**R E S C I E N C E  C**

# [Re] Reproductive pair correlations and the clustering of organisms

Coralie Picoche[1,2, ID] and Frederic Barraquand[1,2, ID]

[1]Institute of Mathematics of Bordeaux, CNRS & University of Bordeaux, Talence, France – [2]Integrative and Theoretical Ecology, LabEx COTE, University of Bordeaux, Pessac, France

## Introduction

In the present work, we attempt to replicate the results of Young et al. 2001 "Reproductive pair correlations and the clustering of organisms" [1], an analysis of the clustering of organisms in an homogeneous environment mimicking marine turbulence. Using an individual-based model of independent, random-walking particles (also called "brownian bugs"), they show that simple ecological processes such as birth and death in a turbulent and viscous flow leads to the formation of elongated aggregates and therefore departure from the usual, homogeneous solution of the advection-diffusion-reaction equation for a large population. This could explain the observed patchiness of communities such as the phytoplankton.

Our first interest in this paper was mostly driven by a major ecological question: the so-called paradox of the plankton [2], i.e. the surprising diversity of phytoplankton species competing for the same resources in a seemingly homogeneous environment. Aggregation of phytoplanktonic organisms, observed at macro- to micro-scales [3, 4], could explain part of their coexistence, as spatial clustering can help reduce interspecific interactions [5] and/or allow species to benefit from different organic substrates [6]. Due to their size, microphytoplankton organisms experience a mostly viscous environment in a laminar shear fields, with random, but homogeneous changes in directions due to turbulence [7]. Ecological phenomena as simple as growth and death, 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 demographic stochasticity and environmental fluctuations at small scales could provide further explanation for the distribution and coexistence of these organisms in turbulent environments.

In this replication, we aim not only to replicate the main results of the paper, but also to clarify and develop the mathematical background behind the main equations governing the dynamics of the model.

# Methods

## Brownian bug model

The brownian bug model is a discrete-time, individual-based model, here presented in its 2D formulation. 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 attribute being used only for representation purposes. Space is a $L \times L$ square with periodic boundary conditions. Each timestep, of duration $\tau$, is divided into three substeps: (1) demographic processes, (2) diffusion, and (3) advection.

(1) Demographic processes take place during the first substep. The discrete-time Markov chain of the brownian bug model can be approximated as a continuous-time spatial birth-death process (described in the Supplementary Material), also referred to in Young et al. [1] as a Galton-Watson process. 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 this paper, $p = q = 0.5$.
(2) Diffusion is modeled as a brownian motion, i.e. $\boldsymbol{x'}_k(t) = \boldsymbol{x}(t) + \delta\boldsymbol{x}(t)$ 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.
(3) The turbulent flow governing advective stirring follows the Pierrehumbert random map.

$$
\begin{array}{rcll}
x_1(t + \tau) & = & x'_1(t) + (U\tau/2)\cos[kx'_2(t) + \phi(t)] & (1) \\
x_2(t + \tau) & = & x'_2(t) + (U\tau/2)\cos[kx_1(t + \tau) + \theta(t)] & (2)
\end{array}
$$

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

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

## Pair correlation 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 finding a pair of brownian bugs with one member in the area $dA_1$ around $\boldsymbol{x}_1$ and the other in the area $dA_2$ around $\boldsymbol{x}_2$. $G(\boldsymbol{r}, t)$ 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 disappear when $r \to \infty$, $g \to 1$.

### Derivation of G(r,t) —

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

where $\boldsymbol{x}$ is the position of the particle.

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

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

The value of $\gamma$ is computed from simulations (see SI.)

**Analytical solution with advection** – In the presence of advection ($\gamma \neq 0$), a steady-state solution can be found.

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

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

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

We can then integrate Eq. (5) over a small area centered on a particle, with radius $\rho$. Let us first note that

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

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

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

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

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

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

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

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

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

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

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

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

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

**Analytical solution without advection** – When $U = 0$, $\gamma = 0$ and there is no steady solution. We can get back to Eq. (4).

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

Assuming an isotropic environment (and switching to the polar 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 the 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).

$$G(\boldsymbol{x},t) \quad = \quad 2\lambda C \int_{\mathbb{R}^2} \int_0^t H(\boldsymbol{x} - \boldsymbol{x}', t') \delta(\boldsymbol{x}') d^2\boldsymbol{x}' dt'$$

$$\Leftrightarrow \quad = \quad 2\lambda C \int_0^t H(\boldsymbol{x}, t') dt' \tag{15}$$

Eq.(15) can be used in Eq. (12):

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

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

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

which is true.

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

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

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

$$G(r,t) = 2\lambda C \left[\frac{E_1\left(\frac{r^2}{8Dt'}\right)}{8D\pi}\right]_0^t \tag{19}$$

where $E_1$ is the exponential integral. Using $G(r,0) = C^2$ and $\lim_{x\to+\infty} E_1 = 0$ in Eq. (19),

$$G(r,t) = 2\lambda C \frac{E_1\left(\frac{r^2}{8Dt}\right)}{8D\pi} + C^2 \tag{20}$$

$$\Leftrightarrow g(r,t) = \frac{2\lambda}{C}\frac{E_1\left(\frac{r^2}{8Dt}\right)}{8D\pi} + 1 \tag{21}$$

# Results

We were able to reproduce the spatial distributions of brownian bugs in each case described in [1].

In Fig. 1, the model has been run without the advection component and we can see the clumping of organisms due to demographic processes. In Fig. 2, the model has been run without its demographic component, but with advection and diffusion, confirming that hydrodynamics alone cannot ensure cluster formation.
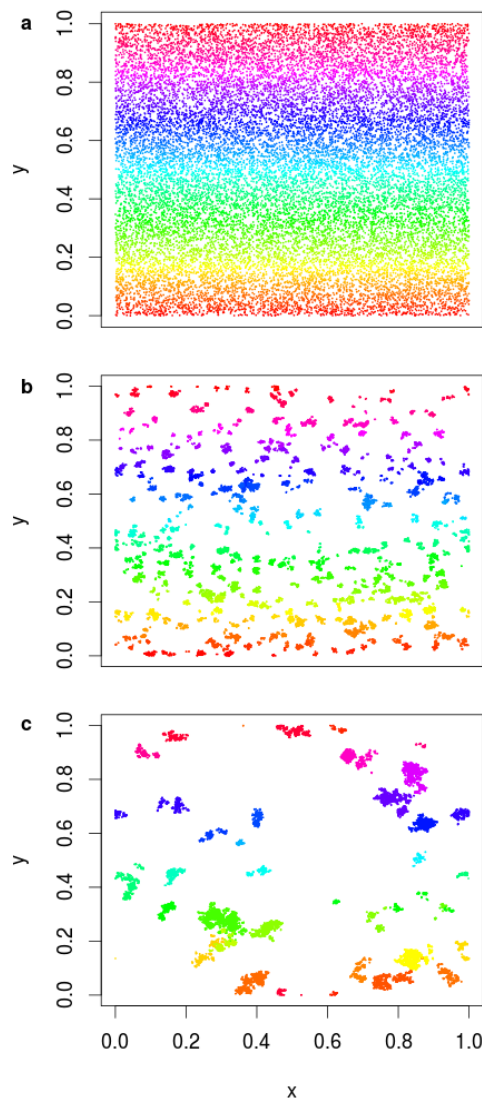


**Figure 1.** Distribution of brownian bugs at different times in a simulation with $\Delta = 10^{-3}$ and $U = 0$: initial conditions with a Poisson spatial distribution (a), $t = 100\tau$ (b) and $t = 1000\tau$ (c). Each particle is identified by a color which corresponds to the initial position on the y axis of its ancestor at t=0.
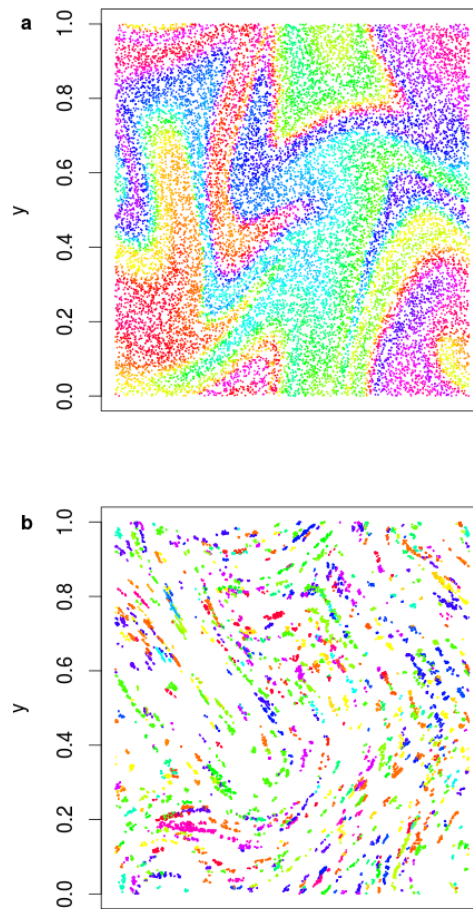
**Figure 2**. Distribution of brownian bugs with different processes in a simulation with $\Delta = 10^{-3}$ and $U\tau/2 = 0.1$: without demographic processes at $t = 30\tau$ (a), and with demographic processes $t = 1000\tau$ (b). Each particle is identified by a color which corresponds to the initial position on the y axis of its ancestor at t=0.

Fig. 3 proved much more challenging. Retrieving the analytical solutions of eq. 4 was the most difficult. However, we also encountered issues when computing the pair correlation functions on simulations: for large values of $r/\Delta$, the pcf was set to 0 when $U = 0$. Despite the missng values, that we can confirm that simulated and analytical pcf match (Fig 3), with a slight underestimation. The pcf also got closer to 1 than the predictions for larger values of $r$.
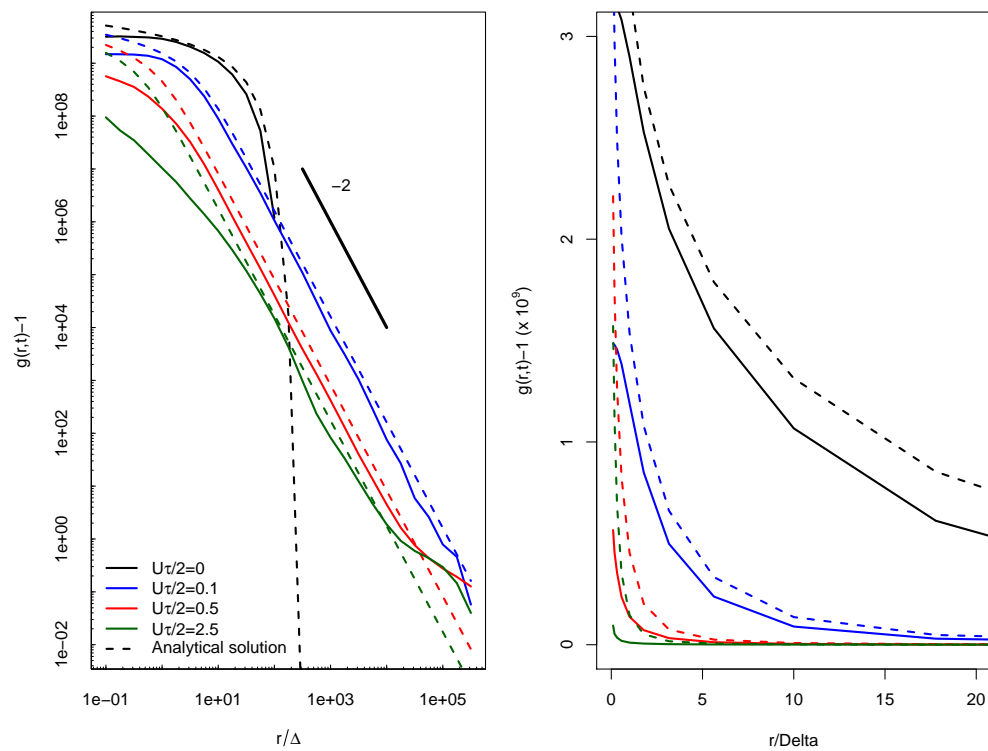
**Figure 3**. Logarithmic (a) and linear (b) plots of $g(r,t)$ versus $r/\Delta$, with $\Delta = 10^{-7}$ and $U\tau/2 = 0, 0.1, 0.5, 2.5$ at $t = 1000\tau$. Solid lines result from simulations, dotted lines correspond to analytical solutions and the solid grey line indicates the $r^{-2}$ scaling predicted by eq. 3.

## Discussion

We successfully replicated both the simulations and retrieved the analytical solutions of [1]. Even though stochasticity prevents us from replicating exactly the same distributions as those seen in Fig. 1 and Fig. 2, we considered the patterns to be close enough to validate the simulations. Fig. 3 was also very close to the one shown in the article, despite a slight underestimation.

The most challenging part of the replication was not actually to replicate the results, but to find the analytical expression of the pair density function $G(r,t)$. In the absence of additional explanations regarding both the derivation and the solution for this function, we could not compute the analytical solutions shown in Fig. 3. William Young helped us find back these solutions, with quick and frequent feedbacks. We hope that the additional details on the theory will help the reader through the article.

The original paper was not strongly quantitative; 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 as the Kolmogorv scale, that is the scale at which viscosity starts dominating turbulence. In the ocean, this value oscillates between 1 and 10 mm [8] (which corresponds to the value which was given by William Young, i.e. 1 cm). If we consider a phytoplankton doubling rate of 0.5-1 d$^{-1}$ [9], $p = 0.5$ means that $\tau = 0.25-0.5$ d. In this case, assuming we can write $U\tau/2 = 0.1L$ (the value used in Fig. 2), $U \approx 5.5$ $10^{-5}$ cm.s$^{-1}$, which seems very low for the oceanic current (around 1-10 cm.s$^{-1}$ near the coast in [5], similar in the North Atlantic Ocean [10]). We should therefore have a lower

reproduction probability $p$, if we wanted to keep the values of the other parameters. The diffusion coefficient can be hard to approximate. Broström [11] suggests a value of $10^{-8}$ $m^2.s^{-1}$, which would lead to $\Delta \approx 10^{-4}$. This is much higher from the value used in Fig. 3 ($10^{-7}$) but only one order of magnitude lower than the value used in Fig. 1 and Fig. 2. A thorough discussion of the parameters is therefore necessary before extrapolating these results to real phytoplanktoni systems.

As the brownian bug model is very general, a next step would be to consider similar dynamics in a 3D-model, for which the pair correlation function can be evaluated easily. Using actual concentrations of phytoplanktonic organisms, between $10^3$ and $10^6$ C/L, this would lead to 1 to $10^3$ organisms if we kept $L$ = 1 cm. We may therefore need to increase the size of the considered square if doing so. With a more quantitative models, this could allow a new modeling of phytoplanktonic organisms.

## Acknoweldgements

## 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. G. E. Hutchinson. "The paradox of the plankton." In: **The American Naturalist** 95.882 (1961), pp. 137–145.
3. S. Lovejoy. "Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity." In: **Journal of Plankton Research** 23.2 (2001), pp. 117–141.
4. B. Pinel-Alloul and A. Ghadouani. "Spatial heterogeneity of planktonic microorganisms in aquatic systems." In: **The spatial distribution of microbes in the environment**. Springer, 2007, pp. 203–310.
5. J. S. Font-Muñoz, A. Jordi, I. Tuval, J. Arrieta, S. Anglès, and G. Basterretxea. "Advection by ocean currents modifies phytoplankton size structure." In: **J. R. Soc. Interface** 14.130 (2017), p. 20170046.
6. J. R. Seymour, Marcos, and R. Stocker. "Resource Patch Formation and Exploitation throughout the Marine Microbial Food Web." In: **The American Naturalist** 173.1 (2009), p. 15.
7. 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.
8. A. D. Barton, B. A. Ward, R. G. Williams, and M. J. Follows. "The impact of fine-scale turbulence on phytoplankton community structure." In: **Limnology and Oceanography: Fluids and Environments** 4.1 (2014), pp. 34–49.
9. 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.
10. M. K. Flatau, L. Talley, and P. P. Niiler. "The North Atlantic Oscillation, Surface Current Velocities, and SST Changes in the Subpolar North Atlantic." In: **Journal of Climate** 16.14 (2003), pp. 2355–2369.
11. G. Broström. "On advection and diffusion of plankton in coarse resolution ocean models." In: **Journal of Marine Systems** 35.1-2 (2002), pp. 99–110.
12. 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.

## Supplementary Material

### Derivation of G(r,t)

**Diffusion and birth/death processes −** In this part, we aim to find back eq. (2) in [1], i.e. eq. 3 in our manuscript. We will first focus on the diffusion and birth/death processes, that is:

$$\frac{\partial G}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial G}{\partial r}\right) + 2(\lambda - \mu)G + 2\lambda C\delta(\boldsymbol{x}) \tag{22}$$

We first define an ensemble of $k$ identical brownian bugs in a d-dimension 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.

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

$$\mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k = \Pr\{k \text{ bugs, with a bug in } d\boldsymbol{x}_1, \text{a bug in } d\boldsymbol{x}_2, \text{etc.}\} \tag{23}$$

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

$$\mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_k, t) = \mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_k, t) \tag{24}$$

The normalization is:

$$\mathcal{P}_0(t) + \int \mathcal{P}_1(\boldsymbol{X}_1, t)d\boldsymbol{x}_1 + \int\int \mathcal{P}_2(\boldsymbol{X}_1, t)d\boldsymbol{x}_1 d\boldsymbol{x}_2 + \ldots + \int_{\mathbb{R}^{2k}} \mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k = 1 \tag{25}$$

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:

$$\rho(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} kb_k(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} k\int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1} \tag{26}$$

The pair correlation function is then:

$$G(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1)\int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{27}$$

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
We show that

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 c_k \tag{28a}$$

$$- k(\lambda + \mu)c_k \tag{28b}$$

$$+ (k+1)\mu c_{k+1} \tag{28c}$$

$$+ 2\frac{\lambda}{k}\delta(\boldsymbol{x} - \boldsymbol{y})b_{k-1}(\boldsymbol{x}) + \lambda\frac{(k-2)(k+1)}{k}c_{k-1} \tag{28d}$$

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

$$\frac{\partial \mathcal{P}_k}{\partial t} = D\nabla_k^2 \mathcal{P}_k \tag{29a}$$

$$- k(\lambda + \mu)\mathcal{P}_k \tag{29b}$$

$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{X}_k, \boldsymbol{y}, t)d\boldsymbol{y} \tag{29c}$$

$$+ \frac{\lambda}{k}\sum_{p=1}^{k}\sum_{q=1,q\neq p}^{k}\delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t}) \tag{29d}$$

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. 29a is the diffusion part of the process, eq. 29b corresponds to the rate at which realizations with $k$ bugs lose a bug by mortality or gain a bug through birth, eq. 29c corresponds to the rate at which realization 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.

Using the derivation of $\mathcal{P}_k(\boldsymbol{X}_k, t)$ and the definition of $c_k$, we have:

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{30a}$$

$$- k(\lambda + \mu)\int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{30b}$$

$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, \boldsymbol{z}, t)d\boldsymbol{X}_{k-2}d\boldsymbol{z} \tag{30c}$$

$$+ \frac{\lambda}{k}\int \sum_{p=1}^{k}\sum_{q=1,q\neq p}^{k}\delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t}) \tag{30d}$$

We will treat one term after the other.

## Diffusion term (30a)

$$D\int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} = D\nabla_k^2 \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{31a}$$

$$= D\nabla_2^2 \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2} \tag{31b}$$

$$= D\nabla_2^2 c_k \tag{31c}$$

because we already integrate over $k - 2$ coordinates.

### Second term (30b)

$$k(\lambda + \mu) \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t) d\boldsymbol{X}_{k-2} = k(\lambda + \mu)c_k \tag{32}$$

### Death term (30c)

$$(k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, \boldsymbol{z}, t) d\boldsymbol{X}_{k-2} d\boldsymbol{z} = (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-1}, t) d\boldsymbol{X}_{k-1} \tag{33a}$$

$$= (k+1)\mu c_{k+1} \tag{33b}$$

### Birth term (30d)

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 $p = 1$.

$$\sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) = \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{34a}$$

$$+ \delta(\boldsymbol{x}_1 - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{34b}$$

$$+ \ldots \tag{34c}$$

$$+ \delta(\boldsymbol{x}_1 - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) \tag{34d}$$

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

$$\int \sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{35a}$$

$$= \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{35b}$$

$$+ \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \ldots \tag{35c}$$

$$+ \int \delta(\boldsymbol{x}_1 - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{35d}$$

$$\tag{35e}$$

As $\int \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)F(\boldsymbol{x}_q)d\boldsymbol{x}_q = F(\boldsymbol{x}_p)$, we note here that

$$\int_{x_3} \int_{\ldots} \int_{x_q} \int_{\ldots} \int_{x_k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k = \tag{36}$$

$$\int_{x_3} \int_{\ldots} \int_{x_{q-1}} \int_{x_{q+1}} \int_{\ldots} \int_{x_k} \mathcal{P}_{k-2}(\boldsymbol{x}_3, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{37}$$

$$\sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t) = \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \quad \text{(38a)}$$

$$+ \int \mathcal{P}_{k-3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_1, \ldots \ldots \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots$$
$$\text{(38b)}$$

$$+ \int \mathcal{P}_{k-3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \ldots \boldsymbol{x}_1, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_1 \quad \text{(38c)}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)\int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \quad \text{(38d)}$$

$$+ \int \mathcal{P}_{k-3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} + \ldots \quad \text{(38e)}$$

$$+ \int \mathcal{P}_{k-3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \quad \text{(38f)}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)b_{k-1}(\boldsymbol{x}_2) + (k-2)c_{k-1} \quad \text{(38g)}$$

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 $p \geq 3$.

$$\int \sum_{q=1,q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \quad \text{(39a)}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \quad \text{(39b)}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \quad \text{(39c)}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_3)\mathcal{P}_{k|p}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \ldots \quad \text{(39d)}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_k)\mathcal{P}_{k|p}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \quad \text{(39e)}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \quad \text{(39f)}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \quad \text{(39g)}$$

$$+ \int \mathcal{P}_{k|p}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots \quad \text{(39h)}$$

$$+ \int \mathcal{P}_{k|p}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{k-1} \quad \text{(39i)}$$

$$= 2 \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \quad \text{(39j)}$$

$$+ (k-3) \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \quad \text{(39k)}$$

$$= 2c_{k-1} + (k-3)c_{k-1} \quad \text{(39l)}$$
$$= (k-1)c_{k-1} \quad \text{(39m)}$$

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

Finally, the birth term is

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

(40)

Combining all terms, we obtain:

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x},\boldsymbol{y},t)=D\nabla_2^2 c_k \tag{41a}$$

$$-k(\lambda+\mu)c_k \tag{41b}$$

$$+(k+1)\mu c_{k+1} \tag{41c}$$

$$+2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x})+\lambda\frac{(k-2)(k+1)}{k}c_{k-1} \tag{41d}$$

## Proposition

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

$$\frac{\partial G}{\partial t}(\boldsymbol{x},\boldsymbol{y},t)=D\nabla_2^2 G+2(\lambda-\mu)G+2\lambda\delta(\boldsymbol{x}-\boldsymbol{y})\rho(\boldsymbol{x}) \tag{42}$$

## Proof

Using the definition of $G(\boldsymbol{x},\boldsymbol{y},t)$ and eq. 28a-28d:

$$\frac{\partial G}{\partial t}(\boldsymbol{x},\boldsymbol{y},t)=\sum_{k=2}^{\infty}k(k-1)D\nabla_2^2 c_k \qquad\qquad \equiv T1 \tag{43a}$$

$$-\sum_{k=2}^{\infty}k(k-1)k(\lambda+\mu)c_k \qquad\qquad \equiv T2 \tag{43b}$$

$$+\sum_{k=2}^{\infty}k(k-1)(k+1)\mu c_{k+1} \qquad\qquad \equiv T3 \tag{43c}$$

$$+\sum_{k=2}^{\infty}k(k-1)2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x}) \qquad\qquad \equiv T4 \tag{43d}$$

$$+\sum_{k=2}^{\infty}\lambda k(k-1)\frac{(k-2)(k+1)}{k}c_{k-1} \qquad\qquad \equiv T5 \tag{43e}$$

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

**T3**

$$T3=\sum_{k=2}^{\infty}k(k-1)(k+1)\mu c_{k+1} \tag{44a}$$

$$=\sum_{k'=3}^{\infty}(k'-1)(k'-2)k'\mu c_{k'} \qquad\qquad \text{with } k'=k+1 \tag{44b}$$

$$=\sum_{k'=2}^{\infty}(k'-1)(k'-2)k'\mu c_{k'} \qquad\qquad \text{if } k'=2, k'-2=0 \tag{44c}$$

T4

$$T4 = 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\sum_{k=2}^{\infty}(k-1)b_{k-1}(\boldsymbol{x}) \tag{45a}$$

$$= 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\sum_{k''=1}^{\infty}k''b_{k''}(\boldsymbol{x}) \qquad \text{with } k'' = k - 1 \tag{45b}$$

$$= 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{x}) \tag{45c}$$

T5

$$T5 = \lambda\sum_{k=2}^{\infty}k(k-1)\frac{(k-2)(k+1)}{k}c_{k-1} \tag{46a}$$

$$= \lambda\sum_{k''=1}^{\infty}k''(k''-1)(k''+2)c_{k''} \qquad \text{with } k'' = k - 1 \tag{46b}$$

$$= \lambda\sum_{k''=2}^{\infty}k''(k''-1)(k''+2)c_{k''} \qquad \text{if } k'' = 1, k'' - 1 = 0 \tag{46c}$$

T2+T3+T5

$$T2 + T3 + T5 = \sum_{k=2}^{\infty}-k(k-1)k(\lambda+\mu)c_k + (k-1)(k-2)k\mu c_k + k(k-1)(k+2)\lambda c_k \tag{47a}$$

$$= \sum_{k=2}^{\infty}k(k-1)c_k(-k(\lambda+\mu) + (k-2)\mu + (k+2)\lambda) \tag{47b}$$

$$= 2(\lambda - \mu)\sum_{k=2}^{\infty}k(k-1)c_k \tag{47c}$$

$$= 2(\lambda - \mu)G_k \tag{47d}$$

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

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

$$+ 2(\lambda - \mu)G_k \tag{48b}$$

$$+ 2\lambda\delta(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{x}) \tag{48c}$$

**Advection process —** We here prove that, using advection only, we have

$$\frac{\partial G}{\partial t} = \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right) \tag{49}$$

We are using here the equations from Kraichnan (1974) [12]. Let $\boldsymbol{r}(t)$ being 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 $Q = r^d G$. We know that (from Kraichnan and Fokker-Planck equation):

$$\frac{\partial Q}{\partial t} = \gamma\frac{\partial^2 Q}{\partial q^2} - \gamma d\frac{\partial Q}{\partial q} \tag{50}$$

where $\gamma$ is the stretching parameter, i.e. the diffusivity and $\gamma d$ is the drift.

We have:

$$\frac{\partial Q}{\partial q} = \frac{\partial(r^d G)}{\partial r}\frac{\partial r}{\partial q} \tag{51a}$$

$$= r\left(dr^{d-1}G + r^d\frac{\partial G}{\partial r}\right) \tag{51b}$$

$$= dr^d G + r^{d+1}\frac{\partial G}{\partial r} \tag{51c}$$

Therefore,

$$\frac{\partial^2 Q}{\partial q^2} = \frac{\partial}{\partial q}\left(\frac{\partial(r^d G)}{\partial r}\frac{\partial r}{\partial q}\right) \tag{52a}$$

$$= r\frac{\partial}{\partial r}\left(dr^d G + r^{d+1}\frac{\partial G}{\partial r}\right) \tag{52b}$$

$$= r\frac{\partial}{\partial r}(r^{d+1}\frac{\partial G}{\partial r}) + rd\frac{\partial(r^d G)}{\partial r} \tag{52c}$$

Using eq. 50,

$$\frac{\partial(r^d G)}{\partial t} = \gamma r\frac{\partial}{\partial r}(r^{d+1}\frac{\partial G}{\partial r}) + \gamma dr\frac{\partial(r^d G)}{\partial r} - \gamma dr\frac{\partial(r^d G)}{\partial r} \tag{53a}$$

$$r^d\frac{\partial G}{\partial t} = \gamma r\frac{\partial}{\partial r}(r^{d+1}\frac{\partial G}{\partial r}) \tag{53b}$$

$$= \gamma r^{1-d}\frac{\partial}{\partial r}(r^{d+1}\frac{\partial G}{\partial r}) \tag{53c}$$

## Stretching parameter $\gamma$

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

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

with $\langle ln(r(t))\rangle$ being 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 $\gamma$ estimates).
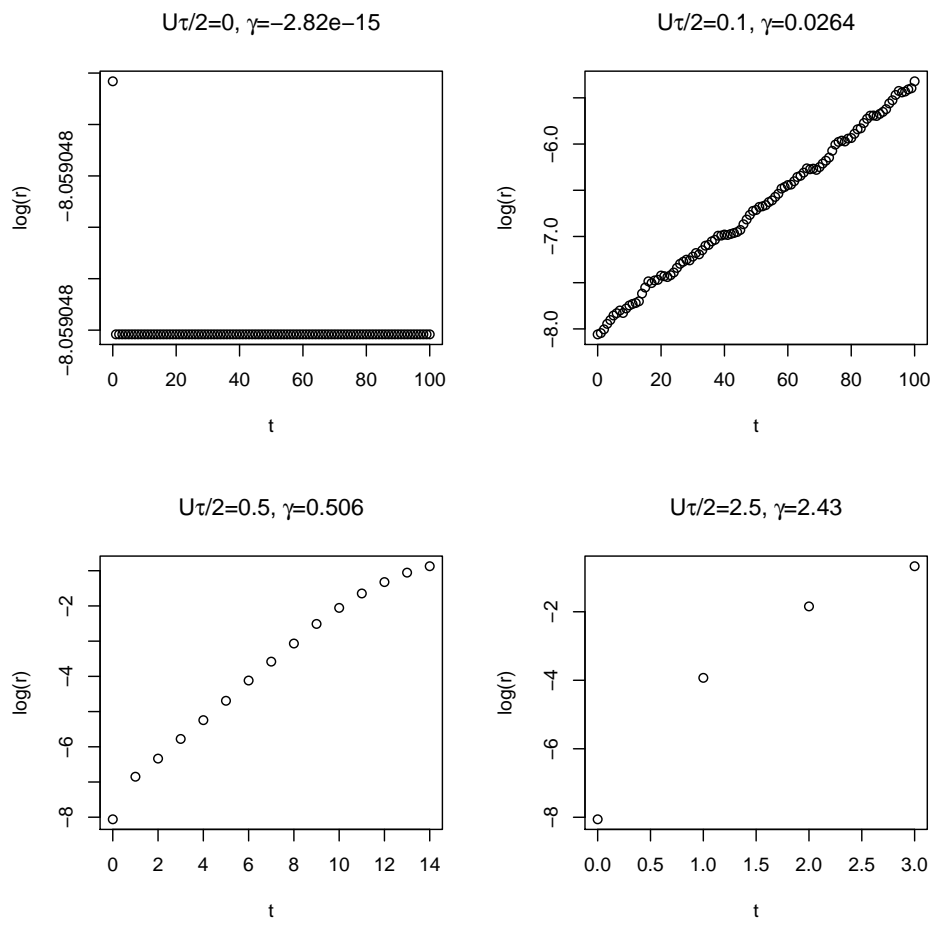
**Figure 4**. Estimates of $\gamma$ for different $U\tau/2$