**RESCIENCE C**

Replication / Ecology

[Re] Reproductive pair correlations and the clustering of organisms

Coralie Picoche[1] [ID], William R. Young[2], and Frédéric Barraquand[1] [ID]

[1]Institute of Mathematics of Bordeaux, CNRS & University of Bordeaux, Talence, France – [2]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0230, USA

Introduction

In the present work, we replicate the results of Young et al. 2001 "Reproductive pair correlations and the clustering of organisms" [1], an analysis of the formation of aggregates in an otherwise homogeneous environment mimicking marine small-scale hydrodynamics. Using an individual-based model of independent, random-walking particles (also called "Brownian bugs"), they show that reproduction by fission in a turbulent [2] and viscous flow leads to the formation of elongated clusters. Spatial patterns therefore depart from the usual, homogeneous solution of the advection-diffusion-reaction equation for a large population.

Due to their size, phytoplankton organisms experience a mostly viscous environment in a laminar shear field, with random but homogeneous changes in directions due to turbulence [3, 4]. Reproduction and limited movement of daughter cells, which occur at the phytoplankton scale, interact with these hydrodynamics processes and can lead to aggregates. In this context, a better understanding of the interactions between demography and small-scale hydrodynamics could provide further explanation for observed spatial distribution of phytoplankton species, and perhaps even their coexistence. This motivated us to revisit Young et al. 2001 [1].

In addition to replicating the numerical and mathematical results of Young et al. 2001, we also wished to present the mathematical derivations that were missing from the original paper, which should make this replication article more accessible to most readers, especially those without a fluid mechanics background.

Brownian bug model

The Brownian bug model is defined as an individual-based model in continuous space and time, here presented in its 2D formulation. For efficient computer simulation, it is implemented in discrete time [1]. Each particle is characterized by the vector of its Cartesian coordinates $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and its original position on the y-axis at $t = 0$ (a child particle inherits this attribute), this last characteristic being used only for representation purposes. Space is a $L \times L$ square with periodic boundary conditions. Each timestep, of duration τ, is divided into three substeps: (1) demographic processes, (2) diffusion,

and (3) advection.

Demographic processes take place during the first substep (1). Each organism has a fixed probability (p) of reproducing, dying (q), or remaining unchanged ($1 - p - q$). When an individual reproduces, a new organism appears on top of the parent. In the following, $p = q = 0.5$. Diffusion is then modeled as a Brownian motion (2), i.e. $\boldsymbol{x}'(t) = \boldsymbol{x}(t) + \delta\boldsymbol{x}(t)$ where each component of $\delta\boldsymbol{x}(t)$ follows a Gaussian distribution $\mathcal{N}(0, \Delta)$ where $D = \frac{\Delta^2}{2\tau}$ is the diffusivity. The discrete-time Markov chain presented here approximates the continuous-time Brownian bug model, which can be thought of as a spatial birth-death or branching process (described in the Supplementary Material), step (1) being referred to in Young et al. [1] as a Galton-Watson process. Finally, (3) the turbulent flow governing advective stirring follows the Pierrehumbert random map [5].

$$
\begin{align}
x_1(t + \tau) &= x_1'(t) + (U\tau/2)\cos[kx_2'(t) + \phi(t)] \tag{1}\\
x_2(t + \tau) &= x_2'(t) + (U\tau/2)\cos[kx_1(t + \tau) + \theta(t)] \tag{2}
\end{align}
$$

where $\phi(t)$ and $\theta(t)$ are random phases uniformly distributed between 0 and 2π, $k = 2\pi/L$ and U is the stretching parameter.

Unless otherwise specified, each simulation is initialized with $N_0 = 20,000$ particles uniformly distributed in a 1×1 square and run for 1000 timesteps.

Pair density function G(r,t)

The pair density function $G(\boldsymbol{x}_i, \boldsymbol{x}_j, t)$ is defined so that $G(\boldsymbol{x}_i, \boldsymbol{x}_j, t)dA_1 dA_2$ is the probability of finding a pair of Brownian bugs with one member in the area dA_1 around $\boldsymbol{x}_i$ and the other in the area dA_2 around $\boldsymbol{x}_j$. $G(\boldsymbol{r}, t)$ is actually called the pair correlation function in [1]. The radial density function $g(r, t)$ is defined as $G(\boldsymbol{x}_i, \boldsymbol{x}_j, t) = C^2 g(r, t)$ with $r = |\boldsymbol{x}_i - \boldsymbol{x}_j|$. As the pair correlation disappears when $r \to \infty$, $g \to 1$.

Derivation of G(r,t) − All details of the derivation of $G(r, t)$ are to be found in the Supplementary Material. We finally obtain:

$$
\frac{\partial G}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial G}{\partial r}\right) + 2(\lambda - \mu)G + \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{x}) \tag{3}
$$

where $\boldsymbol{x}$ is the position of the particle, λ is the birth rate and μ is the death rate.

We focus on the case $d = 2$ and $\lambda = \mu$, which means Eq. 3 can be reduced to

$$
\frac{\partial G}{\partial t} = \frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{x}) \tag{4}
$$

The value of γ is computed from simulations (see Supplementary Material).

Analytical solution with advection − CP and FB could only find the analytical solutions of $G(r, t)$ with and without advection with the indications of WY. In the presence of advection ($\gamma \neq 0$), a steady-state solution can be found; without advection, there is no steady-state and the solution changes through time. Let us first examine the steady-state solution, given by:

$$\frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{x}) \quad = \quad 0$$

$$\Leftrightarrow 2\pi r\left(\frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{x})\right) \quad = \quad 0$$

$$\Leftrightarrow 2\pi\left(2D\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \gamma\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right)\right) + 2\pi r 2\lambda C\delta(\boldsymbol{x}) \quad = \quad 0. \quad (5)$$

We can then integrate Eq. 5 over a small area centered on a particle, with radius ρ. Let us first note that

$$\int_{\mathbb{R}^2}\delta(\boldsymbol{x})d^2\boldsymbol{x} \quad = \quad 1$$

$$\Leftrightarrow \int_0^{2\pi}\int_0^\rho \delta(r')\delta(\theta)r'\,dr'\,d\theta \quad = \quad 1$$

$$\Leftrightarrow 2\pi\int_0^\rho \delta(\boldsymbol{x}')r'\,dr' \quad = \quad 1 \quad (6)$$

where $\boldsymbol{x}'$ is the equivalent of $\boldsymbol{x}$ in polar coordinates. Using Eq. 5 and 6, we can integrate between 0 and ρ,

$$0 \quad = \quad 2\pi\left(2D\rho\frac{\partial G}{\partial r} + \gamma\rho^3\frac{\partial G}{\partial r}\right) + 2\lambda C$$

$$\Leftrightarrow \quad \frac{\partial G}{\partial r} \quad = \quad -\frac{1}{2\pi}\frac{2\lambda C}{2D\rho + \gamma\rho^3}. \quad (7)$$

Eq. 7 can now be integrated between ρ and ∞, knowing that $G(\infty) = C^2$:

$$C^2 - G(\rho) = -\frac{1}{2\pi}\int_\rho^\infty \frac{2\lambda C}{2Dr + \gamma r^3}\,dr. \quad (8)$$

Using the variable change $u = 2Dr + \gamma r^3$, the integral is equivalent to $\int \frac{u'}{u}du$.

$$C^2 - G(\rho) \quad = \quad -\frac{\lambda C}{2\pi}\frac{1}{4D}\left[\log(\gamma) - \log\left(\frac{2D}{r^2} + \gamma\right)\right]_\rho^\infty \quad (9)$$

$$\Leftrightarrow G(\rho) \quad = \quad \frac{\lambda C}{8\pi D}\log\left(\frac{2D + \gamma\rho^2}{\gamma\rho^2}\right) + C^2 \quad (10)$$

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

$$g(r) = \frac{\lambda}{8\pi DC}\log\left(\frac{2D + \gamma r^2}{\gamma r^2}\right) + 1. \quad (11)$$

Analytical solution without advection — When $U = 0$, $\gamma = 0$ and there is no steady solution in 2D. We can get back to Eq. 4:

$$\frac{\partial G}{\partial t} = \frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{x}). \quad (12)$$

Assuming an isotropic environment (and switching to the Cartesian coordinate system), this means

$$\frac{\partial G}{\partial t} - 2D\Delta G = 2\lambda C\delta(\boldsymbol{x}) \tag{13}$$

where $\Delta = \nabla^2$ is the Laplacian operator.

We therefore have

$$\mathcal{L}G(\boldsymbol{x}, t) = 2\lambda C\delta(\boldsymbol{x}) \tag{14}$$

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

We can use a Green's function H, defined with $\mathcal{L}H = \delta(\boldsymbol{x}, t) = \delta(\boldsymbol{x})\delta(t)$.

By definition, we know that $G(y) = \int H(y, s)2\lambda C\delta(s)ds$ (where $y = (\boldsymbol{x}, t)$) is a solution to Eq. 14.

$$
\begin{aligned}
G(\boldsymbol{x}, t) \quad &= \quad 2\lambda C \int_{\mathbb{R}^2} \int_0^t H(\boldsymbol{x} - \boldsymbol{x}', t')\delta(\boldsymbol{x}')d^2\boldsymbol{x}'dt' \\
\Leftrightarrow \quad &= \quad 2\lambda C \int_0^t H(\boldsymbol{x}, t')dt' \tag{15}
\end{aligned}
$$

Eq. 15 can be used in Eq. 12:

$$
\frac{\partial}{\partial t}\left(2\lambda C \int_0^t H(\boldsymbol{x}, t')dt'\right) = 2D2\lambda C\Delta \int_0^t H(\boldsymbol{x}, t')dt' + 2\lambda C\delta(\boldsymbol{x}) \tag{16}
$$

$$
\Leftrightarrow \int_0^t \left(\frac{\partial H(\boldsymbol{x}, t')}{\partial t'} - 2D\Delta H(\boldsymbol{x}, t')\right)dt' = \delta(\boldsymbol{x}) \tag{17}
$$

$$
\Leftrightarrow \int_0^t \delta(\boldsymbol{x})\delta(t')dt' = \delta(\boldsymbol{x}) \tag{18}
$$

which is true.

A solution for the Green's function using $\mathcal{L} = \partial_t - 2D\Delta$ in 2 dimensions is

$$H(r, t) = \frac{1}{4\pi 2Dt}\exp\left(\frac{-r^2}{4 \times 2Dt}\right). \tag{19}$$

$G(r, t)$ can then be computed:

$$G(r, t) = 2\lambda C\left[\frac{E_1\left(\frac{r^2}{8Dt'}\right)}{8D\pi}\right]_0^t \tag{20}$$

where $E_1(x) = \int_x^\infty \frac{e^{-t}}{t}dt$ is the exponential integral. Using $G(r, 0) = C^2$ and $\lim_{x \to +\infty} E_1 = 0$ in Eq. 20, we finally obtain

$$G(r, t) = \lambda C\frac{E_1\left(\frac{r^2}{8Dt}\right)}{4D\pi} + C^2 \tag{21}$$

$$\Leftrightarrow g(r, t) = \frac{\lambda}{C}\frac{E_1\left(\frac{r^2}{8Dt}\right)}{4D\pi} + 1. \tag{22}$$

Results

We were able to reproduce the three figures of Young et al. [1] highlighting the spatial distributions of Brownian bugs.

In Fig. 1, the model has been run without the advection component and we can see the clumping of organisms due to reproduction. In Fig. 2 a), the model has been run without its demographic component, but with advection and diffusion, confirming that hydrodynamics alone cannot ensure cluster formation, while in Fig. 2 b), advection, diffusion and demography are present, in which case organisms form elongated aggregates.

Fig. 3 proved much more challenging. Retrieving the analytical solutions of Eq. 4 was difficult as there was no other equation than Eq. 3 in the original paper. We also encountered issues when computing the pair correlation functions on simulations: for large values of r/Δ, we observed zero values (absent pairs) of the pcf when $U = 0$. This is a sampling effect as pcf values get very low for large distances without advection, even though we multiplied the study area by 10 to produce Fig. 3 and counter such effects. Despite the missing values, we can confirm that simulated and analytical pcf match (Fig. 3), with a slight underestimation by the simulations. The numerical pcf also got closer to 1 than the analytical predictions for large values of r.
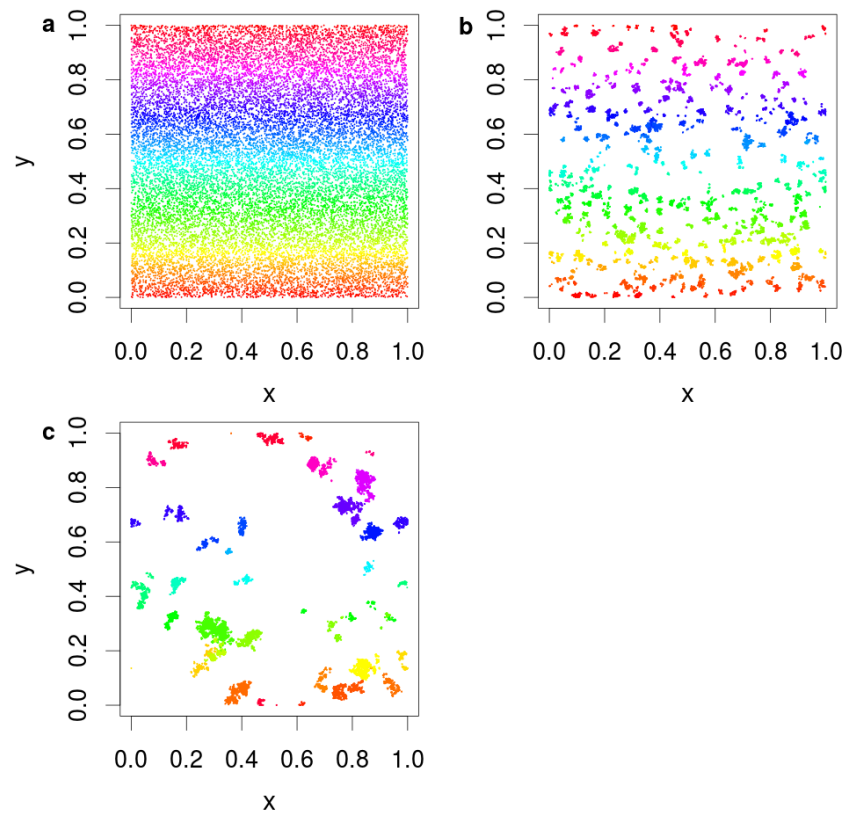
Figure 1. Distribution of Brownian bugs at different times in a simulation with $\Delta = 10^{-3}$ and $U = 0$: initial conditions with a Poisson spatial distribution (a), $t = 100\tau$ (b) and $t = 1000\tau$ (c). Each particle is identified by a color which corresponds to the initial position on the y-axis of its ancestor at $t = 0$.
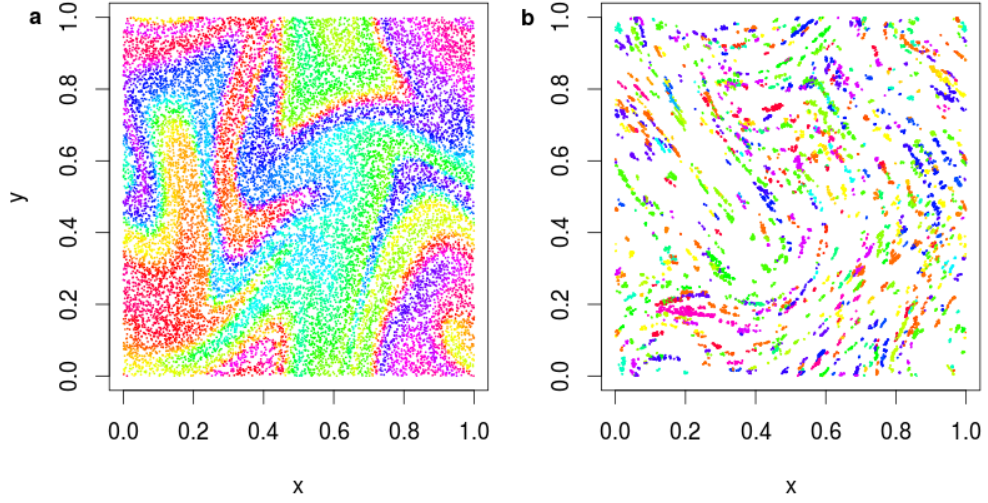
Figure 2. Distribution of Brownian bugs in a simulation with advection and $\Delta = 10^{-3}, U\tau/2 = 0.1$: without demographic processes at $t = 30\tau$ (a), and with demographic processes at $t = 1000\tau$ (b). Each particle is identified by a color which corresponds to the initial position on the y-axis of its ancestor at $t = 0$.
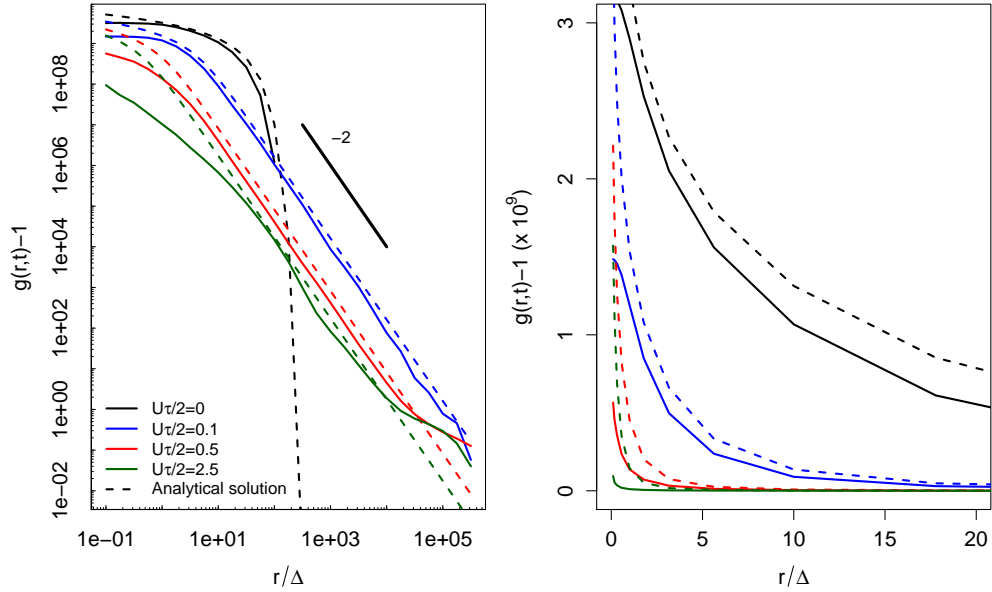
Figure 3. Logarithmic (a) and linear (b) plots of $g(r, t)$ versus r/Δ, with $\Delta = 10^{-7}$ and $U\tau/2 = 0, 0.1, 0.5, 2.5$ at $t = 1000\tau$. To compute simulation-based values of $g(r)$ with a large number of points ($N_0 = 200,000$) and avoid sampling issues, we replicated the 1×1 square 10 times, so that its length is $\sqrt{10}$ while keeping $L = 1$ and $k = 2\pi/L$ in eq. 1 and 2. Solid lines result from simulations, dotted lines correspond to analytical solutions and the solid grey line indicates the r^{-2} scaling predicted by Eq. 3.

Discussion

We successfully replicated both the numerical results and analytical solutions of Young et al. [1]. Even though stochasticity prevents us from replicating exactly the same spatial point patterns as those seen in the original Fig. 1 and Fig. 2, we considered the patterns to be close enough to validate the replication. Fig. 3 was also very close to the one shown in the original article, despite a slight underestimation of the pcf in simulated data.

The most challenging part of the replication was actually not to replicate the numerical results, but to find back the analytical expression of the pair density function $G(r, t)$ dynamics from first principles. How to derive such dynamics was indeed briefly explained in words in the original article, but the many intermediate steps involved (see Supplementary Material) make the additional mathematical derivations presented here worthwhile in our opinion. In addition to providing critical information to CP and FB regarding how to obtain the pair correlation dynamics, WY also communicated the required mathematical steps to find back the analytical solutions for $G(r, t)$ plotted in Fig. 3. We hope that the additional material on the derivation of $G(r, t)$ dynamics (as Supplementary Material) as well as the provided analytical solutions of such dynamics (now presented in the main text) will help readers through both the original and replication articles.

The original article did not provide quantitative values exactly matching marine microbes ecology; we thus wondered about the time and spatial scales that could be used for a realistic phytoplankton model. The length of the square side, L, is defined roughly as the Kolmogorov scale [2], the scale at which viscosity starts dominating turbulence (we use $L = 1$ cm as an upper bound). Here, we consider k the smallest wavenumber corresponding to the largest length scale L, i.e. $k = 2\pi/L$. The chosen length scale defines the Reynolds number which then allows to obtain U, the velocity difference between two points separated by a distance L.

$$\text{Re} \quad = \quad \frac{U}{k\nu} \tag{23}$$

$$\Rightarrow 1 \quad = \quad \frac{UL}{2\pi\nu} \tag{24}$$

$$\Leftrightarrow U \quad = \quad \frac{2\pi\nu}{L} \tag{25}$$

where $\nu = 10^{-6}$ m^2.s^{-1} is the kinematic viscosity for water. These numerical values lead to $U = 6.3 \times 10^{-4}$ m.s^{-1}. Note that U is the speed in the frame of reference of the small square area considered here, which might itself be embedded within larger spatial structures (*e.g.*, large eddies) moving at higher speeds in the ocean or any large waterbody.

To determine the diffusivity of small organisms, we use the Stokes-Einstein equation [6]:

$$D = \frac{RT}{N_A} \frac{1}{6\pi\eta a} \tag{26}$$

where $R = 8.314$ J.K^{-1}.mol^{-1} is the molar gas constant, $T = 293$ K is the temperature of the environment, $N_A = 6.0225 \times 10^{23}$ is Avogadro's number, $\eta = 10^{-3}$ m^{-1}.kg.s^{-1} is the viscosity of water and a is the radius of the organism considered. We apply this formula to microphytoplankton organisms of diameter 50 μm, keeping τ outside of the equation for now:

$$\Delta = \sqrt{2D\tau} \tag{27}$$

$$= \sqrt{\frac{RT}{N_A}\frac{\tau}{3\pi\eta a}} \tag{28}$$

$$= \sqrt{\frac{8.314 \times 293}{6.0225 \times 10^{23}}\frac{1}{3\pi \times 10^{-3} \times 25 \times 10^{-6}}}\sqrt{\tau} \tag{29}$$

$$= 1.3 \times 10^{-7}\sqrt{\tau}\ \mathrm{m}. \tag{30}$$

To compute τ, we can consider a phytoplankton doubling rate of $1\,\mathrm{d}^{-1}$ [7], which means, with $p = 0.5$, that $\tau = 0.5$ d.

This leads to $U\tau/2 \approx 5.4 \times 10^3$ cm.d^{-1} and $\Delta \approx 5 \times 10^{-5}$ cm. These two values are much higher than those used in Fig. 3 ($0.1 < U\tau/2 < 2.5$ and $\Delta = 10^{-7}$). A thorough discussion of the parameters is therefore necessary before extrapolating these results to real phytoplanktonic systems.

As the Brownian bug model is currently fairly theoretical in its 2D formulation, a logical next step would be to consider similar dynamics in a 3D-model, which would render the comparison to real data easier. Using actual concentrations of phytoplanktonic organisms (e.g. diatoms), between 10^3 and 10^6 C/L, this would lead to 1 to 10^3 organisms if we kept $L = 1$ cm. We might therefore need to increase the size of the considered square, or apply the model to small bacteria only. With a closer match between field and simulated concentrations, the model could provide us with a better picture of the likely fine-scale spatial structure of phytoplanktonic populations.

Acknowledgements

References

1. W. R. Young, A. J. Roberts, and G. Stuhne. "Reproductive pair correlations and the clustering of organisms." In: **Nature** 412.6844 (2001), pp. 328–331.
2. H. Tennekes and J. L. Lumley. **A first course in turbulence**. MIT press, 1972.
3. D. B. Dusenbery. **Living at micro scale: the unexpected physics of being small**. Harvard University Press, 2009.
4. F. Peters and C. Marrasé. "Effects of turbulence on plankton: an overview of experimental evidence and some theoretical considerations." In: **Marine Ecology Progress Series** 205 (2000), pp. 291–306.
5. R. Pierrehumbert. "Tracer microstructure in the large-eddy dominated regime." In: **Chaos, Solitons & Fractals** 4.6 (1994), pp. 1091–1110.
6. A. Einstein. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." In: **Annalen der physik** 4 (1905).
7. J. E. Bissinger, D. J. S. Montagnes, J. Harples, and D. Atkinson. "Predicting marine phytoplankton maximum growth rates from temperature: Improving on the Eppley curve using quantile regression." In: **Limnology and Oceanography** 53.2 (2008), pp. 487–493.

Supplementary Material

Derivation of G(r,t)

Diffusion and birth/death processes – In this section, we aim to find back from first principles Eq. 2 in Young et al. [1], i.e. Eq. 3 in our manuscript. We will first focus on the

diffusion and birth/death processes, corresponding to the evolution equation for the pair density:

$$\frac{\partial G}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial G}{\partial r}\right) + 2(\lambda - \mu)G + 2\lambda C\delta(\boldsymbol{x}) \tag{31}$$

We first define an ensemble of k identical Brownian bugs in a d-dimensional space. The bug number p is located at $\boldsymbol{x}_p = [x_1, x_2, ...x_d]$. At time t, the space is defined by (a) the number of Brownian bugs k and (b) the vector of their locations $\boldsymbol{X}_k = [\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_k]$. This is also called the Fock space [8].

The probability distribution over the state space is given by the functions $\mathcal{P}_k(\boldsymbol{X}_k, t)$ such that:

$$\mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k = \Pr\{k \text{ bugs, with a bug in } d\boldsymbol{x}_1, \text{a bug in } d\boldsymbol{x}_2, \text{etc.}\} \tag{32}$$

As bugs are indistinguishable, we can exchange $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ (permutation symmetry):

$$\mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_k, t) = \mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_k, t) \tag{33}$$

The normalization is:

$$\mathcal{P}_0(t) + \int \mathcal{P}_1(\boldsymbol{X}_1, t)d\boldsymbol{x}_1 + \int\int \mathcal{P}_2(\boldsymbol{X}_2, t)d\boldsymbol{x}_1 d\boldsymbol{x}_2 + \ldots + \int_{\mathbb{R}^{2k}} \mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k + \ldots = 1 \tag{34}$$

because having k individuals at time t defines a partition of the sample space for $k = 0, 1, 2, \ldots$. We define $b_k(\boldsymbol{x}, t) = \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1}$, i.e. $b_k(\boldsymbol{x}, t)d\boldsymbol{x}$ is the probability that there are k bugs and bug number 1 is in $d\boldsymbol{x}$.

The density of points is defined as:

$$\rho(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} kb_k(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} k\int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1} \tag{35}$$

The pair correlation function is then:

$$G(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1)\int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{36}$$

We define the two particle distribution functions $c_k(\boldsymbol{x}, \boldsymbol{y}, t) = \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2}$. Note that $G(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1)c_k(\boldsymbol{x}, \boldsymbol{y}, t)$.

Proposition

The time derivative of c_k is given by:

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 c_k \tag{37a}$$

$$- k(\lambda + \mu)c_k \tag{37b}$$

$$+ (k+1)\mu c_{k+1} \tag{37c}$$

$$+ 2\frac{\lambda}{k}\delta(\boldsymbol{x} - \boldsymbol{y})b_{k-1}(\boldsymbol{x}) + \lambda\frac{(k-2)(k+1)}{k}c_{k-1} \tag{37d}$$

Proof

We write the evolution of $\mathcal{P}_k(\boldsymbol{X}_k, t)$:

$$\frac{\partial \mathcal{P}_k}{\partial t} = D\nabla_k^2 \mathcal{P}_k \tag{38a}$$

$$- k(\lambda + \mu)\mathcal{P}_k \tag{38b}$$

$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{X}_k, \boldsymbol{y}, t)d\boldsymbol{y} \tag{38c}$$

$$+ \frac{\lambda}{k} \sum_{p=1}^{k} \sum_{q=1, q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t}) \tag{38d}$$

where $\nabla_k^2 = \frac{\partial^2}{\partial_{x_1}^2} + \frac{\partial^2}{\partial_{y_1}^2} + \ldots + \frac{\partial^2}{\partial_{x_k}^2} + \frac{\partial^2}{\partial_{y_k}^2}$ and $\boldsymbol{X}_{k|p} = \boldsymbol{X}_k$ without $\boldsymbol{x}_p$.

Here, Eq. 38a is the diffusion part of the process, Eq. 38b corresponds to the rate at which realizations with k bugs lose a bug by mortality or gain a bug through birth, Eq. 38c corresponds to the rate at which realizations with $k+1$ bugs lose a bug. Finally, a realization with k bugs can also be produced by a birth in a realization with $k-1$ bugs.

Combining the time derivative of $\mathcal{P}_k(\boldsymbol{X}_k, t)$ with the definition of c_k, we obtain:

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D \int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{39a}$$

$$- k(\lambda + \mu) \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{39b}$$

$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, \boldsymbol{z}, t)d\boldsymbol{X}_{k-2}d\boldsymbol{z} \tag{39c}$$

$$+ \frac{\lambda}{k} \int \sum_{p=1}^{k} \sum_{q=1, q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t})d\boldsymbol{X}_{k-2} \tag{39d}$$

We will treat one term after the other.

Diffusion term (39a)

$$D \int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} = D\nabla_k^2 \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{40a}$$

$$= D\nabla_2^2 \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{40b}$$

$$= D\nabla_2^2 c_k \tag{40c}$$

because we already integrate over $k-2$ coordinates (and thus Laplacians for these coordinates are zero).

Second term (39b)

$$k(\lambda + \mu) \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} = k(\lambda + \mu)c_k \tag{41}$$

Death term (39c)

$$(k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, \boldsymbol{z}, t)d\boldsymbol{X}_{k-2}d\boldsymbol{z} = (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1}$$

(42a)

$$= (k+1)\mu c_{k+1} \tag{42b}$$

Birth term (39d)

In this section, we assume $\boldsymbol{x} = \boldsymbol{x}_1$ and $\boldsymbol{y} = \boldsymbol{x}_2$ (the reasoning is the same for different positions of $\boldsymbol{x}$ and $\boldsymbol{y}$ due to permutation symmetry).

We can decompose the double sum, starting with $p = 1$.

$$\sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) = \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{43a}$$

$$+ \delta(\boldsymbol{x}_1 - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{43b}$$

$$+ \ldots \tag{43c}$$

$$+ \delta(\boldsymbol{x}_1 - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{43d}$$

We integrate over the last $k - 2$ coordinates.

$$\int \sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{44a}$$

$$= \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{44b}$$

$$+ \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \ldots \tag{44c}$$

$$+ \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{44d}$$

which leads to

$$\int \sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{45a}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{45b}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots \tag{45c}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_1, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_1 \tag{45d}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{45e}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} + \ldots \tag{45f}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{45g}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)b_{k-1}(\boldsymbol{x}_2) + (k-2)c_{k-1} \tag{45h}$$

By symmetry, if $p = 2$, we obtain $\delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)b_{k-1}(\boldsymbol{x}_1) + (k-2)c_{k-1}$.

Now, we need to use $p \geq 3$.

$$\int \sum_{q=1, q \neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{46a}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{46b}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \tag{46c}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \ldots \tag{46d}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{46e}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \tag{46f}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \tag{46g}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots \tag{46h}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_p, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{k-1} \tag{46i}$$

$$= 2 \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{46j}$$

$$+ (k-3) \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{46k}$$

$$= 2c_{k-1} + (k-3)c_{k-1} \tag{46l}$$
$$= (k-1)c_{k-1} \tag{46m}$$

Thus, $\sum_{p=3}^{k} \sum_{q \neq p} \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p})d\boldsymbol{X}_{k-2} = (k-2)(k-1)c_{k-1}$.

Finally, the birth term is

$$\frac{\lambda}{k}\left(2\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x}) + 2(k-2)c_{k-1} + (k-2)(k-1)c_{k-1}\right) = 2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x}) + \lambda\frac{(k-2)(k+1)}{k}c_{k-1} \tag{47}$$

Combining all terms, we obtain the expected result:

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 c_k \tag{48a}$$

$$- k(\lambda + \mu)c_k \tag{48b}$$
$$+ (k+1)\mu c_{k+1} \tag{48c}$$

$$+ 2\frac{\lambda}{k}\delta(\boldsymbol{x} - \boldsymbol{y})b_{k-1}(\boldsymbol{x}) + \lambda\frac{(k-2)(k+1)}{k}c_{k-1}. \tag{48d}$$

Proposition

In cartesian coordinates, the pair density admits the following evolution equation:

$$\frac{\partial G}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 G + 2(\lambda - \mu)G + 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{x}). \tag{49}$$

Proof

Using the definition of $G(\boldsymbol{x}, \boldsymbol{y}, t)$ and Eq. 37a-37d:

$$\frac{\partial G}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1)D\nabla_2^2 c_k \qquad\qquad \equiv T1 \tag{50a}$$

$$-\sum_{k=2}^{\infty} k(k-1)k(\lambda + \mu)c_k \qquad\qquad \equiv T2 \tag{50b}$$

$$+\sum_{k=2}^{\infty} k(k-1)(k+1)\mu c_{k+1} \qquad\qquad \equiv T3 \tag{50c}$$

$$+\sum_{k=2}^{\infty} k(k-1)2\frac{\lambda}{k}\delta(\boldsymbol{x} - \boldsymbol{y})b_{k-1}(\boldsymbol{x}) \qquad\qquad \equiv T4 \tag{50d}$$

$$+\sum_{k=2}^{\infty} \lambda k(k-1)\frac{(k-2)(k+1)}{k}c_{k-1} \qquad\qquad \equiv T5 \tag{50e}$$

T1 $\quad T1 = D\nabla_2^2 \sum_{k=2}^{\infty} k(k-1)c_k = D\nabla_2^2 G_k$

T3

$$T3 = \sum_{k=2}^{\infty} k(k-1)(k+1)\mu c_{k+1} \tag{51a}$$

$$= \sum_{k'=3}^{\infty} (k'-1)(k'-2)k'\mu c_{k'} \qquad\qquad \text{with } k' = k+1 \tag{51b}$$

$$= \sum_{k'=2}^{\infty} (k'-1)(k'-2)k'\mu c_{k'} \qquad\qquad \text{if } k' = 2, k'-2 = 0 \tag{51c}$$

T4

$$T4 = 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\sum_{k=2}^{\infty} (k-1)b_{k-1}(\boldsymbol{x}) \tag{52a}$$

$$= 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\sum_{k''=1}^{\infty} k''b_{k''}(\boldsymbol{x}) \qquad\qquad \text{with } k'' = k-1 \tag{52b}$$

$$= 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{x}) \tag{52c}$$

T5

$$T5 = \lambda\sum_{k=2}^{\infty} k(k-1)\frac{(k-2)(k+1)}{k}c_{k-1} \tag{53a}$$

$$= \lambda\sum_{k''=1}^{\infty} k''(k''-1)(k''+2)c_{k''} \qquad\qquad \text{with } k'' = k-1 \tag{53b}$$

$$= \lambda\sum_{k''=2}^{\infty} k''(k''-1)(k''+2)c_{k''} \qquad\qquad \text{if } k'' = 1, k''-1 = 0 \tag{53c}$$

T2+T3+T5

$$T2 + T3 + T5 = \sum_{k=2}^{\infty} -k(k-1)k(\lambda + \mu)c_k + (k-1)(k-2)k\mu c_k + k(k-1)(k+2)\lambda c_k \tag{54a}$$

$$= \sum_{k=2}^{\infty} k(k-1)c_k(-k(\lambda + \mu) + (k-2)\mu + (k+2)\lambda) \tag{54b}$$

$$= 2(\lambda - \mu)\sum_{k=2}^{\infty} k(k-1)c_k \tag{54c}$$

$$= 2(\lambda - \mu)G_k \tag{54d}$$

T1+T2+T3+T4+T5   Combining all terms, we have

$$\frac{\partial G}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 G \tag{55a}$$

$$+ 2(\lambda - \mu)G_k \tag{55b}$$

$$+ 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{x}). \tag{55c}$$

Advection process – Using the Kraichnan equations [9], we prove that, taking only advection into account:

$$\frac{\partial G}{\partial t} = \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right). \tag{56}$$

Let $\boldsymbol{r}(t)$ be the distance between two points as a function of time, which follows a geometric Brownian motion, and $\boldsymbol{q}(t) = \log(\boldsymbol{r}(t))/\log(\boldsymbol{r}(0))$. Kraichnan defines $Q(\boldsymbol{q})$ as the probability distribution of $\boldsymbol{q}$. We have $Q = r^d G$. From Kraichnan [9] and Fokker-Planck equations, we know that:

$$\frac{\partial Q}{\partial t} = \gamma\frac{\partial^2 Q}{\partial q^2} - \gamma d\frac{\partial Q}{\partial q} \tag{57}$$

where γ is the stretching parameter, i.e. the diffusivity and γd is the drift.

We have:

$$\frac{\partial Q}{\partial q} = \frac{\partial(r^d G)}{\partial r}\frac{\partial r}{\partial q} \tag{58a}$$

$$= r\left(dr^{d-1}G + r^d\frac{\partial G}{\partial r}\right) \tag{58b}$$

$$= dr^d G + r^{d+1}\frac{\partial G}{\partial r}. \tag{58c}$$

Therefore,

$$\frac{\partial^2 Q}{\partial q^2} = \frac{\partial}{\partial q}\left(\frac{\partial(r^d G)}{\partial r}\frac{\partial r}{\partial q}\right) \tag{59a}$$

$$= r\frac{\partial}{\partial r}\left(dr^d G + r^{d+1}\frac{\partial G}{\partial r}\right) \tag{59b}$$

$$= r\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right) + rd\frac{\partial(r^d G)}{\partial r}. \tag{59c}$$

Using Eq. 57,

$$\frac{\partial (r^d G)}{\partial t} = \gamma r \frac{\partial}{\partial r} \left(r^{d+1} \frac{\partial G}{\partial r} \right) + \gamma dr \frac{\partial (r^d G)}{\partial r} - \gamma dr \frac{\partial (r^d G)}{\partial r} \tag{60a}$$

$$r^d \frac{\partial G}{\partial t} = \gamma r \frac{\partial}{\partial r} \left(r^{d+1} \frac{\partial G}{\partial r} \right) \tag{60b}$$

$$= \gamma r^{1-d} \frac{\partial}{\partial r} \left(r^{d+1} \frac{\partial G}{\partial r} \right). \tag{60c}$$

Writing Eq. 60c with Eq. 31, we obtain Eq. 3.

Stretching parameter γ

γ is computed with simulations, using the formula $r(t) \propto \exp(\gamma dt) \to \frac{1}{2} \ln(r(t)) = \gamma t$ if $d = 2$, with r the separation between pairs of particles. γ is estimated as the slope of

$$\frac{1}{2} \langle \ln(r(t)) \rangle = f(t)$$

with $\langle \ln(r(t)) \rangle$ the average obtained from 800 pairs of particles.

$$\forall t, \langle \ln(r(t)) \rangle = \frac{1}{800} \sum_{p=1}^{800} \ln(r(\boldsymbol{x}_{1,p}(t) - \boldsymbol{x}_{2,p}(t)))$$

where $r(\boldsymbol{x}_{1p}(t) - \boldsymbol{x}_{2p}(t))$ is the distance between a particle $1p$ at position $\boldsymbol{x}_{1p}$ and its counterpart $2p$, initialized with $r(0) = 10^{-7} \forall p$ (see Fig. 4 for γ estimates).
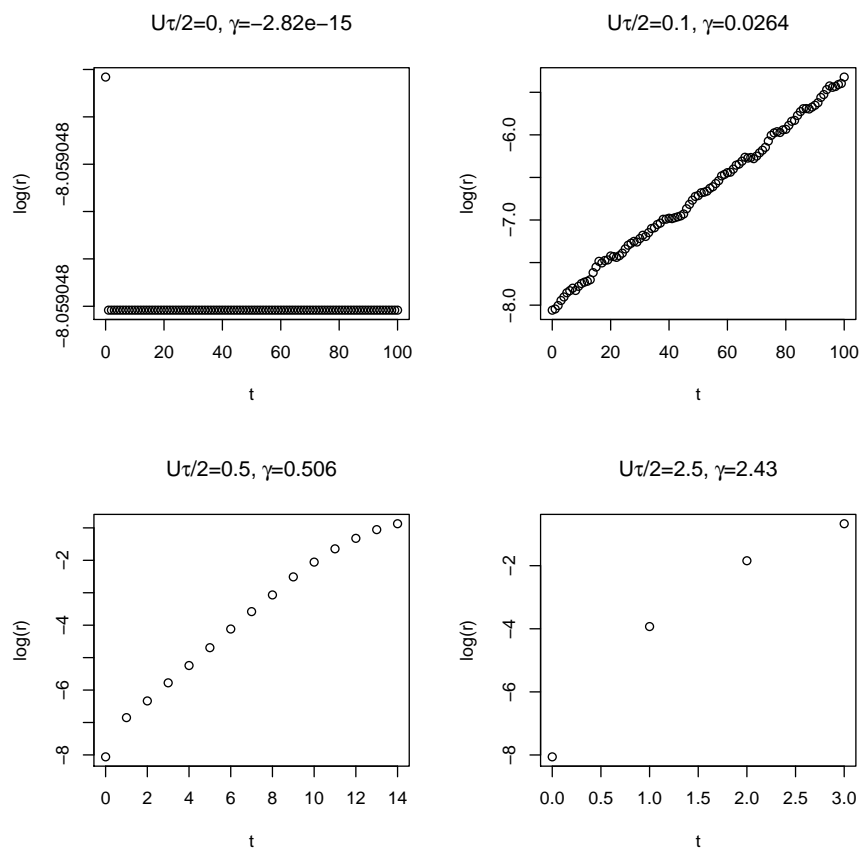
Figure 4. Estimates of γ for different $U\tau/2$.

References

8. D. A. Birch and W. R. Young. "A master equation for a spatial population model with pair interactions." In: **Theoretical Population Biology** 70.1 (2006), pp. 26–42.
9. R. H. Kraichnan. "Convection of a passive scalar by a quasi-uniform random straining field." In: **J. Fluid Mech.** 64.4 (1974), pp. 737–762.