Answers to comments during discussion (05/01/22)

Thomas process

We show that interspecific pcf which seemed underestimated are actually close to 1 when using the right bandwidth (here, the default value used in the R package `spatstat`, ie $\delta = 0.26/\lambda^{1/3}$). We also confirm that results are the same when directly using `spatstat`.

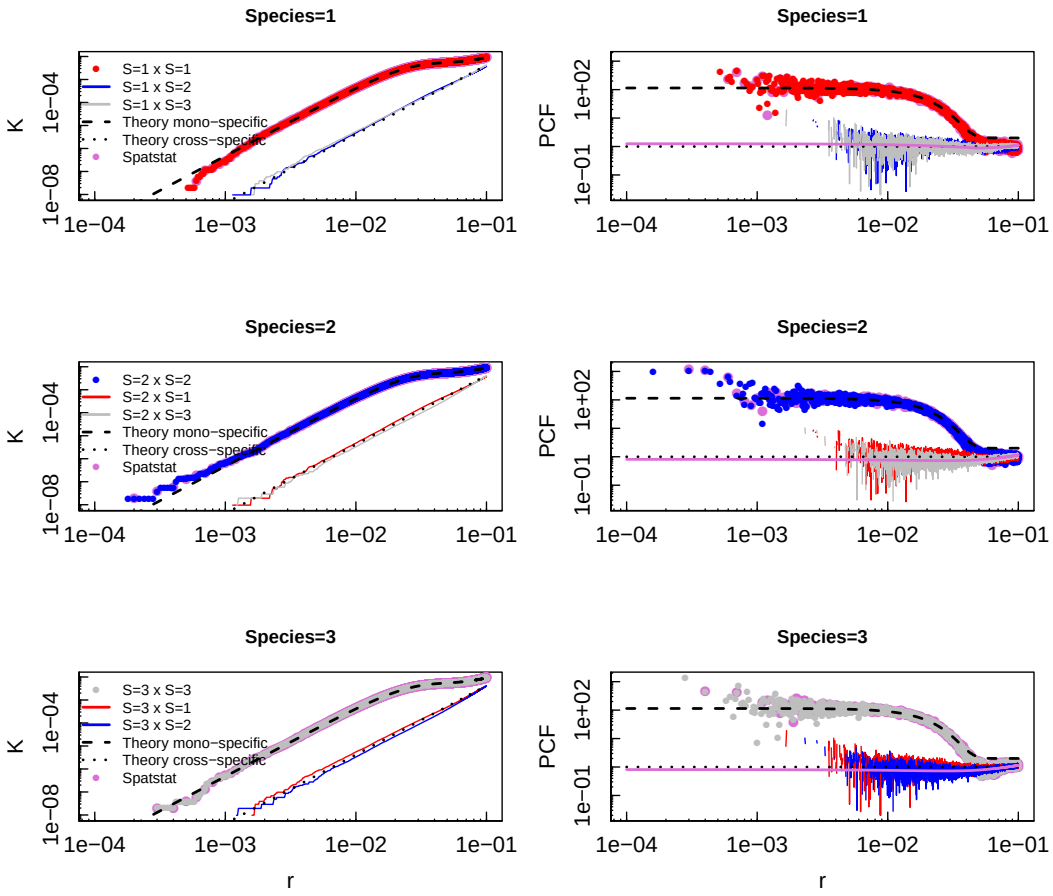

Figure 1: Mono- and cross-specific Ripley's K function and PCF values for 3 species following a Thomas process with parent intensity $\lambda_p = 200$ cm^{-3}, number of children per parent $N_c = 50$, in a volume of 1 cm^3, $\sigma = 0.01$ and $\delta \approx 0.012$.

Minimum distance between points

Theory

One of the reason why estimating K and g(r) is difficult is that for small radii (below 10^{-2}), we can find very few observed distances between pairs of points. As a first proxy, we want to estimate the minimum expected distance

between points when they are uniformly distributed.

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

$$\begin{aligned} Pr(\min(R_1, ..., R_n) > r) &= Pr(R_1 > r, ..., R_n > r) \\ &= \Pi_i Pr(R_1 > r) \\ &= \Pi_i \exp(-b_d r^d \lambda) \\ &= \exp(-b_d r^d \Sigma_i \lambda) \end{aligned} \tag{1}$$

We can then conclude that the distribution of the minimum distance follows the same distribution as the nearest neighbour distribution, but with intensity $n\lambda$.

Clark & Evans (1979) show that a variable with probability distribution (with notations changed to fit our own) $f(r) = \frac{d\lambda\pi^{d/2}r^{d-1}}{\Gamma(\frac{d}{2}+1)}\exp(-\frac{\lambda\pi^{d/2}r^d}{\Gamma(\frac{d}{2}+1)}) = d\lambda b_d r^{d-1}\exp(-\lambda b_d r^d)$[1] has a mean of $\frac{\left(\Gamma(\frac{d}{2}+1)\right)^{1/d}\Gamma(\frac{1}{d}+1)}{\lambda^{1/d}\pi^{1/2}}$.

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

In 3D, $\mu = (n\lambda)^{-1/3}\frac{\left(\Gamma(\frac{3}{2}+1)\right)^{1/3}\Gamma(\frac{1}{3}+1)}{\pi^{1/2}} = (n\lambda)^{-1/3}\left(\frac{3}{2}\Gamma(3/2)\right)^{1/3}\frac{1}{3}\Gamma(1/3)\frac{1}{\pi^{1/2}} \approx 0.554\frac{1}{(n\lambda)^{1/3}}$.

This needs to be taken into account when defining λ. As a first proxy, using λ =10^4cm^{-3} and n ≈10^4 if we model a 1 cm^3 cube, we would expect the smallest distance to be 1.2×10^{-3} cm.

Simulations

(inconclusive for now because monospecific minimum distances can either be above or below interspecific minimum distances, and we only have 3 species)

[1] Remember that b_d, the volume of the unit sphere is equal to $\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$

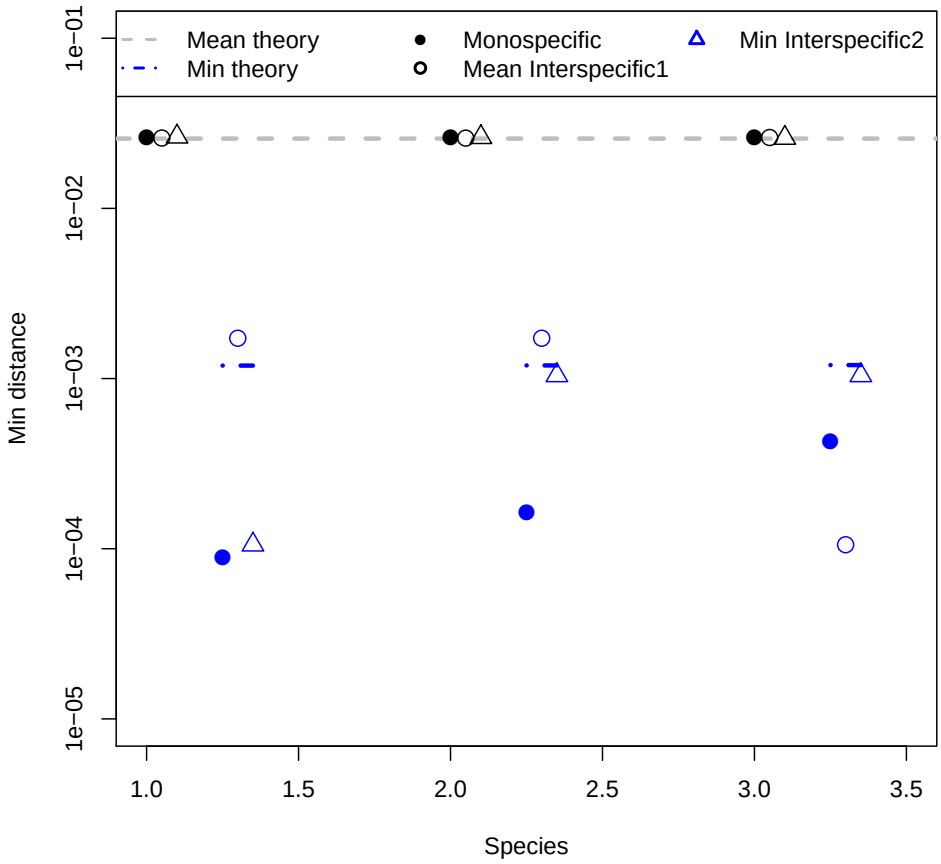

Figure 2: Mean and minimum distance to the nearest neighbour of the same species, or of a different species, compared to the prediction of a uniform distribution. Values correspond to a simulation with 3 species using the Brownian Bug Model with concentration $\lambda = 10^4 \text{cm}^{-3}$ in a volume of 1 cm^3, with advection, after 1000 time steps

Brownian Bug Model

Warning: we use the diatom parameters, with one big change: as we want to focus on small radii, we increase the concentration from 10 cm^{-3} to 10^4 cm^{-3}.

Comparison with spatstat package

We show that we find the exact same result as the package `spatstat`.

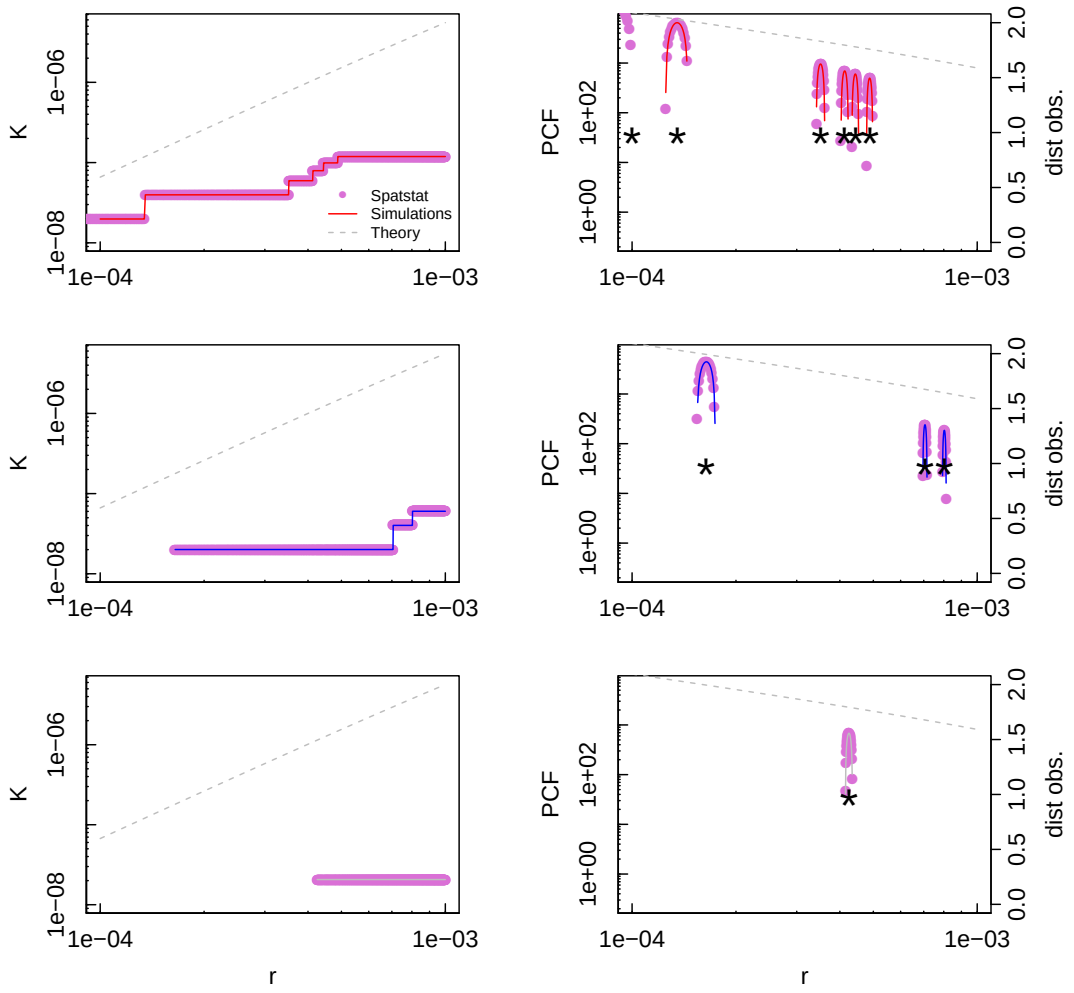

Figure 3: Monospecific Ripley's K function and PCF values for 3 species using the Brownian Bug Model with concentration $\lambda = 10^4 \text{cm}^{-3}$ in a volume of 1 cm^3, with advection, after 1000 time steps, using a bandwidth $\delta = 10^{-5}$ and focusing on small radii around a particle. Black stars correspond to observed distances in the simulation.

Current issues with estimated pcf

Bandwidth

In Fig. 3, we show that the parabole observed correspond to the shape of the kernel when points are too rare and the bandwidth is too small ($\delta = 10^{-5}$). This also explains the sudden absence of points we observed previously.

We thus compared the effect of different bandwidths in the same range of radius in Fig. 4.

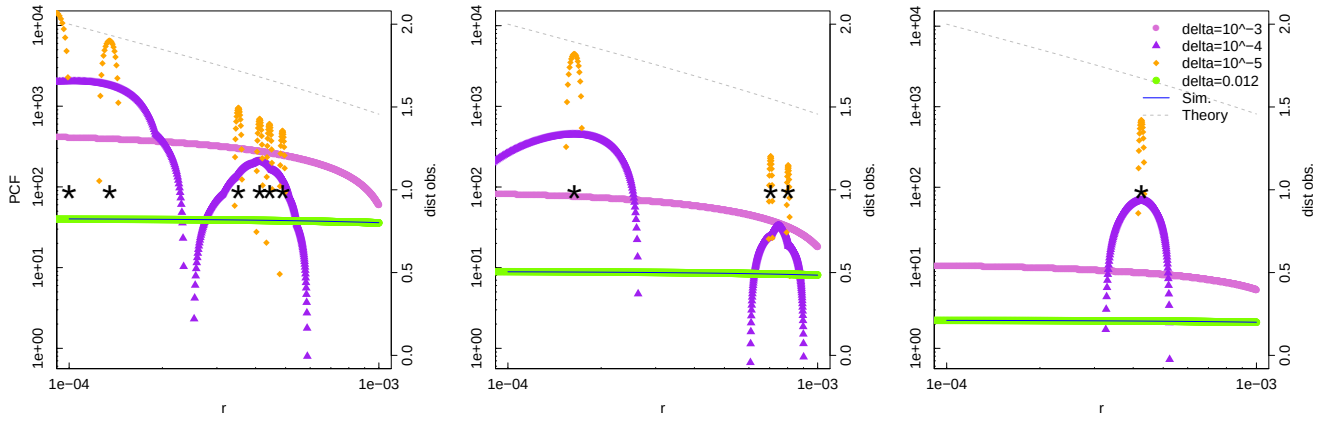

Figure 4: Monospecific PCF values for 3 species using the Brownian Bug Model with concentration $\lambda = 10^4 \text{cm}^{-3}$ in a volume of 1 cm^3, with advection, after 1000 time steps, using different bandwidth δ for small radius ranges (simulations were done with default δ values, i.e. $\delta \approx 0.012$). Black stars correspond to observed distances in the simulation.

We can also look at the whole range of radius values we want to study in Fig. 5.

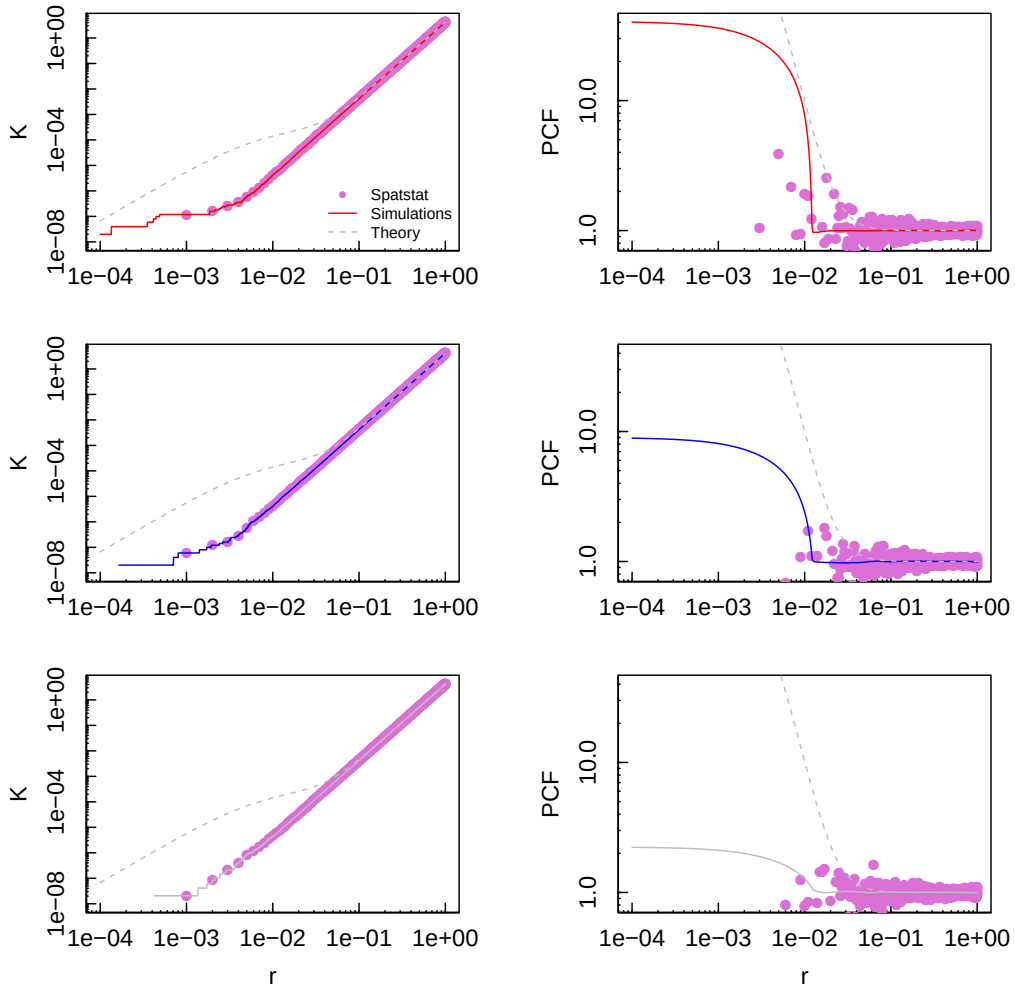

Figure 5: Monospecific Ripley's K function and PCF values for 3 species using the Brownian Bug Model with concentration $\lambda = 10^4 \text{cm}^{-3}$ in a volume of 1 cm^3, with advection, after 1000 time steps, using $\delta \approx 0.012$ with all ranges we want to consider.

I suggest to stop computing pcf from now. The search for the right bandwidth gets us nowhere, and our analyses and conclusions depend on the dominance index, which is computed with Ripley's function.

Ripley's function

Theoretical values

By integrating the pcf, we found an analytical formula for Ripley's function.

With advection,

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

Warning: for now, this formula gives aberrant results (including negative dominance for rare species), which tend to show that it is most certainly wrong (and simulations are also far from these values in Fig. 5, which confirm the conclusion). By default, we stick to it while waiting for Bill's algebra.

Without advection,

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

Simulations

Monospecific K We show in Fig. 3 and 5 that for now, simulated values of K are far from the theoretical values (the fit for larger r is only due to the prevalence of the $\frac{4}{3}\pi r^3$ term and only illustrate the fit to the uniform distribution when $r \to \infty$).

Interspecific K

We can at least find the expected values of interspecific values of K (Fig. 6).

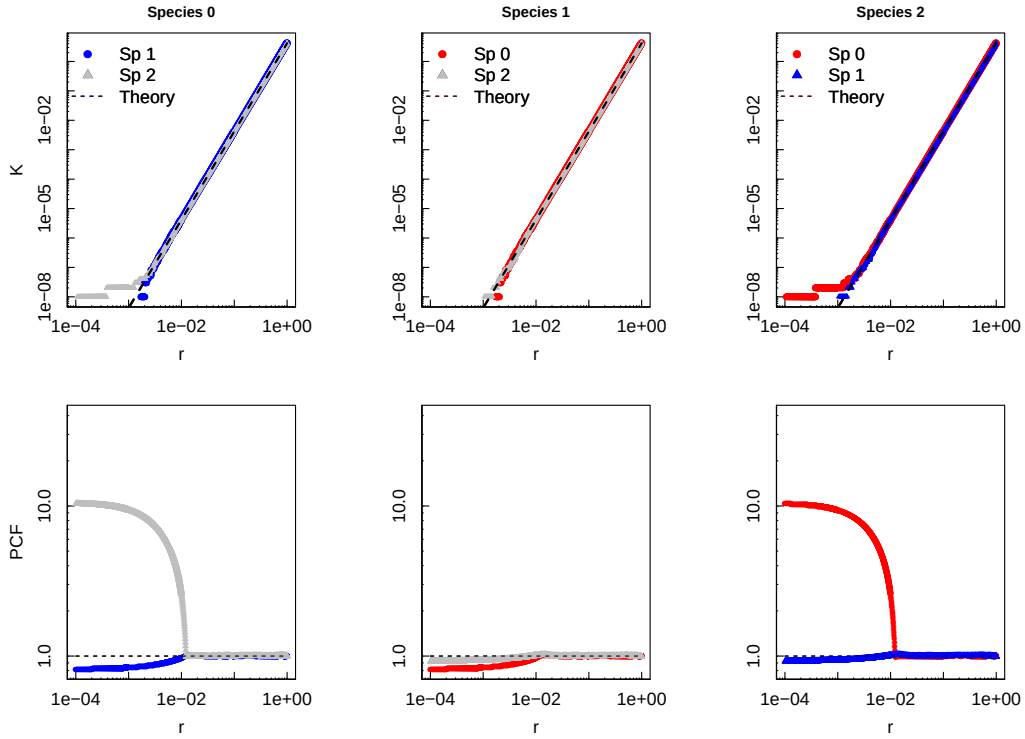

Figure 6: Interspecific Ripley's K function and PCF values for 3 species using the Brownian Bug Model with concentration λ =10^4cm^{-3} in a volume of 1 cm^3, with advection, after 1000 time steps, using $\delta \approx 0.012$ with all ranges we want to consider.

References

Clark, P.J. & Evans, F.C. (1979). Generalization of a nearest neighbor measure of dispersion for use in k dimensions. 60, 316–317.