2</sup>, Frédéric Barraquand<sup>1,\*</sup>

<sup>1</sup>Institute of Mathematics of Bordeaux, University of Bordeaux and CNRS, Talence, France

<sup>2</sup>Scripps Institution of Oceanography, La Jolla, California, USA

## Abstract

The coexistence of multiple phytoplankton species despite their reliance on similar resources is often explained with mean-field models assuming mixed populations. In reality, observations of phytoplankton indicate spatial aggregation at all scales, including at the scale of a few individuals. Local spatial aggregation can hinder competitive exclusion since individuals then interact mostly with other individuals of their own species, rather than competitors from different species. To evaluate how microscale spatial aggregation might explain phytoplankton diversity maintenance, an individual-based, multispecies representation of cells in a hydrodynamic environment is required. We formulate a three-dimensional and multispecies individual-based model of phytoplankton population dynamics at the Kolmogorov scale. The model is studied through both simulations and the derivation of spatial moment equations, in connection with point process theory. The spatial moment equations show a good match between theory and simulations. We parameterized the model based on phytoplankters' ecological and physical characteristics, for both large and small phytoplankton. Defining a zone of potential interactions as the overlap between nutrient depletion volumes, we show that local species composition—within the range of possible interactions—depends on the size class of phytoplankton. In large phytoplankton, individuals are surrounded by cells from other species, while in small phytoplankton, individuals remain in mostly monospecific clusters. Spatial structure therefore favours intra- over inter-specific interactions for small phytoplankton, which likely contributes to coexistence mechanisms. Other factors behind diversity maintenance must be examined for large phytoplankton.

**Keywords:** aggregation; coexistence; individual-based model; phytoplankton; spatial moment equations; spatial point process

\*corresponding authors: [coralie.picoche@u-bordeaux.fr](mailto:coralie.picoche@u-bordeaux.fr) & [frederic.barraquand@u-bordeaux.fr](mailto:frederic.barraquand@u-bordeaux.fr)

## **1 Introduction**

2 Phytoplankton communities are among the most important photosynthetic groups on Earth, being at the bottom  
3 of the marine food chain, and responsible for approximately half the global primary production (Field *et al.*,  
4 1998). Their contribution to ecosystem functions is only matched by their contribution to biodiversity. Indeed,  
5 phytoplankton communities are characterized by a surprisingly high number of species. For example, a single  
6 sample as small as a few mL can contain up to seventy species (REPHY, 2017; Widdicombe & Harbour, 2021).  
7 This observation is usually called the “paradox of the plankton” (Hutchinson, 1961), which refers to the conflict  
8 between the observed diversity of species competing for similar resources in a seemingly homogeneous environment,  
9 and models predicting that only a few species will persist by outcompeting the others (MacArthur & Levins, 1964;  
10 Huisman & Weissing, 1999; Schippers *et al.*, 2001). Phytoplankton models for coexistence are now almost as diverse  
11 as their model organisms (Record *et al.*, 2014), but they often describe only a handful of species, which does  
12 not correspond to the diversity observed in the field. When modeling rich communities ( $> 10$  species), classical  
13 answers to the plankton paradox involving temporal fluctuations (e.g., Li & Chesson, 2016; Chesson, 2018) are  
14 not sufficient to maintain a realistic diversity. For instance, we found that a phytoplankton community dynamics  
15 model with environmental fluctuations and storage effect still requires extra niche differentiation for coexistence,  
16 which manifests in stronger intraspecific than interspecific interactions (Picoche & Barraquand, 2019). However,  
17 it is not clear that we should resort to hidden niches to explain phytoplankton coexistence, as most models also  
18 make hidden simplifying assumptions that could be relaxed. One that we relax here is mean-field dynamics at the  
19 microscale. Indeed, field observations have revealed phytoplankton patchiness for more than a century (Bainbridge,  
20 1957; Stocker, 2012), from the macro- to the micro-scale (Leonard *et al.*, 2001; Doubell *et al.*, 2006; Font-Muñoz  
21 *et al.*, 2017).

22 Phytoplankton patchiness can at least be partly explained by the hydrodynamics of their environment: the size  
23 of these organisms is mostly below the size of the smallest eddy (i.e., the Kolmogorov scale). In a typical aquatic  
24 environment such as the ocean, phytoplankton individuals are embedded in viscous micro-structures (Peters &  
25 Marrasé, 2000) while phytoplankton populations are displaced by a turbulent flow at slightly larger scales (Martin,  
26 2003; Prairie *et al.*, 2012). Phytoplankton organisms therefore live in an environment where fluid viscosity dominates  
27 at the scale of an individual but turbulent dispersion dominates on length scales characteristic of a small population  
28 of those individuals (Estrada *et al.*, 1987; Prairie *et al.*, 2012).

29 This leads us to consider demography in the context of this environmental variation created by hydrodynamic  
30 processes. Individual-based models provide a convenient depiction of population dynamics and movement at the  
31 microscale (Hellweger & Bucci, 2009). In this framework, population growth is a result of individual births and  
32 deaths. Aggregation of individuals can emerge from local reproduction coupled with limited dispersal, which can  
33 happen in a fluid where turbulence and diffusion are not strong enough to disperse kin aggregates (Young *et al.*,  
34 2001). The resulting local aggregation can then affect the community dynamics at larger spatial scales, even when

35 all competitors are equivalent (i.e., with equal interaction strengths irrespective of species identity). Indeed, the  
36 combination of local dispersal after reproduction and local interactions leads to stronger intraspecific interactions  
37 than interspecific interactions at the population level (Dettlo & Muller-Landau, 2016). This mechanism stabilizes  
38 the community, as a high intra-to-interspecific interaction strength ratio makes a species control its abundance more  
39 than it controls the abundance of other species, which is associated with coexistence in theoretical models (Levine  
40 & HilleRisLambers, 2009; Barabás *et al.*, 2017) and often observed in the field at the population level (Adler *et al.*,  
41 2018; Picoche & Barraquand, 2020). Therefore, the microscale spatial distribution of individuals likely affects the  
42 interaction structure within a community (Haegeman & Rapaport, 2008), and may sustain diversity.

43 Existing models of phytoplankton populations near the Kolmogorov scale — between 1 mm and 1 cm in an  
44 oceanic environment (Barton *et al.*, 2014) — focus on a single species and the clustering of its individuals (Young  
45 *et al.*, 2001; Birch & Young, 2006; Bouderbala *et al.*, 2018; Breier *et al.*, 2018). They share similarities to dynamic  
46 point process models (Law *et al.*, 2003; Bolker & Pacala, 1999; Plank & Law, 2015) developed initially with larger  
47 organisms in mind. When phytoplankton individual-based models consider multiple types of organisms, they focus  
48 for now on how organisms with opposite characteristics (e.g., increase versus decrease in density with turbulence in  
49 Borgnino *et al.*, 2019; Arrieta *et al.*, 2020) segregate spatially, or on coexistence for species that have contrasting  
50 trait values (e.g., size in Benczik *et al.*, 2006). This is useful as an explanation of how species with marked differences  
51 might coexist. The difficulty of the coexistence problem, however, is that we have to explain how closely related  
52 species or genera (e.g., within diatoms), many of whom have similar size, buoyancy, chemical composition, etc.,  
53 manage to coexist within a single trophic level. This requires modelling *similar* species in a spatially realistic  
54 environment and objectively quantifying whether they aggregate or segregate in space.

55 To do so, we build a multispecies version of the Brownian Bug Model (BBM) of Young *et al.* (2001), an  
56 individual-based model which includes an advection process mimicking a turbulent fluid flow, passive diffusion of  
57 organisms, as well as stochastic birth and death processes. The initial version of this model (Young *et al.*, 2001)  
58 coupled limited dispersal and local reproduction with ocean-like microscale hydrodynamics, and showed spatial  
59 clusters of individuals of the same species. The original BBM was limited to a single species and was illustrated  
60 with two-dimensional simulations. The model was not strongly quantitative (Picoche *et al.*, 2022) in the sense that  
61 parameters were not informed by current knowledge on phytoplankton biology (numbers of cells per liter, diffusion  
62 characteristics, etc.). As phytoplankton organisms live in a three-dimensional environment, informing the model  
63 with more realistic parameters requires us to shift to three dimensions. We also extend the model to multiple species,  
64 and consider two size classes for our phytoplankton communities, which are either made of nanophytoplankton ( $3 \mu\text{m}$   
65 diameter,  $\approx 10^6$  cells  $\text{L}^{-1}$ ) or microphytoplankton ( $50 \mu\text{m}$ ,  $\approx 10^4$  cells  $\text{L}^{-1}$ ). We populate each community with 3  
66 to 10 different species.

67 The Brownian Bug model (in its original single-species form as in the multispecies version considered here) is  
68 related to spatial branching processes. Without advection, it combines a continuous-time, discrete-state model for

69 population growth and a continuous-time, continuous-space Brownian motion for particle diffusion (Birch & Young,  
70 2006). It is further complexified by a turbulent flow in Young *et al.* (2001); Picoche *et al.* (2022) as well as here.  
71 In spite of this complexity, it remains possible to derive the dynamics of pair density functions, which quantify  
72 the degree of intra- and interspecific clustering of organisms, via correlations between positions of organisms (see  
73 next section). Thus we can understand emergent spatial structure in analytic detail and compare these predictions  
74 to the results from three-dimensional simulations. Furthermore, because we do not consider direct interactions  
75 between organisms, the multispecies spatial point process that represents the stable state of the BBM is a random  
76 superposition of spatial point processes for each species (Illian *et al.*, 2008). This enables us to derive, in addition  
77 to pair correlation functions, analytical formulas for the species composition in the neighbourhood of an individual,  
78 which are more readily ecologically interpreted than pair density or correlation functions.

## 79 Model and spatial statistics

### 80 Brownian Bug Model

81 The Brownian Bug Model (BBM) describes the dynamics of individuals in a turbulent and viscous environment,  
82 including demographic processes. The model is continuous in space and time. Here we extend the mostly two-  
83 dimensional, monospecific version in Young *et al.* (2001), to three dimensions and  $S$  species.

84 Each individual is characterized by its species identity  $i$  and its position  $\mathbf{x}^T = (x, y, z)$ . The population dynamics  
85 are modelled by a linear birth-death process with birth rate  $\lambda_i$  and death rate  $\mu_i$ . Each individual independently  
86 follows a Brownian motion with diffusivity  $D_i$ , and is advected by a common stochastic and chaotic flow modelling  
87 turbulence. The model applies in the Batchelor regime, which means that the separation  $s(t)$  between two individuals  
88  $k$  and  $l$  grows exponentially with time with stretching parameter  $\gamma$ , i.e.  $s(t) = \ln(|\mathbf{x}_k - \mathbf{x}_l|(t)) \propto 3\gamma t$  (Kraichnan,  
89 1974; Young *et al.*, 2001).

90 Within a given community (the set of all individuals of the  $S$  species), all species share the same parameters:  
91  $\lambda_i$ ,  $\mu_i$  and  $D_i$  values can change between communities, as we later consider small and large phytoplankton, but are  
92 set to common values within a community. On the contrary,  $\gamma$  describes the environment and is not community-  
93 specific, i.e., all individuals are displaced by the same turbulent stirring. For numerical simulations, time needs to  
94 be discretized (this is required for diffusion and advection modelling). The approximated model advances through  
95 time in small steps of duration of  $\tau$ . During each interval, events unroll as follows:

- 96 1. Demography: each individual can either reproduce with probability  $p_i = \lambda_i \tau$  (forming a new individual of the  
97 same species  $i$  at the same position  $\mathbf{x}$  as the parent), die with probability  $q_i = \mu_i \tau$ , or remain unchanged with  
98 probability  $1 - p_i - q_i$ .
- 99 2. Diffusion: each individual moves to a new position  $\mathbf{x}(t') = \mathbf{x}(t) + \delta\mathbf{x}(t)$ , with  $t < t' < t + \tau$ . The random

100 displacement  $\delta\mathbf{x}(t)$  is drawn from a Gaussian distribution  $\mathcal{N}(0, \Delta_i^2)$  with  $D_i = \Delta_i^2/2\tau$  the diffusivity. This  
 101 diffusive step separates the initially coincident pairs produced by reproduction in step 1 above.

102 3. Turbulence: each individual is displaced by a turbulent flow, modelled with the Pierrehumbert map (Pierre-  
 103 humbert, 1994), adapted to three dimensions following Ngan & Vanneste (2011). Thus given the position at  
 104 time  $t'$  the updated position at time  $t + \tau$  is

$$\begin{aligned} x(t + \tau) &= x(t') + \frac{U\tau}{3} \cos(ky(t') + \phi(t)) \\ y(t + \tau) &= y(t') + \frac{U\tau}{3} \cos(kz(t') + \theta(t)) \\ z(t + \tau) &= z(t') + \frac{U\tau}{3} \cos(kx(t + \tau) + \psi(t)). \end{aligned} \quad (1)$$

105 Above,  $U$  is the velocity of the flow,  $k = 2\pi/L_s$  is the wavenumber for the flow at the length scale  $L_s$  (see below)  
 106 and  $\phi(t), \theta(t), \psi(t)$  are random phases drawn from a uniform distribution between 0 and  $2\pi$ ; these phases remain  
 107 constant during the interval between  $t$  and  $t + \tau$ . The shift from continuous to discrete-time turbulence modelling  
 108 is described in Section S1 in the Supplementary Information. The velocity  $U$  is related to  $\gamma$ . As the separation  
 109 between two points grows exponentially with parameter  $3\gamma$  due to turbulence, the exponent  $\gamma$  can be estimated as  
 110 the slope of  $1/3 \langle \ln(s(t)) \rangle = f(t)$  in the absence of diffusion and demography (Young *et al.*, 2001; Picoche *et al.*,  
 111 2022).

112 Individuals are distributed in a cube of side length  $L$ , with periodic boundary conditions. The cube dimen-  
 113 sions are determined to balance computing costs and realistic concentrations of individuals; they represent the  
 114 accumulation of a few volumes of scale  $L_s$ .

## 115 Characterization of the spatial distribution

116 Let  $W$  be the observation window (in our case, the whole cube, which we never subsample hereafter). The state of the  
 117 system at time  $t$  can be described as a collection of  $S$  populations, where the population of species  $i$  is made of  $n_i$  indi-  
 118 viduals randomly distributed in  $W$ , with positions  $\mathbf{X}_i(t) = [\mathbf{x}_{1,i}(t), \mathbf{x}_{2,i}(t), \dots, \mathbf{x}_{n_i,i}(t)]$ .  $\mathbf{X}(t) = [\mathbf{X}_1(t), \dots, \mathbf{X}_S(t)]$   
 119 arises from a stochastic and spatial individual-based model changing through time, but can also be analyzed as a  
 120 spatial point process at time  $t$ . We note that the point distributions remain the same for all spatial translations  $\xi$   
 121 (i.e., the point process described by the set  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]$  is the same as  $\mathbf{X}_\xi = [\mathbf{x}_1 + \xi, \mathbf{x}_2 + \xi, \dots, \mathbf{x}_k + \xi]$ ): the  
 122 process is stationary.

123 A useful method to characterize a spatial point process is the use of spatial moments (illustrated in Section S2  
 124 of the SI for simple spatial point processes). These can be theoretically derived and used to check simulations. The  
 125 spatial moments of a process are, however, merely statistical indicators which then need to be related to more easily  
 126 ecologically interpretable quantities. This is the role of the dominance index, which we present below.

<sup>127</sup> **Spatial moments**

<sup>128</sup> The first-order moment is the intensity of the process, or mean concentration of individuals, whose empirical estimate  
<sup>129</sup> is  $C_i = \frac{\widehat{N_i(W)}}{V(W)}$ , where  $\widehat{N_i(W)}$  is the empirical number of individuals of species  $i$  in the cube  $W$  and  $V(W) = L^3$  is  
<sup>130</sup> the volume of the cube; it does not give any information regarding the spatial distribution of individuals, and their  
<sup>131</sup> spatial correlations.

<sup>132</sup> The second-order product density, or pair density  $G(r, t)$ , is the expected density of pairs of points separated  
<sup>133</sup> by a distance  $r$  (Law *et al.*, 2003). A similar characteristic can be used for marked spatial point process. In our  
<sup>134</sup> case, the marks are the species' identities, and we can define  $G_{ij}(r, t)$ , so that  $G_{ij}(r, t)d\mathbf{x}_Ad\mathbf{x}_B$  is the probability of  
<sup>135</sup> finding an individual of species  $i$  in volume  $d\mathbf{x}_A$  and an individual of species  $j$  in volume  $d\mathbf{x}_B$ , with the distance  
<sup>136</sup> between the centers of  $d\mathbf{x}_A$  and  $d\mathbf{x}_B$  equal to  $r$  (pages 219 and 325 in Illian *et al.*, 2008). We define  $\xi$  as the vector  
<sup>137</sup> connecting the center of  $d\mathbf{x}_A$  to the center of  $d\mathbf{x}_B$ , while  $r = |\xi|$  is the radial distance. We show in Picoche *et al.*  
<sup>138</sup> (2022) that the intraspecific pair density  $G_{ii}(r, t)$ , in three dimensions, is a solution of

$$\frac{\partial G_{ii}}{\partial t}(r, t) = \frac{2D_i}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial G_{ii}}{\partial r} \right) + \frac{\gamma}{r^2} \frac{\partial}{\partial r} \left( r^4 \frac{\partial G_{ii}}{\partial r} \right) + 2(\lambda_i - \mu_i)G_{ii} + 2\lambda_i C_i \delta(\xi). \quad (2)$$

<sup>139</sup> The pair correlation function  $g_{ij}(r, t)$ , or pcf, can be derived from the pair density and is defined as

$$g_{ij}(r, t) = \frac{G_{ij}(r, t)}{C_i C_j}. \quad (3)$$

<sup>140</sup> The pcf is equal to one when the spatial distribution of species  $i$  individuals is random relative to species  $j$  individuals.  
<sup>141</sup> To compute the intraspecific pcf  $g_{ii}(r, t)$  at steady state, considering a population at equilibrium, we integrate Eq.  
<sup>142</sup> 2 (see Appendices, Eqs. 19-30) with  $\lambda_i = \mu_i$  and obtain

$$g_{ii}(r) = 1 + \frac{\lambda_i}{4\pi D_i C_i \ell_{B,i}} \left( \frac{\ell_{B,i}}{r} + \arctan \left( \frac{r}{\ell_{B,i}} \right) - \frac{\pi}{2} \right), \quad (4)$$

<sup>143</sup> where  $\ell_{B,i} = \sqrt{2D_i/\gamma}$  approximates the Batchelor scale for species  $i$ .

<sup>144</sup> The system converges rapidly to the solution in Eq. 4 in the presence of advection. However, when there is no  
<sup>145</sup> turbulent advection, convergence is much slower, to the point that an equilibrium assumption requires unrealistically  
<sup>146</sup> long timeframes (see Section S3 in the SI). We therefore need a time-dependent formula for the pcf in the absence of  
<sup>147</sup> advection, which can be obtained in the case where  $\gamma = 0$  using a Green's function (see derivation in the Appendices,  
<sup>148</sup> Eqs. 31-37),

$$g_{ii}(r, t) = 1 + \frac{\lambda_i}{4\pi r D_i C_i} \left\{ 1 - \operatorname{erf} \left( \frac{r}{\sqrt{8D_i t}} \right) \right\}. \quad (5)$$

<sup>149</sup> The above equations match when  $\gamma \rightarrow 0$  and  $t \rightarrow +\infty$ .

<sup>150</sup> As populations of different species do not directly interact, each population is an independent realization of a

151 point process, which means that the distribution of all individuals within the community at time  $t$  is a random  
 152 superposition of stationary point processes and thus  $g_{ij}(r, t) = 1$  if  $i \neq j$  (Illian *et al.*, 2008, p. 326, eq. 5.3.13).

153 Related to the pair correlation function is Ripley's  $K$ -function  $K(r)$ . Using its marked version,  $C_j K_{ij}(r)$  is the  
 154 average number of points of species  $j$  surrounding an individual of species  $i$  within a sphere of radius  $r$  (Illian *et al.*,  
 155 2008), i.e.,

$$\forall r \geq 0, K_{ij}(r) = \frac{1}{C_j} \mathbb{E}_i (N_j(b(o, r) \setminus \{o\})), \quad (6)$$

156 where  $\mathbb{E}_i$  is the expectation with respect to individuals of species  $i$  and  $N_j(b(o, r) \setminus \{o\})$  is the number of individuals  
 157 of species  $j$  in the sphere of radius  $r$  centered on  $o$ , not counting  $o$  itself.  $K_{ij}(r)$  is related to  $g_{ij}(r)$  as

$$g_{ij}(r) = \frac{K'_{ij}(r)}{4\pi r^2}. \quad (7)$$

158 Combining Eq. 7 and, when  $U > 0$ , Eq. 4, we can show that (see Appendices, Eqs. 38-44)

$$K_{ii}(r) = \frac{4}{3}\pi r^3 + \frac{\lambda_i r^3}{3D_i C_i \ell_{B,i}} \left( \frac{\ell_{B,i}}{r} + \frac{\ell_{B,i}^3 \log\left(\frac{r^2}{\ell_{B,i}^2} + 1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{B,i}}\right) - \frac{\pi}{2} \right). \quad (8)$$

159 When  $U = 0$ , we need a time-dependent solution corresponding to our simulation duration, i.e. (see Appendices,  
 160 Eq. 46-51)

$$K_{ii}(r, t) = \frac{4}{3}\pi r^3 + \frac{\lambda_i r^2}{C_i D_i} \left( \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right) \left( 1 - \frac{4D_i}{r^2} t \right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi} r} e^{-\frac{r^2}{8D_i t}} \right). \quad (9)$$

161 For random superposition of stationary point processes,  $K_{ij}(r, t) = \frac{4}{3}\pi r^3$  if  $i \neq j$  (Illian *et al.*, 2008, p. 324, eq.  
 162 5.3.5).

### 163 Dominance index

164 The dominance index (defined in Table S1 in the Supporting Information of Wiegand *et al.*, 2007) is the ratio  
 165 between the number of conspecifics and the number of individuals of all species surrounding a given individual.

166 Let  $M_{ij}(r)$  be the average number of individuals of species  $j$  within a circle of radius  $r$  around an individual  
 167 of species  $i$ , which can also be written with Ripley's  $K$ -function as  $M_{ij}(r) = C_j K_{ij}(r)$ .  $M_{ii}(r)$  corresponds to the  
 168 conspecific neighbourhood and  $M_{io}(r) = \sum_{j=1, j \neq i}^S M_{ij}(r)$  corresponds to individuals of all other species. We can  
 169 then define  $\mathcal{D}_i$  as

$$\begin{aligned} \mathcal{D}_i(r) &= \frac{M_{ii}(r)}{M_{ii}(r) + M_{io}(r)} \\ &= \frac{C_i K_{ii}(r)}{\sum_{j=1}^S C_j K_{ij}(r)}. \end{aligned} \quad (10)$$

170 When individuals of the same species  $i$  tend to cluster,  $\mathcal{D}_i(r)$  tends to 1 while it tends to the proportion of individuals  
 171 of species  $i$  in the whole community when the distribution is uniform (Section S2 of the SI).

172 Using Eq. 9 and 10, we obtain the formula for the dominance index in the presence of advection as

$$\mathcal{D}_i(r) = \frac{\frac{\lambda_i}{3D_i\ell_{B,i}} \left( \frac{\ell_{B,i}}{r} + \frac{\ell_{B,i}^3 \log\left(\frac{r^2}{\ell_{B,i}^2} + 1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{B,i}}\right) - \frac{\pi}{2} \right) + \frac{4}{3}\pi C_i}{\frac{\lambda_i}{3D_i\ell_{B,i}} \left( \frac{\ell_{B,i}}{r} + \frac{\ell_{B,i}^3 \log\left(\frac{r^2}{\ell_{B,i}^2} + 1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{B,i}}\right) - \frac{\pi}{2} \right) + \sum_{j=1}^S \frac{4}{3}\pi C_j}. \quad (11)$$

173 In the absence of advection ( $U = 0, \gamma = 0$ ), we use the time-dependent dominance index, computed similarly:

$$\mathcal{D}_i(r, t) = \frac{\frac{\lambda_i}{D_i r} \left( \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right) \left(1 - \frac{4D_i}{r^2}t\right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi r}} e^{-\frac{r^2}{8D_i t}} \right) + \frac{4}{3}\pi C_i}{\frac{\lambda_i}{D_i r} \left( \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right) \left(1 - \frac{4D_i}{r^2}t\right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi r}} e^{-\frac{r^2}{8D_i t}} \right) + \sum_{j=1}^S \frac{4}{3}\pi C_j}. \quad (12)$$

## 174 Parameters

175 We model two types of organisms: microphytoplankton (defined by a diameter between 20 and 200  $\mu\text{m}$ , here 50  
 176  $\mu\text{m}$ ) and nanophytoplankton (defined by a diameter between 2 and 20  $\mu\text{m}$ , here 3  $\mu\text{m}$ ). These two groups are  
 177 characterized respectively by a low diffusivity, slow growth and lower concentration vs. high diffusivity, fast growth  
 178 and higher concentration. Organisms are displaced by a turbulent fluid whose velocity defines the time scale of the  
 179 discretized model: we give here the reasoning behind parameter values, keeping in mind that our model parameters  
 180 are only approximate. Main parameter definitions and values are given in Table 1.

## 181 Advection

182 We first consider the advection process, due to the turbulence of the environment. We only consider the Batchelor-  
 183 Kolmogorov regime, i.e., the size of the volume  $W$  is below the size of the smallest eddy, but above the smallest  
 184 length scale of fluctuations in nutrient concentrations. The defining scale of the environment therefore corresponds  
 185 to a Reynolds number

$$\operatorname{Re} = \frac{U}{k\nu} \approx 1 \quad (13)$$

186 where  $\nu = 10^{-6} \text{ m}^2 \text{ s}^{-1}$  is the kinematic viscosity for water. The smallest wavenumber  $k$  corresponds to the largest  
 187 length scale  $L_s$  (Kolmogorov scale), i.e.,  $k = 2\pi/L_s$ , with  $L_s \approx 1 \text{ cm}$  in the ocean (Barton *et al.*, 2014). The  
 188 definition of the Reynolds number leads to

$$\begin{aligned} 1 &\approx \frac{UL_s}{2\pi\nu} \\ \Leftrightarrow U &\approx \frac{2\pi\nu}{L_s}. \end{aligned} \quad (14)$$

189 This means that  $U = 6.3 \times 10^{-4} \text{ m s}^{-1} = 5.4 \times 10^3 \text{ cm d}^{-1}$ . Using  $U\tau/3 = 0.5 \text{ cm}$  as in Young *et al.* (2001),  
 190 we have  $\tau = 2.8 \times 10^{-4} \text{ d} = 24 \text{ s}$ . When  $U\tau/3 = 0$ , the environment is only diffusive, we keep the same value for  $\tau$ .  
 191 For  $U\tau/3 = 0.5 \text{ cm}$ ,  $\gamma = 1231 \text{ d}^{-1}$ .

192 **Diffusion**

193 If we use the Stokes-Einstein equations (Einstein, 1905, cited from Dusenbery, 2009), diffusivity can be computed  
194 with

$$D_i = \frac{RT}{N_A} \frac{1}{6\pi\eta a_i} \quad (15)$$

195 where  $R = 8.314 \text{ J K}^{-1} \text{ mol}^{-1}$  is the molar gas constant,  $T = 293 \text{ K}$  is the temperature,  $N_A = 6.0225 \times 10^{23}$  is  
196 Avogadro's number,  $\eta = 10^{-3} \text{ m}^{-1} \text{ kg s}^{-1}$  is the dynamic viscosity of water and  $a_i$  is the radius of the organism  
197 considered.

198 Using  $D_i = \frac{\Delta_i^2}{2\tau}$ , we find that

$$\begin{aligned} \Delta_i &= \sqrt{2\tau D_i} \\ \Leftrightarrow \Delta_i &= \sqrt{\frac{RT}{N_A} \frac{\tau}{3\pi\eta a_i}}. \end{aligned} \quad (16)$$

199 We consider  $a_n = 1.5 \mu\text{m}$  for nanophytoplankton individuals and  $a_m = 25 \mu\text{m}$  for microphytoplankton individ-  
200 uals, which allows us to compute  $\Delta_n$  and  $\Delta_m$  (see Table 1).

201 **Ecological processes**

202 We study the community at equilibrium, with the birth rate equal to the death rate, i.e.,  $p_i = q_i \forall i$ . We use a micro-  
203 phytoplankton doubling rate of  $1 \text{ d}^{-1}$  (Bissinger *et al.*, 2008) and consider the fastest-growing nanophytoplankton  
204 species, corresponding to a diameter of  $3 \mu\text{m}$  (Bec *et al.*, 2008), for which the doubling rate is between  $2$  and  $3 \text{ d}^{-1}$   
205 (set to  $2.5 \text{ d}^{-1}$  here).

Parameter	Definition	Value
$p_m, q_m$	Probability of reproducing/dying for microphytoplankton individuals	$2.8 \times 10^{-4}$
$p_n, q_n$	Probability of reproducing/dying for nanophytoplankton individuals	$6.9 \times 10^{-4}$
$U$	Turbulent advection speed	$\{0, 0.06\} \text{ cm.s}^{-1}$
$\Delta_m$	Diffusion parameter for microphytoplankton individuals	$6.4 \times 10^{-5} \text{ cm}$
$\Delta_n$	Diffusion parameter for nanophytoplankton individuals	$2.6 \times 10^{-4} \text{ cm}$

Table 1: Definitions and values of the main parameters used in the three-dimensional BBM, assuming the duration of a time step  $\tau$  is 24 seconds.

206 **Range of interaction**

207 As we examine individual aggregation and its potential effects on interactions between species, we have to ascertain  
208 the volume in which an individual can be affected by the presence of other individuals, or affect other individuals.  
209 We only consider here interactions due to competition for nutrients, and therefore need to define a nutrient depletion  
210 volume. We approximate this volume as the sphere of radius  $r$  where  $C(r) \leq 90\% C_\infty$  with  $C_\infty$  the background  
211 concentration of the nutrient. The radius of this nutrient depletion volume is maximized when the individual is in  
212 stagnant water so that diffusion is the only hydrodynamic process. In this case, the depletion radius corresponds to  
213 10 times the radius of the individual (Jumars *et al.*, 1993; Karp-Boss *et al.*, 1996). We define the maximum distance

214 which allows for potential interactions (due to competition for resources) between two individuals of radius  $a_i$  and  
215  $a_j$  as  $d_{\text{threshold}}$ , and the corresponding volume of potential interactions around an organism as  $V_{\text{int}} = 4/3\pi d_{\text{threshold}}^3$   
216 with

$$d_{\text{threshold}} = 10a_i + 10a_j. \quad (17)$$

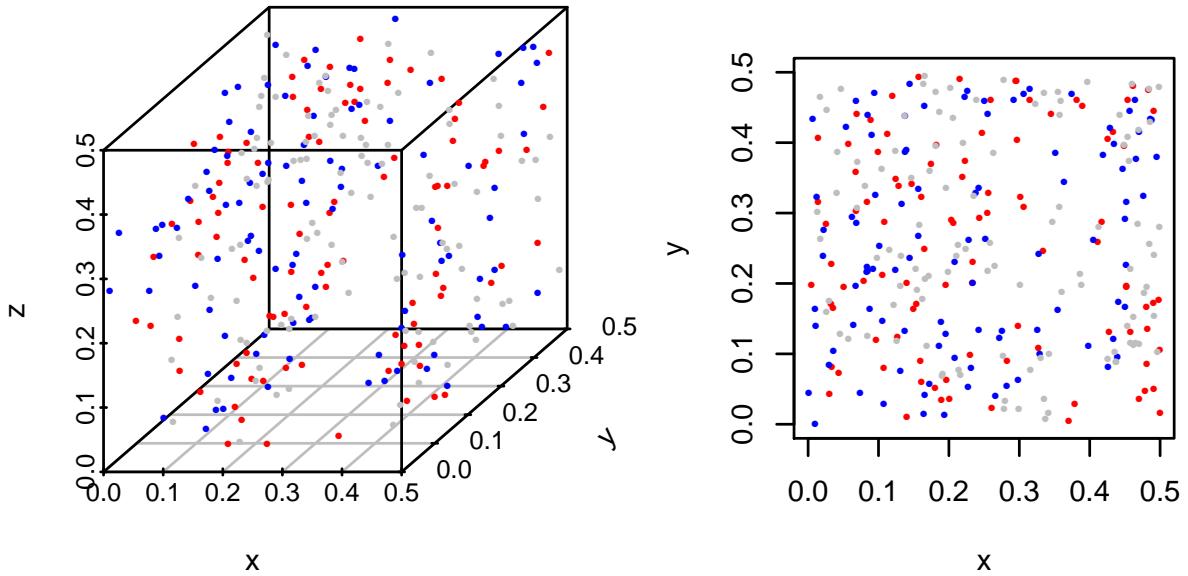
217 We consider this maximum value as our baseline, keeping in mind that turbulence reduces the size of the nutrient  
218 depletion volume and increases the nutrient flux to the cell (Arnott *et al.*, 2021). We caution that determination  
219 of the shape of the nutrient depletion volume in the presence of turbulence is too complex to be addressed here  
220 (Karp-Boss *et al.*, 1996).

221 We consider a total volume of  $1000 \text{ cm}^3$  for microphytoplankton and  $10 \text{ cm}^3$  for nanophytoplankton (volumes are  
222 adapted to balance realistic concentrations and computation time) with periodic boundary conditions. Individuals  
223 are uniformly distributed in the cube at the beginning of the simulation. We run an idealized simulation with 3  
224 species with an even abundance distribution of about  $10^4 \text{ cells L}^{-1}$  for microphytoplankton (Picoche & Barraquand,  
225 2020) and  $10^6 \text{ cells L}^{-1}$  for nanophytoplankton individuals (Edwards, 2019). We then model a more realistic com-  
226 munity with 10 species having a skewed abundance distribution (between 55,000 and 400 cells  $\text{L}^{-1}$  for microphyto-  
227 plankton, according to observations of field abundance distributions in Picoche & Barraquand, 2020, and multiplied  
228 by  $10^2$  for nanophytoplankton). All simulations are run for 1000 time steps of duration  $\tau$  (corresponding to approx-  
229 imately 6h40). The computation of  $g$  and  $K$  of simulated distributions is explained in Section S4 of the SI. The  
230 code for all simulations and analyses can be found at [https://github.com/CoraliePicoche/brownian\\_bug\\_3D/](https://github.com/CoraliePicoche/brownian_bug_3D/).

## 231 Results

232 We show an example of nanophytoplankton spatial distributions with and without advection at the end of a sim-  
233 ulation in Fig. 1: clustering is not visible to the naked eye, even when zooming in on the observation volume, in  
234 the presence of advection, but removing turbulence helps visualising small aggregates of conspecifics. Microphyto-  
235 plankton distributions are not so easy to analyse as no clusters can be detected from visual observations (although  
236 it may actually be present), whether advection is included or not (Section S5 of the SI). Statistics are therefore  
237 needed to go further in detecting patterns of aggregation.

## Advection



## No advection

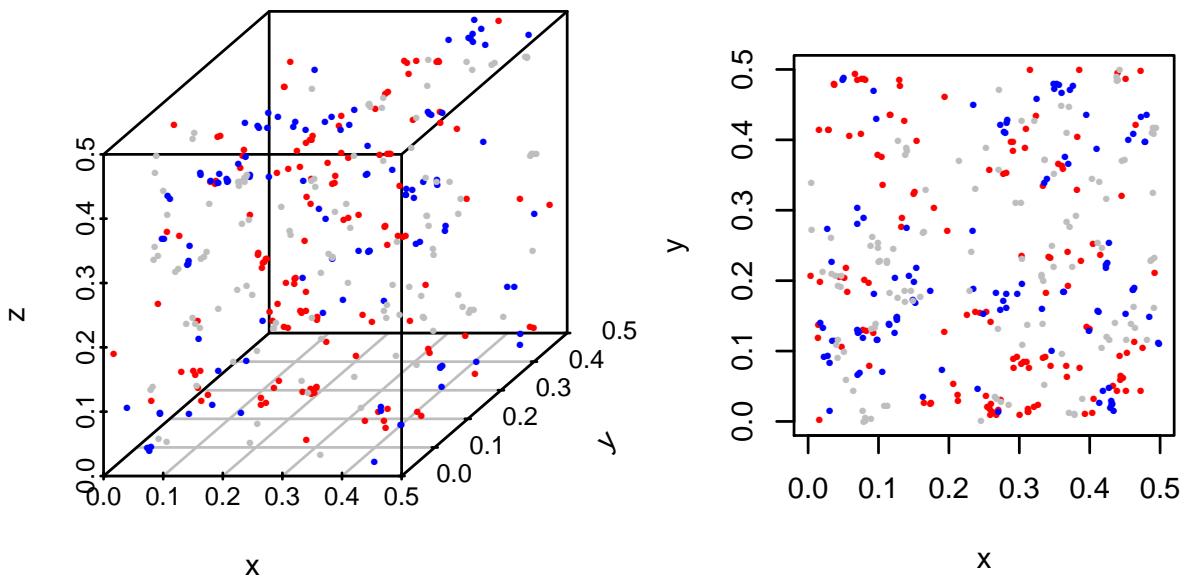


Figure 1: Spatial distributions of a 3-species community of nanophytoplankton with and without advection with density  $C = 10^3$  cells  $\text{cm}^{-3}$  after 1000 time steps. Each color corresponds to a different species. On the left-hand side, only a zoom on a  $0.5 \times 0.5 \times 0.5 \text{ cm}^3$  cube is shown, and its projection on the x-y plane is shown on the right-hand side.

<sup>238</sup> Ripley's  $K$ -functions extracted from numerical simulations match theoretical formula (Fig. 2) for both types of  
<sup>239</sup> organisms, which also indicates that dominance indices extracted from the simulations match theoretical expecta-  
<sup>240</sup> tions.

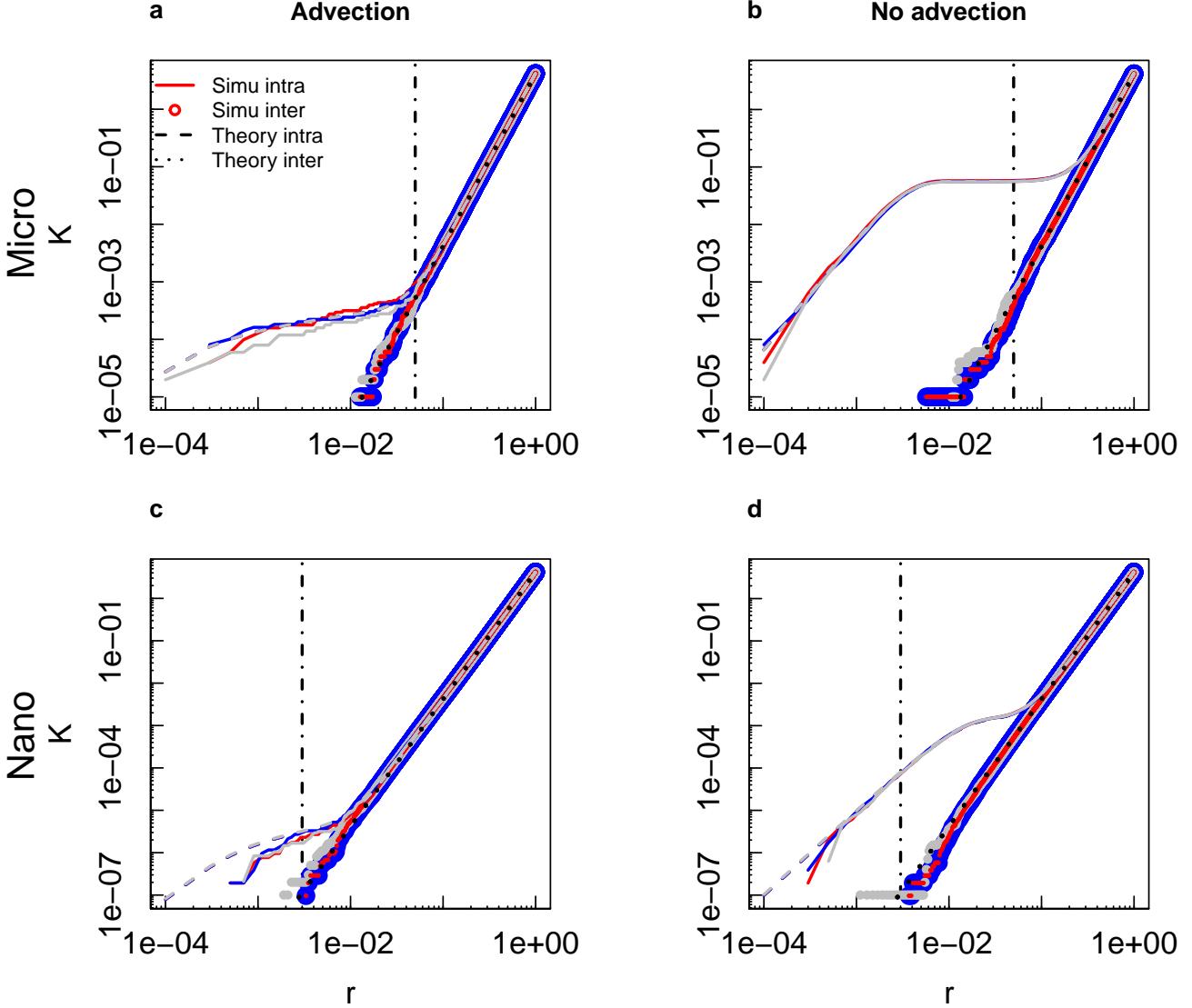


Figure 2: Comparison of theoretical and simulated Ripley’s  $K$ -functions as a function of distance (in cm) for microphytoplankton (a-b) and nanophytoplankton (c-d) in a 3-species community with even abundance distributions after 1000 timesteps, with (a, c) and without (b, d) advection. Each color represents a different species. Intraspecific  $K$ -functions are shown with dashed (theoretical values) and solid (simulated values) lines. Interspecific  $K$ -functions are shown with dotted lines (theoretical values) and circles (simulated values). The black dash-dotted line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

241 Dominance indices all follow a similar pattern (Fig. 3 and 4). The dominance index is close to 1 for small  
 242 distances: there is always a scale at which an organism is surrounded almost only by conspecifics. The index  
 243 then decreases sharply to converge at large distances (close to 1 cm) to the proportion of the focus species in the  
 244 whole community, as it would for a uniform spatial distribution. Patterns differ at intermediate ranges of distances  
 245 between organisms.

246 In the presence of advection, the dominance index starts decreasing for a distance between 5 and 10 times smaller  
 247 than when advection is absent, which indicates that organisms are closer to heterospecifics when their environment

is turbulent. A quasi-uniform distribution is also reached for smaller distances with advection than without. Microphytoplankton species start mixing for distances larger than for nanophytoplankton species irrespective of the hydrodynamic regime surrounding them.

In a 3-species community with the same initial abundances, in the presence of advection, microphytoplankton dominance indices are between 0.37 and 0.47 at the distance threshold for potential interactions, while they are between 0.80 and 0.94 for nanophytoplankton species. In the absence of turbulence, dominance indices are all above 0.98 when the distance threshold is reached (Fig. 3). Microphytoplankton organisms are therefore as likely to share their depletion volume with conspecifics as they are with heterospecifics, but only when turbulent advection is accounted for, whereas nanophytoplankton organisms always have almost only conspecifics around them.

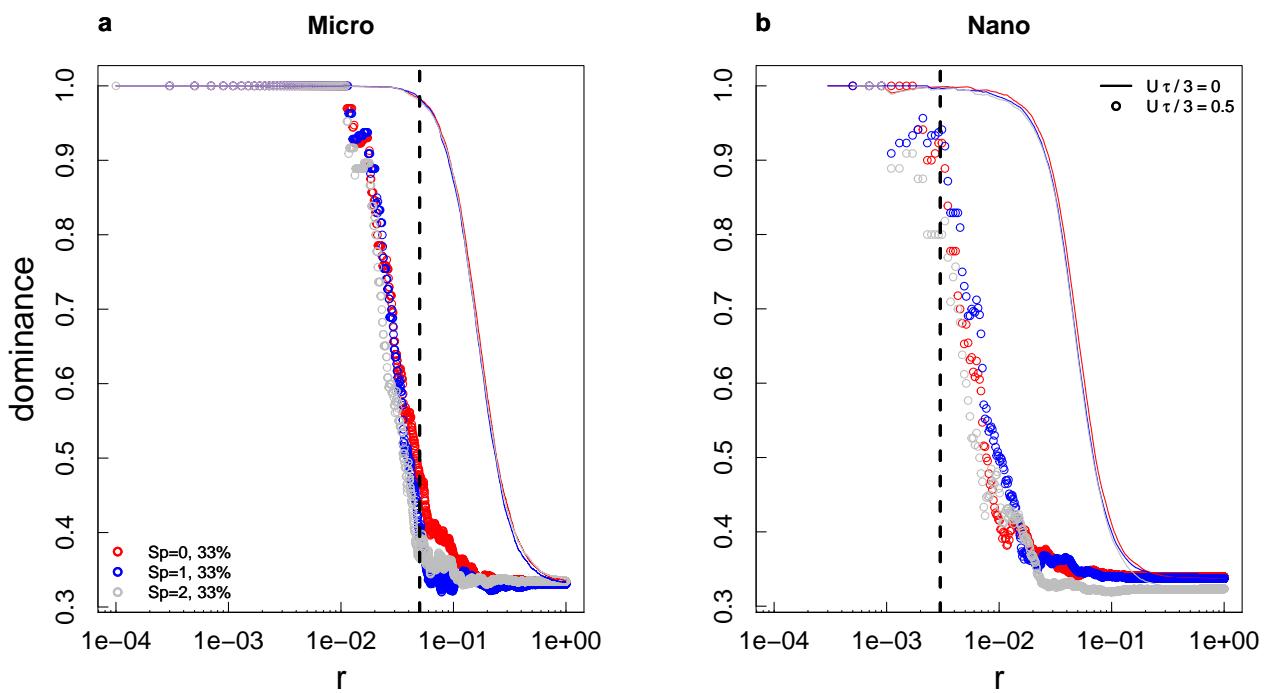


Figure 3: Dominance indices as a function of distance (in cm) for microphytoplankton (a) and nanophytoplankton (b) in a 3-species community with even abundance distributions (final proportions in the community are indicated in the figure) after 1000 timesteps, with (circles) and without (lines) advection. Each color represents a different species. The black dashed line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

More mixing in microphytoplankton than nanophytoplankton, and more mixing with advection, also holds when considering a 10 species-community with a skewed abundance distribution (Fig. 4), but dominance indices are overall lower in communities with more species and with less even abundances. In the presence of advection, microphytoplankton dominance indices at the distance threshold are between 0.34 (for the most abundant species) and 0.033 (for one of the least abundant species), while they are between 0.90 and 0.85 when advection is not taken into account. Nanophytoplankton species, too, are more mixed than in the 3 species-community: dominance indices vary between 0.54 and 0.2 when the depletion threshold is reached (with an exception of 0 for one particular species

which had no conspecific for distances below  $10^{-2}$  cm) when organisms are displaced by turbulence, while the same quantity is between 1 and 0.97 when they are only subject to diffusion.

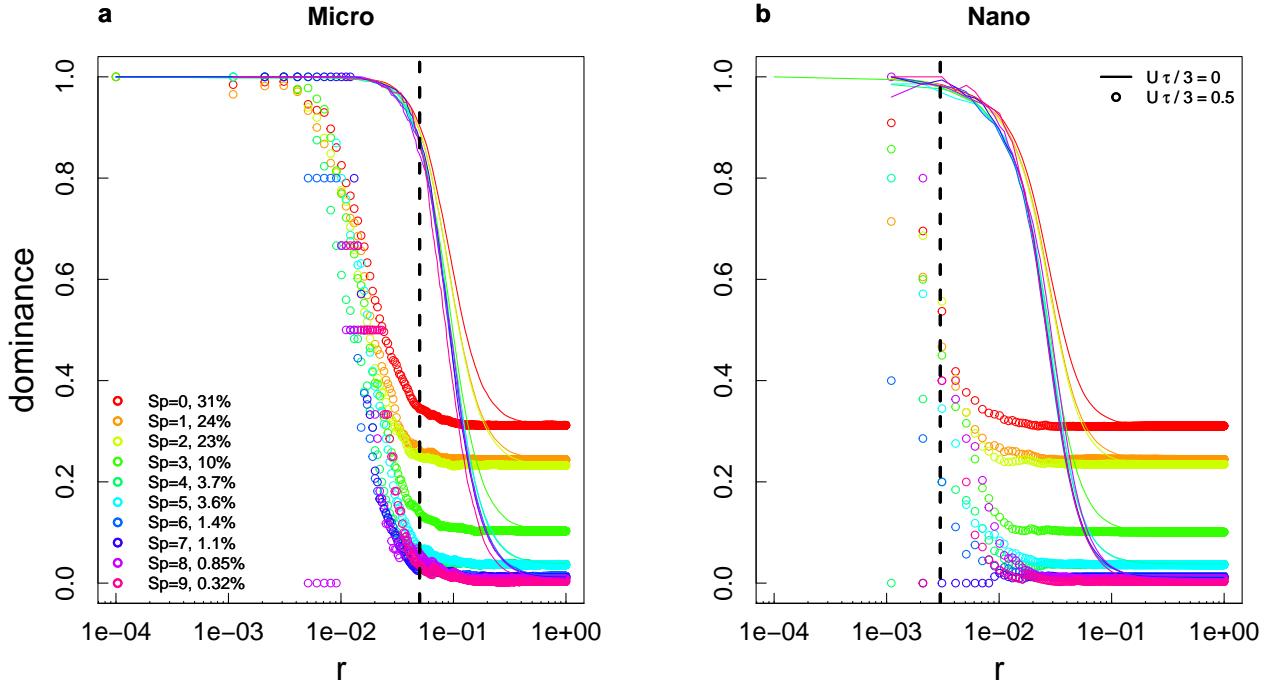


Figure 4: Dominance indices as a function of distance (in cm) for microphytoplankton (a) and nanophytoplankton (b) in a 10-species community with a skewed abundance distribution (final proportions in the community are indicated in the figure) after 1000 timesteps, with (circles) and without (lines) advection. Each color represents a different species. The black dashed line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

Differences in spatial distributions are not only due to organism sizes, which determine their demographic and hydrodynamic properties, but also to their abundances (here set through initial values). In the presence of turbulence, the threshold distance at which dominance falls below 95% is smaller for more abundant species (Fig. 5 a-b). Abundant species tend to be present nearly everywhere when they are mixed in the environment. Therefore, they are also more likely to be close to a heterospecific, but still have more conspecifics close to them than the less abundant species ( $\mathcal{D}(d_{\text{threshold}})$  increases with abundance, Fig. 5 c-d). However, this increase is less marked for nanophytoplankton than for microphytoplankton (Fig. 5 c-d). When turbulence is absent, the relationships with abundance are unclear, possibly affected by sampling effects, and we refrain from interpreting them.

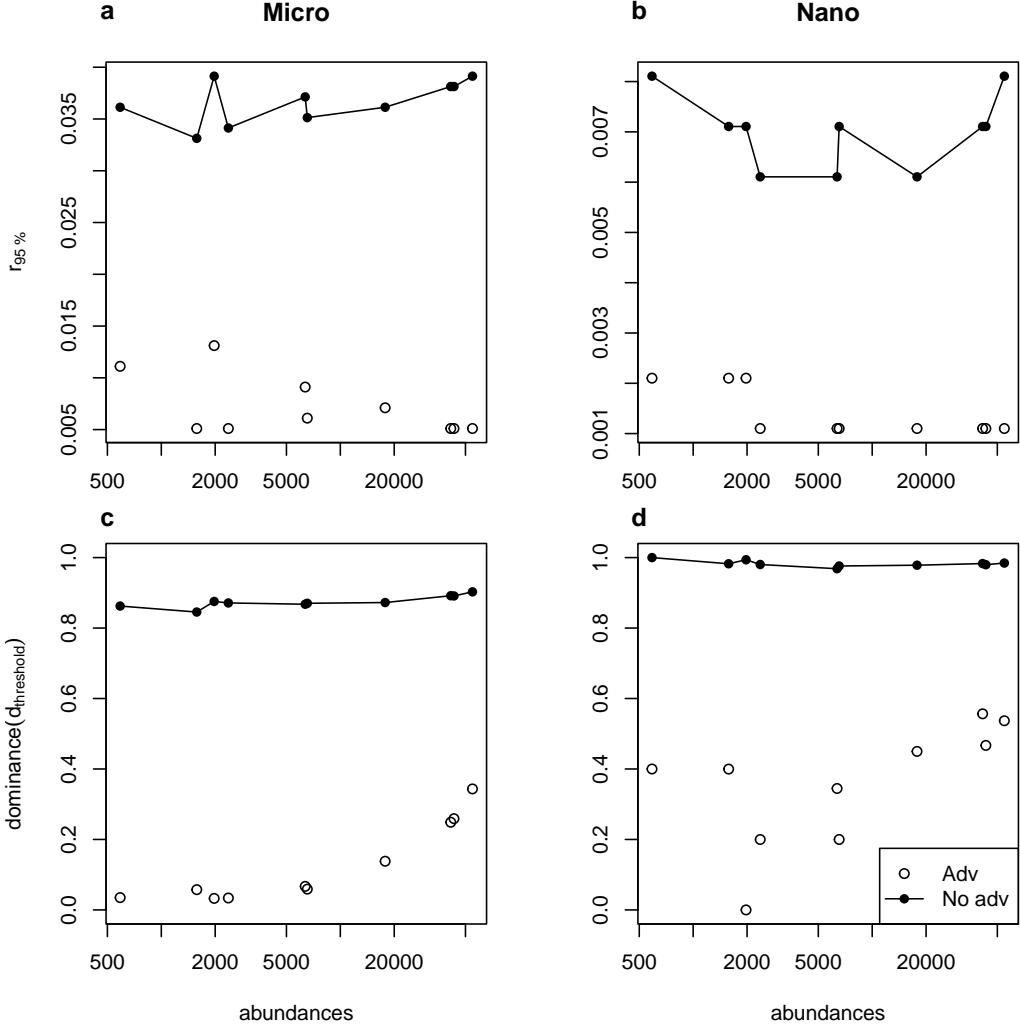


Figure 5: Minimum distances (in cm) between points for dominance to drop below 95% (a and b) and dominance at a distance corresponding to the threshold for competition (c and d) as a function of abundances (note the logarithmic scale on the x-axis) for microphytoplankton and nanophytoplankton. We consider cases with and without advection in a 10-species community with a skewed abundance distribution.

## 274 Discussion

We designed a stochastic, three-dimensional, individual-based model of the spatial distribution of multiple species in a viscous and turbulent flow. We conducted both mathematical analyses and numerical simulations to quantify spatial correlations in the distribution of organisms. We focused on the pair correlation function and Ripley's  $K$ -function, for which numerical and theoretical analyses showed a good agreement, and extracted a more ecologically-oriented metric from them, i.e., the dominance index. This statistic is the *local* average ratio of conspecifics, i.e., the number of organisms of the focal species in the neighbourhood of an individual of the same species, divided by the total number of organisms in that neighbourhood. Intraspecific clustering corresponds to a dominance index close to 1, which decreases when interspecific mixing increases. The choice of this index was motivated by two

283 reasons: (1) it is at its core a proportion of a focus species in a certain volume, i.e. a scale-dependent, localized  
284 metric bounded between 0 and 1 as opposed to other statistics whose values are less directly interpreted, and (2) it  
285 is easy to relate to coexistence theory as it describes the environment of an organism in terms of heterospecifics and  
286 conspecifics, which can, under certain hypotheses that we discuss below, be related to interspecific and intraspecific  
287 interactions. Comparing the distributions of organisms of different sizes, we showed that the presence of turbulence  
288 always increased mixing (results are robust to slight modifications in the computation of advection velocity  $U$ , shown  
289 in Section S7 of the SI). The species composition around an organism depended on its size, which mechanically  
290 determines its hydrodynamic properties (diffusivity), and is linked with its ecological characteristics (growth rate  
291 and density). Microphytoplankters (20 to 200  $\mu\text{m}$ ), larger cells with lower diffusivity, growth rate and abundance,  
292 were on average further away from other cells, due to their lower concentrations (Figure S10 of the SI), than  
293 nanophytoplankters (2 to 20  $\mu\text{m}$ ). However, they were surrounded by more heterospecifics than conspecifics within  
294 a volume of potential interactions, whose radius is defined as the maximum distance for which nutrient depletion  
295 volumes of two different individuals may overlap. If we consider that interactions between species (not modelled  
296 directly here because of timescale issues, see below) could occur with equal probability at all distances within  
297 the volume of potential interactions, we would conclude that microphytoplankters are more likely to interact with  
298 individuals from other species than with individuals of their own species. This affirmation is, however, conditional  
299 upon interactions at 10 cell diameters from an individual being equally likely than at 1 diameter from an individual.  
300 If we keep in mind that interactions are more likely or stronger at very short distances, microphytoplankters may  
301 still experience more frequent effects of conspecifics than heterospecifics.

302 To see this, let us first focus on the smallest distances between organisms. The nearest neighbour of an organism  
303 was always an organism of the same species, and the minimum distance between conspecifics was always lower than  
304 expected for a uniform distribution (Section S6 of the SI). The dominance index remained close to 1 for distances  
305 below  $10^{-2}$  cm or  $10^{-3}$  cm for microphytoplankton and nanophytoplankton respectively. There was therefore always  
306 *some* intraspecific aggregation, i.e. conspecifics were always closer than heterospecifics at the smallest distances.  
307 This is due to the prevalence of demographic processes at individual scales, because an individual acts as a source  
308 point for other organisms of the same species, and hydrodynamic processes do not separate conspecifics fast enough  
309 to prevent aggregation. If we consider that interaction strengths are a smoothly decaying function of distance, a  
310 common assumption in spatial coexistence models (e.g., Bolker & Pacala, 1999; Law *et al.*, 2003), this implies  
311 that population-level intraspecific interactions could be stronger than interspecific interactions due to intraspecific  
312 micro-scale aggregation. However, the mechanisms of competition at this scale are poorly known, likely relying on  
313 multiple types of resources with different distributions in the environment, effects on the cell, uptakes, etc. Rather  
314 than weighting much more heavily the potential interactions with the closest neighbour(s) through an interaction  
315 kernel, we therefore chose conservatively to define a maximum distance for two organisms to possibly affect the  
316 concentrations of elements in the environment of each other. We consider that, at all distances below this threshold,

317 interactions could happen between organisms. We continue the discussion with that simplification in mind, and  
318 explicitly mention when it is relaxed.

319 Dominance indices began to decrease at distances above  $10^{-3}$  cm, still below the maximum distance for inter-  
320 actions. At this distance and above, the balance between heterospecifics and conspecifics was much more sensitive  
321 to different phytoplankters' demographic and hydrodynamic traits. The species composition of an organism's  
322 neighbourhood depended on its size: nanophytoplankton organisms mainly shared their volume of potential in-  
323 teractions with conspecifics (the dominance index remained close to 1, even near the distance threshold, i.e. the  
324 maximum distance for the overlap of nutrient depletion volumes) while microphytoplankton organisms could affect  
325 both conspecifics and heterospecifics (the dominance index was often below 0.5 at the distance threshold, i.e. an  
326 individual's depletion zone probably overlapped with more heterospecifics' than conspecifics'). Microphytoplankters  
327 were therefore more likely to share their depletion volume with heterospecifics than nanophytoplankters. The rate  
328 of production of new microphytoplankton conspecifics was not sufficient to compensate for the mixing induced by  
329 turbulence and diffusivity, even though the diffusivity of microphytoplankters was smaller than that of nanophy-  
330 toplankters. There may therefore be different mechanisms at play at the community level for microphytoplankton  
331 and nanophytoplankton to maintain coexistence. For nanophytoplankton, the spatial structure likely leads to more  
332 interactions between conspecifics than between heterospecifics. The spatial distribution of microphytoplankton  
333 species, on the contrary, encourages more interactions between heterospecifics. If we consider that local interaction  
334 strengths are equal within the volume of potential interactions, scaling to the population level, we would likely  
335 observe stronger intra- over interspecific interactions for nanophytoplankton (a key factor in coexistence theory,  
336 Barabás *et al.*, 2017) but not necessarily so for microphytoplankton. Using a timescale separation argument, we  
337 show in Section S8 in the SI how stronger interactions at population level than individual level may arise in a  
338 Lotka-Volterra model whose spatial structure is summed up by the dominance indices evidenced here. Stronger  
339 intra- than interspecific competition may arise at population level even when assuming that all local interaction  
340 strengths between individuals are equal, regardless of the identity of competitors.

341 All of the above discussion is based on a microphytoplankter's neighbourhood in its nutrient depletion volume.  
342 To simplify the computation, we used maximum volumes of potential interactions, corresponding to a diffusive-only  
343 flow of nutrient particles. But when fluid turbulence increases, nutrient uptake increases, and the size of the depletion  
344 zone decreases (Karp-Boss *et al.*, 1996). The proportion of change in the depletion volume increases with the size  
345 of organisms: a 10  $\mu\text{m}$ -diameter organism might not experience any change, while the uptake of a 100  $\mu\text{m}$ -diameter  
346 organism would increase by at least 50% (Karp-Boss *et al.*, 1996). Therefore the volume of potential interactions  
347 shrinks in the presence of turbulence for microphytoplankton, but not necessarily for nanophytoplankton. This  
348 could be one additional reason why microphytoplankters might still be surrounded by conspecifics at ecologically  
349 meaningful distances and interacting more frequently with them.

350 Up to now, we have only focused on the dominance index, a localized proportion of conspecifics. However,

interactions also depend on the absolute densities of individuals. Mechanically, when density decreases, the distances between neighbours increase, which explains that the distances between the low-abundance microphytoplankters tended to be greater than distances between the more abundant nanophytoplankters (Section S6 of the SI). Explicit mathematical models using pair densities to express interaction rates (e.g. Law *et al.*, 2003; Plank & Law, 2015) may be able to incorporate those effects; however, as we highlight below, the timescales and spatial correlations that are seen in such models may not necessarily represent faithfully phytoplankton community dynamics.

Contrary to other similar models (e.g., Birch & Young, 2006; Bouderbala *et al.*, 2018), we did not consider explicit effects of local density on survival and fertility rates. Outside of simply maintaining analytical tractability, we had another, more biological reason to do so: we cannot be sure that these local density-dependencies make sense in our phytoplankton context. To understand why, consider that even if a species abundance is locally tripled, competition might not directly ensue at the time scales covered by our model ( $\approx 7$  h), if nutrient depletion has not had time to set in yet. Even if we considered longer time frames, we would need lagged local density-dependencies, which are to our knowledge not leading to tractable spatial branching or dynamic point processes. We could, of course, directly model nutrients, perhaps as resource “points” with a dynamics of their own (Murrell, 2005; North & Ovaskainen, 2007), which in turn change the reproduction or death rate of individuals. If the resource points risk being depleted, this entails a negative spatial correlation between organisms and their resources (Murrell, 2005; Barraquand & Murrell, 2012). And that is where such models might be inadequate. The phycosphere, a micro-environment at the periphery of a phytoplankton organism where communities of bacteria interact (Seymour *et al.*, 2017), can also impact phytoplankton fitness, both positively (cross-feeding) and negatively (algicidal activities of bacteria). This can sometimes lead to an accumulation of key resources close to the phytoplankter. This will lead to positive spatial correlations between consumers and their resources, and we currently do not have theoretical models to represent this process (short of modelling precisely the spatial distribution of these bacteria).

Our model should be viewed as a first model of spatial distributions of multiple phytoplankton species in a realistic, three-dimensional environment at the microscale, describing only basic hydrodynamic and demographic processes. Using this model, we were able to predict whether phytoplankters could be in contact with individuals of their own or other species, and emit reasonable conjectures regarding potential intra vs interspecific interactions between species, emerging at the population level through spatial distributions (Detto & Muller-Landau, 2016). It is worthwhile to keep in mind that there are many remaining features of phytoplankton physiology and life histories which we do not address here, but which may affect spatial distributions. Many phytoplankters are able to move actively in three dimensions, which can favour cluster formation (Breier *et al.*, 2018). Even those who are believed to move passively actually often move along the vertical dimension by regulating their buoyancy (Reynolds, 2006), and can at times aggregate to form pairs (Font-Muñoz *et al.*, 2019). Finally, a part of spatial structure is explained by the partially colonial nature of microphytoplankton (Kiørboe *et al.*, 1990). This clearly calls for viewing our model as a null model to which more complex mechanistic models and their spatial outputs can be compared.

385 **Acknowledgments**

386 FB and CP were supported by the grant ANR-20-CE45-0004. CP was supported by a PhD grant from the French  
 387 Ministry of Research.

388 **Appendices**

389 **Derivation of the spatial characteristics of the Brownian Bug Model**

390 We show here how to compute the monospecific pair correlation function and Ripley's  $K$ -function of the Brownian  
 391 Bug Model (see Young *et al.*, 2001 and Picoche *et al.*, 2022 for a detailed derivation of the master equation). As  
 392 these formula only apply to intraspecies pairs, we ignore species' index in the following for the sake of clarity. Similar  
 393 formula for well-known spatial point processes are given in the Supplementary Information, for readers who want  
 394 to understand better the properties of these spatial statistics.

395 **Proof of Eq. 4 and Eq. 5**

396 In three dimensions, when the birth rate  $\lambda$  is the same as the mortality rate  $\mu$ , the pair density  $G(r)$  is a solution of

$$\frac{\partial G}{\partial t} = \frac{2D}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial G}{\partial r} \right) + \frac{\gamma}{r^2} \frac{\partial}{\partial r} \left( r^4 \frac{\partial G}{\partial r} \right) + 2\lambda C \delta(\xi). \quad (18)$$

**Steady-state solution** We first compute the steady-state solution, *i.e.*

$$\begin{aligned} 0 &= \frac{2D}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial G}{\partial r} \right) + \frac{\gamma}{r^2} \frac{\partial}{\partial r} \left( r^4 \frac{\partial G}{\partial r} \right) + 2\lambda C \delta(\xi) \\ 0 &= 4\pi r^2 \left( \frac{2D}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial G}{\partial r} \right) + \frac{\gamma}{r^2} \frac{\partial}{\partial r} \left( r^4 \frac{\partial G}{\partial r} \right) + 2\lambda C \delta(\xi) \right) \\ 0 &= 4\pi \left( 2D \frac{\partial}{\partial r} \left( r^2 \frac{\partial G}{\partial r} \right) + \gamma \frac{\partial}{\partial r} \left( r^4 \frac{\partial G}{\partial r} \right) \right) + 4\pi r^2 2\lambda C \delta(\xi). \end{aligned} \quad (19)$$

We can then integrate Eq. 18 over a small sphere centered on an individual, with radius  $\rho$ . Let us first note that

$$\begin{aligned} &\int_{\mathbb{R}^3} \delta(\xi) d\xi = 1 \\ \Leftrightarrow &\int_0^{2\pi} \int_0^\pi \int_0^\rho \delta(r) \delta(\phi) \delta(\theta) r^2 \sin(\phi) dr d\phi d\theta = 1 \\ \Leftrightarrow &4\pi \int_0^\rho \delta(r) r^2 dr = 1. \end{aligned} \quad (20)$$

Using Eq. 19 and 20,

$$\begin{aligned} 0 &= 4\pi \left( 2Dr^2 \frac{\partial G}{\partial r} + \gamma r^4 \frac{\partial G}{\partial r} \right) + 2\lambda C \\ \Leftrightarrow \frac{\partial G}{\partial r} &= -\frac{1}{4\pi} \frac{2\lambda C}{2Dr^2 + \gamma r^4}. \end{aligned} \quad (21)$$

We can integrate Eq. 21 between  $\rho$  and  $\infty$ . As  $G(\infty) = C^2$ ,

$$C^2 - G(\rho) = -\frac{\lambda C}{2\pi} \int_{\rho}^{\infty} \frac{1}{2Dr^2 + \gamma r^4} dr. \quad (22)$$

We first compute the primitive  $A = \int \frac{1}{2Dr^2 + \gamma r^4} dr$ .

$$A = \int \frac{1}{r^2(2D + \gamma r^2)} dr \quad (23)$$

$$= \int \frac{1}{2Dr^2} - \frac{\gamma}{2D(2D + \gamma r^2)} dr \quad (24)$$

$$= -\frac{1}{2Dr} - \frac{\gamma}{2D} \int \frac{1}{2D \left( 1 + (\sqrt{\frac{\gamma}{2D}} r)^2 \right)} dr. \quad (25)$$

<sup>397</sup> With a change of variable  $u = \sqrt{\frac{\gamma}{2D}} r$ , using  $\int \frac{1}{1+u^2} = \arctan(u)$ , we have

$$A = -\frac{1}{2Dr} - \frac{\sqrt{\gamma} \arctan \left( \frac{\sqrt{\gamma}r}{\sqrt{2D}} \right)}{2\sqrt{2}D\sqrt{D}} + K \quad (26)$$

<sup>398</sup> where  $K$  is a constant. We can now compute  $B = [A]_{\rho}^{\infty}$ .

$$B = -\frac{\sqrt{\gamma}\pi}{4\sqrt{2}D\sqrt{D}} + \frac{1}{2D\rho} + \frac{\sqrt{\gamma} \arctan \left( \frac{\sqrt{\gamma}\rho}{\sqrt{2D}} \right)}{2\sqrt{2}D\sqrt{D}}. \quad (27)$$

This leads to

$$G(\rho) = C^2 + \frac{\lambda C}{2\pi} B \quad (28)$$

$$= C^2 + \frac{\lambda C}{2\pi} \left[ \frac{1}{2D\rho} + \frac{\sqrt{\gamma} \arctan \left( \frac{\sqrt{\gamma}\rho}{\sqrt{2D}} \right)}{2\sqrt{2}D\sqrt{D}} - \frac{\sqrt{\gamma}\pi}{4\sqrt{2}D\sqrt{D}} \right]. \quad (29)$$

<sup>399</sup> Finally, the pair correlation function  $g = G/C^2$  is defined as

$$g(\rho) = \frac{\lambda}{4\pi CD} \left( \frac{\sqrt{\gamma} \arctan \left( \frac{\sqrt{\gamma}\rho}{\sqrt{2D}} \right)}{\sqrt{2D}} + \frac{1}{\rho} - \frac{\pi\sqrt{\gamma}}{2\sqrt{2D}} \right) + 1. \quad (30)$$

400 **Time-dependent solution** In the absence of advection by turbulent diffusion ( $U = 0, \gamma = 0$ ), convergence to the  
 401 steady-state solution can be very slow (more than a week, see Section S3 in the SI). In order to keep a realistic  
 402 timeframe, we need to compute a time-dependent solution. We can get back to Eq. 18 with  $\gamma = 0$ , which yields

$$\frac{\partial G}{\partial t} = \frac{2D}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial G}{\partial r} \right) + 2\lambda C \delta(\xi). \quad (31)$$

403 Assuming an isotropic environment, this means

$$\frac{\partial G}{\partial t} - 2D\Delta G = 2\lambda C \delta(\xi) \quad (32)$$

404 where  $\Delta = \nabla^2$  is the Laplacian operator. We therefore have

$$\mathcal{L}G(\xi, t) = 2\lambda C \delta(\xi) \quad (33)$$

405 where  $\mathcal{L}$  is the linear differential operator  $\partial_t - 2D\Delta$ .

406 Using the Green's function theory, we know that  $G(y) = \int H(y, s) 2\lambda C \delta(s) ds$  where  $H(y, s) = H(y - s)$  is the  
 407 Green kernel (heat kernel). We can therefore write

$$\begin{aligned} G(\xi, t) &= 2\lambda C \int_{\mathbb{R}^3} \int_0^t H(\xi - \xi', t') \delta(\xi') d\xi' dt' \\ \Leftrightarrow G(\xi, t) &= 2\lambda C \int_0^t H(\xi, t') dt'. \end{aligned} \quad (34)$$

408 A solution for the Green's function using  $\mathcal{L} = \partial_t - 2D\Delta$  in three dimensions is  $H(r, t) = \left(\frac{1}{8\pi Dt}\right)^{3/2} \exp\left(-\frac{r^2}{8Dt}\right)$ .

409  $G(r, t)$  can then be computed as

$$G(r, t) = 2\lambda C \left( \frac{-\text{erf}\left(\frac{r}{\sqrt{8Dt}}\right)}{8\pi Dr} + K \right) \quad (35)$$

410 where  $\text{erf}$  is the error function. Using  $G(r, 0) = C^2$  and  $\lim_{x \rightarrow +\infty} \text{erf}(x) = 1$  in Eq. 35,

$$\begin{aligned} C^2 &= 2\lambda C \left( \frac{1}{8\pi Dr} + K \right) \\ \Leftrightarrow \frac{C}{2\lambda} - \frac{1}{8\pi Dr} &= K. \end{aligned} \quad (36)$$

411 We can finally compute  $G(r, t)$ :

$$\begin{aligned} G(r, t) &= 2\lambda C \left( -\frac{\text{erf}\left(\frac{r}{\sqrt{8Dt}}\right)}{8\pi Dr} + \frac{C}{2\lambda} + \frac{1}{8D\pi r} \right) \\ &= \frac{\lambda C}{4\pi Dr} \left\{ 1 - \text{erf}\left(\frac{r}{\sqrt{8Dt}}\right) \right\} + C^2 \\ \Leftrightarrow g(r, t) &= \frac{\lambda}{4D\pi r C} \left\{ 1 - \text{erf}\left(\frac{r}{\sqrt{8Dt}}\right) \right\} + 1. \end{aligned} \quad (37)$$

<sup>412</sup> **Proof of Eq. 8 and Eq. 9**

<sup>413</sup> We can integrate the pcf formula to compute Ripley's  $K$ -function, as  $g(r) = \frac{K'(r)}{4\pi r^2}$ .

<sup>414</sup> **Steady-state solution** From Eq. 30,

$$K(\rho) = 4\pi \int_0^\rho r^2 + \frac{\lambda}{2\pi C} \left[ \frac{r}{2D} + \frac{\sqrt{\gamma}r^2 \arctan\left(\frac{\sqrt{\gamma}r}{\sqrt{2D}}\right)}{2\sqrt{2D}\sqrt{D}} - \frac{\sqrt{\gamma}\pi r^2}{4\sqrt{2D}\sqrt{D}} \right] dr. \quad (38)$$

<sup>415</sup> We define  $A = \int_0^\rho r^2 dr$ ,  $B = \int_0^\rho \frac{r}{2D} dr$ ,  $C = \int_0^\rho r^2 \arctan\left(\frac{\sqrt{\gamma}r}{\sqrt{2D}}\right) dr$  and  $E = \int_0^\rho \frac{\sqrt{\gamma}\pi r^2}{4\sqrt{2D}\sqrt{D}} dr$ .

$$\begin{aligned} A &= \frac{1}{3}\rho^3. \\ B &= \frac{\rho^2}{4D}. \\ E &= \frac{\sqrt{\gamma}\pi\rho^3}{12\sqrt{2D}\sqrt{D}}. \end{aligned} \quad (39)$$

<sup>416</sup> We can also compute  $C = \int_0^\rho r^2 \arctan\left(\frac{\sqrt{\gamma}r}{\sqrt{2D}}\right) dr$ . We first change variable, with  $u = \frac{r}{\sqrt{2D}}$ ,  $dr = \sqrt{2D}du$ , and  
<sup>417</sup> obtain

$$C = (2D)^{3/2} \int_0^{\rho/\sqrt{2D}} u^2 \arctan(\sqrt{\gamma}u) du. \quad (40)$$

<sup>418</sup> We can integrate by parts, with  $f = \arctan(\sqrt{\gamma}u)$  and  $g' = u^2$ , which leads to

$$C = (2D)^{3/2} \left( \frac{\rho^3}{3(2D)^{3/2}} \arctan\left(\sqrt{\frac{\gamma}{2D}}\rho\right) - \frac{\sqrt{\gamma}}{3} \int_0^{\rho/\sqrt{2D}} \frac{u^3}{(\gamma u^2 + 1)} du \right). \quad (41)$$

<sup>419</sup> We then substitute  $v = \gamma u^2 + 1$ ,  $du = \frac{1}{2\gamma u} dv$ , and have

$$\begin{aligned} \int_0^{\rho/\sqrt{2D}} \frac{u^3}{(\gamma u^2 + 1)} du &= \frac{1}{2\gamma^2} \int_1^{\gamma\rho^2/2D+1} \frac{v-1}{v} dv \\ &= \frac{1}{2\gamma^2} \int_1^{\gamma\rho^2/2D+1} 1 - \frac{1}{v} dv \\ &= \frac{1}{2\gamma^2} \left( \gamma \frac{\rho^2}{2D} - \log\left(\gamma \frac{\rho^2}{2D} + 1\right) \right). \end{aligned} \quad (42)$$

<sup>420</sup> Going back to C, we obtain

$$\begin{aligned} C &= \frac{\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{3} - (2D)^{3/2} \frac{\sqrt{\gamma}}{3} \frac{1}{2\gamma^2} \left( \frac{\gamma}{2D} \rho^2 - \log\left(\gamma \frac{\rho^2}{2D} + 1\right) \right) \\ &= \frac{\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{3} - \frac{\sqrt{2D}}{6\sqrt{\gamma}} \rho^2 + \frac{\sqrt{2D}^{3/2}}{3\gamma^{3/2}} \log\left(\gamma \frac{\rho^2}{2D} + 1\right). \end{aligned} \quad (43)$$

<sup>421</sup> Combining all equations,

$$\begin{aligned} K(\rho) &= \frac{4}{3}\pi\rho^3 + \frac{2\lambda}{C} \left( \frac{\rho^2}{4D} + \frac{\sqrt{\gamma}\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{6\sqrt{2}D^{3/2}} - \frac{\rho^2}{12D} + \frac{\log\left(\gamma\frac{\rho^2}{2D}+1\right)}{6\gamma} - \frac{\sqrt{\gamma}\pi\rho^3}{12\sqrt{2}D\sqrt{D}} \right) \\ &= \frac{4}{3}\pi\rho^3 + \frac{\lambda}{3C} \left( \frac{\rho^2}{D} + \frac{\sqrt{\gamma}\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{\sqrt{2}D^{3/2}} + \frac{\log\left(\gamma\frac{\rho^2}{2D}+1\right)}{\gamma} - \frac{\sqrt{\gamma}\pi\rho^3}{2\sqrt{2}D\sqrt{D}} \right). \end{aligned} \quad (44)$$

<sup>422</sup> Note that in the absence of advection,

$$\begin{aligned} g(r) &= \frac{\lambda}{4\pi CD r} + 1 \\ \Rightarrow K'(r) &= \frac{\lambda r}{CD} + 4\pi r^2 \\ \Leftrightarrow K(r) &= \frac{\lambda r^2}{2CD} + \frac{4}{3}\pi r^3. \end{aligned} \quad (45)$$

<sup>423</sup> **Time-dependent solution** In the absence of advection ( $U = 0, \gamma = 0$ ), we need to compute a time-dependent solution. From eq. 37,

$$\begin{aligned} K(\rho) &= \frac{\lambda}{DC} \int_0^\rho r \left\{ 1 - \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right) \right\} + 4\pi r^2 dr \\ &= \frac{\lambda}{CD} \left( \frac{\rho^2}{2} - \int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right) dr \right) + \frac{4}{3}\pi\rho^3. \end{aligned} \quad (46)$$

<sup>425</sup> We first compute the primitive for  $\int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right) dr$ . We define  $u = \frac{r}{\sqrt{8Dt}}$ ,  $dr = \sqrt{8Dt}du$ , then

$$\int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right) dr = 8Dt \int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u) du. \quad (47)$$

<sup>426</sup> We can integrate by parts, with  $f = \operatorname{erf}(u)$  and  $g' = u$ , and obtain

$$8Dt \int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u) du = 8Dt \left( \frac{\rho^2}{2} \frac{1}{8Dt} \operatorname{erf}\left(\frac{\rho}{\sqrt{8Dt}}\right) - \frac{1}{\sqrt{\pi}} \int_0^{\rho/\sqrt{8Dt}} u^2 e^{-u^2} du \right). \quad (48)$$

<sup>427</sup> We integrate by parts again, this time with  $f = u$  and  $g' = ue^{-u^2}$ , which leads to

$$\int u^2 e^{-u^2} du = -\frac{ue^{-u^2}}{2} + \frac{1}{2} \int e^{-u^2} du = -\frac{ue^{-u^2}}{2} + \frac{\sqrt{\pi} \operatorname{erf}(u)}{4}. \quad (49)$$

<sup>428</sup> If we use Eq. 49 in Eq. 48,

$$\begin{aligned} 8Dt \int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u) du &= 8Dt \left( \frac{\rho^2}{2} \frac{1}{8Dt} \operatorname{erf}\left(\frac{\rho}{\sqrt{8Dt}}\right) - \frac{\operatorname{erf}\left(\frac{\rho}{\sqrt{8Dt}}\right)}{4} + \frac{1}{2\sqrt{\pi}} \frac{\rho}{\sqrt{8Dt}} e^{-\frac{\rho^2}{8Dt}} \right) \\ \Leftrightarrow \int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right) dr &= \frac{1}{2} \operatorname{erf}\left(\frac{\rho}{\sqrt{8Dt}}\right) (\rho^2 - 4Dt) + \frac{\sqrt{2Dt}}{\sqrt{\pi}} \rho e^{-\frac{\rho^2}{8Dt}}. \end{aligned} \quad (50)$$

<sup>429</sup> We can now compute  $K(\rho)$ :

$$K(\rho) = \frac{\lambda}{CD} \left( \frac{\rho^2}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\rho}{\sqrt{8Dt}}\right) (\rho^2 - 4Dt) - \frac{\sqrt{2Dt}\rho}{\sqrt{\pi}} e^{-\frac{\rho^2}{8Dt}} \right) + \frac{4}{3}\pi\rho^3. \quad (51)$$

430 **References**

- 431 Adler, P.B., Smull, D., Beard, K.H., Choi, R.T., Furniss, T., Kulmatiski, A., Meiners, J.M., Tredennick, A.T. &  
432 Veblen, K.E. (2018). Competition and coexistence in plant communities: intraspecific competition is stronger  
433 than interspecific competition. *Ecology Letters*, 21, 1319–1329.
- 434 Arnott, R.N., Cherif, M., Bryant, L.D. & Wain, D.J. (2021). Artificially generated turbulence: a review of phyco-  
435 logical nanocosm, microcosm, and mesocosm experiments. *Hydrobiologia*, 848, 961–991.
- 436 Arrieta, J., Jeanneret, R., Roig, P. & Tuval, I. (2020). On the fate of sinking diatoms: the transport of active  
437 buoyancy-regulating cells in the ocean. *Philosophical Transactions of the Royal Society A*, 378, 20190529.
- 438 Bainbridge, R. (1957). The size, shape and density of marine phytoplankton concentrations. *Biological Reviews*,  
439 32, 91–115.
- 440 Barabás, G., Michalska-Smith, M.J. & Allesina, S. (2017). Self-regulation and the stability of large ecological  
441 networks. *Nature Ecology & Evolution*, 1, 1870–1875.
- 442 Barraquand, F. & Murrell, D.J. (2012). Evolutionarily stable consumer home range size in relation to resource  
443 demography and consumer spatial organization. *Theoretical Ecology*, 5, 567–589.
- 444 Barton, A.D., Ward, B.A., Williams, R.G. & Follows, M.J. (2014). The impact of fine-scale turbulence on phyto-  
445 plankton community structure. *Limnology and Oceanography: Fluids and Environments*, 4, 34–49.
- 446 Bec, B., Collos, Y., Vaquer, A., Mouillot, D. & Souchu, P. (2008). Growth rate peaks at intermediate cell size in  
447 marine photosynthetic picoeukaryotes. *Limnology and Oceanography*, 53, 863–867.
- 448 Benczik, I.J., Károlyi, G., Scheuring, I. & Tél, T. (2006). Coexistence of inertial competitors in chaotic flows.  
449 *Chaos*, 16, 043110.
- 450 Birch, D.A. & Young, W.R. (2006). A master equation for a spatial population model with pair interactions.  
451 *Theoretical Population Biology*, 70, 26–42.
- 452 Bissinger, J.E., Montagnes, D.J.S., Harples, J. & Atkinson, D. (2008). Predicting marine phytoplankton maxi-  
453 mum growth rates from temperature: improving on the Eppley curve using quantile regression. *Limnology and  
454 Oceanography*, 53, 487–493.
- 455 Bolker, B.M. & Pacala, S.W. (1999). Spatial moment equations for plant competition: understanding spatial  
456 strategies and the advantages of short dispersal. *The American Naturalist*, 153, 575–602.
- 457 Borgnino, M., Arrieta, J., Boffetta, G., De Lillo, F. & Tuval, I. (2019). Turbulence induces clustering and segregation  
458 of non-motile, buoyancy-regulating phytoplankton. *Journal of the Royal Society Interface*, 16, 20190324.

- 459 Bouderbala, I., El Saadi, N., Bah, A. & Auger, P. (2018). A 3D individual-based model to study effects of chemotaxis,  
460 competition and diffusion on the motile-phytoplankton aggregation. *Acta Biotheoretica*, 66, 257–278.
- 461 Breier, R.E., Lalescu, C.C., Waas, D., Wilczek, M. & Mazza, M.G. (2018). Emergence of phytoplankton patchiness  
462 at small scales in mild turbulence. *Proceedings of the National Academy of Sciences*, 115, 12112–12117.
- 463 Chesson, P. (2018). Updates on mechanisms of maintenance of species diversity. *Journal of Ecology*, 106, 1773–1794.
- 464 Detto, M. & Muller-Landau, H.C. (2016). Stabilization of species coexistence in spatial models through the  
465 aggregation-segregation effect generated by local dispersal and nonspecific local interactions. *Theoretical Popu-*  
466 *lation Biology*, 112, 97–108.
- 467 Doubell, M.J., Seuront, L., Seymour, J.R., Patten, N.L. & Mitchell, J.G. (2006). High-resolution fluorometer for  
468 mapping microscale phytoplankton distributions. *Applied and Environmental Microbiology*, 72, 4475–4478.
- 469 Dusenberry, D. (2009). *Living at the microscale*. Harvard University Press.
- 470 Edwards, K.F. (2019). Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of*  
471 *the National Academy of Sciences*, p. 201814860.
- 472 Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden  
473 flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4.
- 474 Estrada, M., Alcaraz, M. & Marrasé, C. (1987). Effects of turbulence on the composition of phytoplankton assem-  
475 blages in marine microcosms. *Marine Ecology Progress Series*, 38, 267–281.
- 476 Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. (1998). Primary production of the biosphere:  
477 integrating terrestrial and oceanic components. *Science*, 281, 237–240.
- 478 Font-Muñoz, J.S., Jeanneret, R., Arrieta, J., Anglès, S., Jordi, A., Tuval, I. & Basterretxea, G. (2019). Collective  
479 sinking promotes selective cell pairing in planktonic pennate diatoms. *Proceedings of the National Academy of*  
480 *Sciences*, 116, 15997–16002.
- 481 Font-Muñoz, J.S., Jordi, A., Tuval, I., Arrieta, J., Anglès, S. & Basterretxea, G. (2017). Advection by ocean  
482 currents modifies phytoplankton size structure. *Journal of the Royal Society Interface*, 14, 20170046.
- 483 Haegeman, B. & Rapaport, A. (2008). How flocculation can explain coexistence in the chemostat. *Journal of*  
484 *Biological Dynamics*, 2, 1–13.
- 485 Hellweger, F.L. & Bucci, V. (2009). A bunch of tiny individuals – individual-based modeling for microbes. *Ecological*  
486 *Modelling*, 220, 8–22.

- 487 Huisman, J. & Weissing, F.J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402,  
488 407–410.
- 489 Hutchinson, G.E. (1961). The paradox of the plankton. *The American Naturalist*, 95, 137–145.
- 490 Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*.  
491 vol. 70. John Wiley & Sons.
- 492 Jumars, P.A., Deming, J., Hill, P., Karp-Boss, L., Yager, P. & Dade, W. (1993). Physical constraints on marine  
493 osmotrophy in an optimal foraging context. *Marine Microbial Food Webs*, 7, 121–159.
- 494 Karp-Boss, L., Boss, E. & Jumars, P.A. (1996). Nutrient fluxes to planktonic osmotrophs in the presence of fluid  
495 motion. *Oceanography and Marine Biology: An Annual Review*, 34, 71–107.
- 496 Kiørboe, T., Andersen, K.P. & Dam, H.G. (1990). Coagulation efficiency and aggregate formation in marine  
497 phytoplankton. *Marine Biology*, 107, 235–245.
- 498 Kraichnan, R.H. (1974). Convection of a passive scalar by a quasi-uniform random straining field. *Journal of Fluid  
499 Mechanics*, 64, 737–762.
- 500 Law, R., Murrell, D.J. & Dieckmann, U. (2003). Population growth in space and time: spatial logistic equations.  
501 *Ecology*, 84, 252–262.
- 502 Leonard, C.L., Bidigare, R.R., Seki, M.P. & Polovina, J.J. (2001). Interannual mesoscale physical and biological  
503 variability in the North Pacific Central Gyre. *Progress in Oceanography*, 49, 227–244.
- 504 Levine, J.M. & HilleRisLambers, J. (2009). The importance of niches for the maintenance of species diversity.  
505 *Nature*, 461, 254–257.
- 506 Li, L. & Chesson, P. (2016). The effects of dynamical rates on species coexistence in a variable environment: the  
507 paradox of the plankton revisited. *The American Naturalist*, 188, E46–E58.
- 508 MacArthur, R. & Levins, R. (1964). Competition, habitat selection and character displacement in a patchy envi-  
509 ronment. *Proceedings of the National Academy of Sciences*, 51, 1207–1210.
- 510 Martin, A.P. (2003). Phytoplankton patchiness: the role of lateral stirring and mixing. *Progress in Oceanography*,  
511 57, 125–174.
- 512 Murrell, D. (2005). Local spatial structure and predator-prey dynamics: counterintuitive effects of prey enrichment.  
513 *The American Naturalist*, 166, 354–367.
- 514 Ngan, K. & Vanneste, J. (2011). Scalar decay in a three-dimensional chaotic flow. *Physical Review E*, 83, 056306.

- 515 North, A. & Ovaskainen, O. (2007). Interactions between dispersal, competition, and landscape heterogeneity.  
516 *Oikos*, 116, 1106–1119.
- 517 Peters, F. & Marrasé, C. (2000). Effects of turbulence on plankton: an overview of experimental evidence and some  
518 theoretical considerations. *Marine Ecology Progress Series*, 205, 291–306.
- 519 Picoche, C. & Barraquand, F. (2019). How self-regulation, the storage effect, and their interaction contribute to  
520 coexistence in stochastic and seasonal environments. *Theor Ecol*, 12, 489–500.
- 521 Picoche, C. & Barraquand, F. (2020). Strong self-regulation and widespread facilitative interactions in phytoplank-  
522 ton communities. *Journal of Ecology*, 108, 2232–2242.
- 523 Picoche, C., Young, W.R. & Barraquand, F. (2022). [Re] Reproductive pair correlations and the clustering of  
524 organisms. *ReScience C*, 8.
- 525 Pierrehumbert, R.T. (1994). Tracer microstructure in the large-eddy dominated regime. *Chaos, Solitons & Fractals*,  
526 4, 1091–1110.
- 527 Plank, M.J. & Law, R. (2015). Spatial point processes and moment dynamics in the life sciences: a parsimonious  
528 derivation and some extensions. *Bulletin of Mathematical Biology*, 77, 586–613.
- 529 Prairie, J.C., Sutherland, K.R., Nickols, K.J. & Kaltenberg, A.M. (2012). Biophysical interactions in the plankton:  
530 a cross-scale review. *Limnology and Oceanography: Fluids and Environments*, 2, 121–145.
- 531 Record, N.R., Pershing, A.J. & Maps, F. (2014). The paradox of the “paradox of the plankton”. *ICES Journal of  
532 Marine Science*, 71, 236–240.
- 533 REPHY (2017). *REPHY dataset - French Observation and Monitoring program for Phytoplankton and Hydrology  
534 in coastal waters. 1987-2016 Metropolitan data.* <https://www.seanoe.org/data/00361/47248/>.
- 535 Reynolds, C.S. (2006). *The ecology of phytoplankton*. Cambridge University Press.
- 536 Schippers, P., Verschoor, A.M., Vos, M. & Mooij, W.M. (2001). Does “supersaturated coexistence” resolve the  
537 “paradox of the plankton”? *Ecology Letters*, 4, 404–407.
- 538 Seymour, J.R., Amin, S.A., Raina, J.B. & Stocker, R. (2017). Zooming in on the phycosphere: the ecological  
539 interface for phytoplankton–bacteria relationships. *Nature Microbiology*, 2, 17065.
- 540 Stocker, R. (2012). Marine microbes see a sea of gradients. *Science*, 338, 628–633.
- 541 Widdicombe, C. & Harbour, D. (2021). Phytoplankton taxonomic abundance and biomass time-series at Plymouth  
542 Station L4 in the Western English Channel, 1992-2020.

- 543 Wiegand, T., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N. & Huth, A. (2007). How individual species structure  
544 diversity in tropical forests. *Proceedings of the National Academy of Sciences*, 104, 19029–19033.
- 545 Young, W.R., Roberts, A.J. & Stuhne, G. (2001). Reproductive pair correlations and the clustering of organisms.  
546 *Nature*, 412, 328–331.