



**HAL**  
open science

## Patchy patterns due to group dispersal

Samuel Soubeyrand, Lionel Roques, Jérôme Coville, Julien Fayard

► **To cite this version:**

Samuel Soubeyrand, Lionel Roques, Jérôme Coville, Julien Fayard. Patchy patterns due to group dispersal. *Journal of Theoretical Biology*, 2011, 271 (1), pp.87. 10.1016/j.jtbi.2010.11.047. hal-00657590

**HAL Id: hal-00657590**

**<https://hal.science/hal-00657590>**

Submitted on 7 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Author's Accepted Manuscript

Patchy patterns due to group dispersal

Samuel Soubeyrand, Lionel Roques, Jérôme Coville,  
Julien Fayard

PII: S0022-5193(10)00642-9  
DOI: doi:10.1016/j.jtbi.2010.11.047  
Reference: YJTBI6269

To appear in: *Journal of Theoretical Biology*

Received date: 24 February 2010  
Revised date: 30 November 2010  
Accepted date: 30 November 2010

Cite this article as: Samuel Soubeyrand, Lionel Roques, Jérôme Coville and Julien Fayard, Patchy patterns due to group dispersal, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.11.047](https://doi.org/10.1016/j.jtbi.2010.11.047)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

## Patchy patterns due to group dispersal

Samuel Soubeyrand<sup>1</sup>, Lionel Roques, Jérôme Coville and Julien Fayard

*INRA, UR546 Biostatistics and Spatial Processes, F-84914 Avignon, France*

---

### Abstract

The presence of multiple foci in population patterns may be due to various processes arising in the population dynamics. Group dispersal, which has been lightly investigated for airborne species, is one of these processes.

We built a stochastic model generating the dispersal of groups of particles. This model may be viewed as an extension of classical dispersal models based on parametric kernels. It has a hierarchical structure: at the first stage group centers are drawn under a classical dispersal kernel; at the second stage the particles are diffused around their group centers. Analytic and simulation results show that group dispersal is a sufficient condition to generate patterns with multiple foci, i.e. patchy patterns, even if the population can remain particularly concentrated.

*Key words:* Brownian motion, Dispersal kernel, Multiple foci, Population dynamics, Population pattern

---

<sup>1</sup>Corresponding Author; E-mail address: Samuel.Soubeyrand@avignon.inra.fr; Tel.: +33 432722185; Fax.: +33 432722182)

## 1. Introduction

### *1.1. Patterns with foci and groups of particles*

Snapshots of epidemics and invasions often show multiple foci (for real examples, see Heesterbeek and Zadoks, 1987; Nash et al., 1995). A typical situation is the presence of one primary focus and several secondary foci which will eventually merge with the primary focus (Shigesada and Kawasaki, 1997, chap. 5). The reasons why such patterns are generated are diverse. For organisms taken into account in this article, i.e. airborne plant pathogens and plant propagules, the main reasons invoked to explain multiple foci patterns may be (i) spatial heterogeneity in host receptivity or environmental conditions (Kauffman and Jules, 2006; Laine and Hanski, 2006; Soubeyrand et al., 2009), (ii) long distance dispersal (Aylor, 1987; Cannas et al., 2006; Ferrandino, 1993; Filipe and Maule, 2004; Marco et al., 2009; Minogue, 1989), (iii) stratified dispersal (Sapoukhina et al., 2010) and (iv) density-dependence (Bolker and Pacala, 1997, 1999; Dieckmann and Law, 2000). Here, we are interested in another potential reason: group dispersal.

Group dispersal occurs when several entities move or are moved together from one area to another one. This mechanism has been mainly studied for animals, e.g. juvenile tarantulas (Reichling, 2000) and female viscachas (Branch et al., 1993), and for seeds eaten and transported by large mammals (Howe, 1989; Takahashi et al., 2008) or birds (Pizo and Simão, 2001). Here, we are interested in group dispersal of propagules like seeds, pollen grains or spores dispersed by wind. It is assumed that several of these propagules may be released because of a single wind gust, transported in the air into a more or less limited volume and deposited over a more or less limited

area. Such a group of propagules deposited far enough from the primary focus will generate, in a single generation, a secondary focus. Then, because of local reproduction, the secondary focus will be amplified. Intuitively, to generate a secondary focus distinguishable from the primary focus, a group of propagules does not need to be dispersed as far as a single propagule since the group has a numerical advantage that the single individual has not.

The complete observation of groups of very small particles during the dispersal process (release, transport and deposit) is not easy to achieve. However, there are several articles presenting the visualization of the joint release of particles, often due to an explosive discharge; e.g. see Roper et al. (2010) and Trail et al. (2005) for fungal spores, Aylor et al. (2003), Bianchini and Pacini (1996) and Whitaker et al. (2007) for pollen grains, Whitaker and Edwards (2010) and Sundberg (2010) for sphagnum spores. Then, during the transport stage under the wind effect (turbulence), there is no clear evidence that the group formed at the release stage remains in a limited air volume. Experiments assessing the cohesion (or, conversely, the break up) of the group should be carried out. Nevertheless, in favor of the group dispersal hypothesis, there is the special case of particle clumps (e.g. spores attached by a mucilaginous liquid): the dispersal of this sort of groups have been observed until the deposit stage (Aylor and Ferrandino, 1986; Martin et al., 2009). The model that we propose in this article encompasses this special case and the case where the spore groups break up at a determined speed during the transport.

### *1.2. An approach for the modeling of group dispersal*

In propagation models for airborne plant pathogens and plants, propagule transports are usually assumed to be independently and identically drawn from a density probability function called dispersal kernel. The dispersal kernel generally decreases along the radial directions (Austerlitz et al., 2004; Klein et al., 2006a; Tufto et al., 1997) and can be anisotropic (Klein et al., 2003; Soubeyrand et al., 2007, 2008). If group dispersal occurs, then propagule transports are not independent anymore. Therefore, we resort to an assumption reflecting a hierarchical structure of dependence: at the first stage of the hierarchy, groups are independently dispersed; at the second stage, propagules within each group are dispersed independently but conditionally on the group transport.

More explicitly, the group dispersal model (GDM) that we propose relies on the following principles: (i) the numbers of propagules in the groups are independently generated from a counting distribution and the barycenters of the groups are independently transported according to a dispersal kernel; then, (ii) within each group, propagules follow independent Brownian motions. The Brownian motions are centered around the group barycenter and are stopped when the group is deposited. The stopping time is assumed to be proportional to the distance between the source and the deposit location of the group barycenter (the dispersal distance is viewed as a time surrogate or a time proxy). Thus, the further the group is deposited, the larger the deposit area is. This is justified as follows: The paths of the propagules of a single group deviate, time after time, from the group barycenter (these deviations are relativistic). Therefore, the volume of air containing all the propagules

of a single group increases with time. When the parameter determining the intra-group diffusion is zero, the propagules of any group are all located at the group barycenter and the model can generate patterns with clumps.

Based on this model, group dispersal and dispersal under the independence assumption are compared in a spatial context (dispersal from a single point source) and in a spatiotemporal context (multiple generations). Using probabilistic and simulation tools, we especially show that group dispersal is a sufficient condition to generate secondary foci in propagation processes.

## 2. Group dispersal model (GDM) in the case of a single source

### 2.1. Deposit equation for particles

Consider a single point source of particles located at the origin of the planar space  $\mathbb{R}^2$ . Let  $J$  denote the number of groups of particles released by the source and  $N_j$  the number of particles in group  $j \in \{1, \dots, J\}$ .

The deposit location vector  $X_{jn}$  of the  $n$ -th particle of group  $j$  is assumed to satisfy

$$X_{jn} = X_j + B_{jn}(\nu \|X_j\|), \quad (1)$$

where  $X_j$  is the final location vector of the center of group  $j$ ,  $B_{jn}$  is a centered Brownian motion describing the relative movement of the  $n$ -th particle in group  $j$  with respect to the group center,  $\nu$  is a positive parameter and  $\|\cdot\|$  denotes the Euclidean distance.

We use the term “group center” to denote  $X_j$  for the sake of shortness. Indeed,  $X_j$  is not the average of the particle locations  $X_{jn}$ ,  $n = 1, \dots, N_j$ , but only the conditional expectation of  $X_{jn}$  given  $X_j$ , i.e.  $X_j = E(X_{jn} | X_j)$ .

2.2. Assumptions about the deposit equation (1)

The random variables  $J$ ,  $N_j$ ,  $X_j$  and the random processes  $\{B_{jn} : n = 1, \dots, N_j\}$  are mutually independent. It has to be noted that the motion  $B_{jn}$  is independent from  $X_j$  but, obviously, the value of this motion  $B_{jn}(\nu||X_j||)$  at the stopping time  $\nu||X_j||$  is not.

The number of groups  $J$  is Poisson distributed with mean value  $\lambda$ .

The  $N_j$  are independently drawn from the counting distribution  $p_{\mu, \sigma^2}$  defined over  $\mathbb{N}$  with mean and variance parameters  $\mu$  and  $\sigma^2$ , respectively.

The group center locations  $X_j$  are independently and identically drawn from the probability density function (p.d.f.)  $f_{X_j} : \mathbb{R}^2 \mapsto \mathbb{R}_+$ . This function can be characterized by features usually associated with classical dispersal kernels: the decrease of  $f_{X_j}$  at the origin is more or less steep, the tail of  $f_{X_j}$  is more or less heavy, the shape of  $f_{X_j}$  is more or less anisotropic.

The Brownian motions  $B_{jn}$  defined over  $\mathbb{R}^2$  are centered, independent and with independent components. They are stopped at time  $t_j = \nu||X_j||$ . The distance between the source and the location  $X_j$  is used, up to the scaling parameter  $\nu$ , as a time surrogate. Thus, the further a group is transported, the most the particules forming the group are spread with respect to the group center. The value of  $\nu$  determines the strength of the relative spread from the group center. It follows that  $B_{jn}(\nu||X_j||)$  follow independent and centered normal distributions with variance matrices  $\nu||X_j||I$  where  $I$  is the  $2 \times 2$  identity matrix.

Remark: Klein et al. (2003) and Stockmarr (2002) represent particle transports as 3D Brownian motions, the vertical dimension being used to define the times when the motions are stopped. Here, the particle motions



are stopped with the use of the final locations  $X_j$  of the group centers. Appendix A makes the link between the two approaches and motivates the deterministic linear relationship between  $\|X_j\|$  and the stopping times of the Brownian motions  $B_{j_n}$ . It has to be noted here that this relationship could be nonlinear and stochastic and could even depend not only on the distance  $\|X_j\|$  but on the components of  $X_j$ .

### 2.3. Probability density functions characterizing the dispersal

The conditional p.d.f.  $f_{X_{j_n}|X_j}$  of the deposit location  $X_{j_n}$  of a particle given the group center  $X_j$  is the p.d.f. of the normal law with mean vector  $X_j$  and variance matrix  $\nu\|X_j\|I$ :

$$\begin{aligned} f_{X_{j_n}|X_j}(x | y) &= \phi_{\nu,y}(x) \\ &= \frac{1}{2\pi\nu\|y\|} \exp\left(-\frac{(x-y)'(x-y)}{2\nu\|y\|}\right), \end{aligned}$$

where  $(x-y)'$  denotes the transpose vector of  $x-y$ .

Let  $X$  denote the deposit location of a particle whose group it belongs to is unknown. The conditional p.d.f.  $f_{X|\{X_j:j=1,\dots,J\}}$  of  $X$  given the locations of the group centers is the p.d.f. of a mixture of normal distributions, the mixture weights being similar:

$$f_{X|\{X_j:j=1,\dots,J\}}(x | \{x_j\}_j) = \frac{1}{J} \sum_{j=1}^J \phi_{\nu,x_j}(x).$$

Figure 1 shows the density of such a mixture (left) and the deposit locations of the particles (right) obtained by simulating the GDM. The components of the mixture has the same weights ( $1/J$ ) but the further the group center from the point source, the larger the variance of the corresponding mixture component.

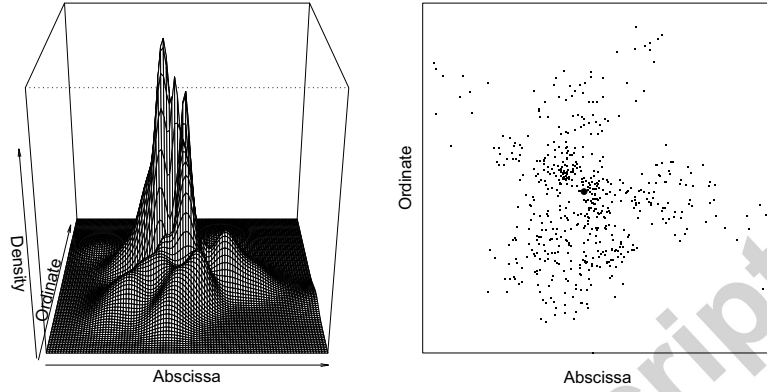


Figure 1: Simulation of the group dispersal model with a single point source. Left: conditional density of the deposit location of a particle whose group it belongs to is unknown given the locations of the group centers. Right: deposit locations of particles obtained under the density shown on the left.

The unconditional p.d.f.  $f_{X_{jn}}$  of  $X_{jn}$ , that is to say the dispersal kernel of any particle, is

$$\begin{aligned} f_{X_{jn}}(x) &= \int_{\mathbb{R}^2} f_{X_{jn}|X_j}(x | y) f_{X_j}(y) dy \\ &= \int_{\mathbb{R}^2} \phi_{\nu,y}(x) f_{X_j}(y) dy. \end{aligned}$$

The particles are not independently but identically distributed (n.i.i.d.) from this p.d.f. while in the classical dispersal models the particles are independently and identically distributed (i.i.d.) from a dispersal kernel which may be of the form of  $f_{X_j}$  or  $f_{X_{jn}}$ .

### 3. Comparison with independent dispersal models (IDM)

In propagation models for airborne plant pathogens and plants, propagule transports are usually assumed to be independent. Thereafter, such models are called independent dispersal models (IDM). This section is devoted to the comparison of our GDM with two IDMs which are particular cases of the GDM:

- IDM1: the number of particles in each group is assumed to be one (i.e.  $\mu = 1$  and  $\sigma^2 = 0$ ). Thus, particles are independently drawn under the p.d.f.  $f_{X_{j_n}}$ .
- IDM2: the number of particles in each group is assumed to be one and the Brownian motions are deleted (i.e.  $\mu = 1$ ,  $\sigma^2 = 0$  and  $\nu = 0$ ). Thus, particles are independently drawn under the p.d.f.  $f_{X_j}$ .

#### 3.1. Analysis of moments

Moments allows the characterization of patterns generated by dynamical models (Bolker and Pacala, 1997, 1999; Dieckmann and Law, 2000). The three models (GDM, IDM1 and IDM2) are compared using six criteria based on moments:

1. The expectation  $E(X)$  of the deposit location  $X$  of any particle;
2. The variance matrix  $V(X)$  of the deposit location of any particle;
3. The expectation of the squared dispersal distance;
4. The expectation of the number of particles  $Q(x + dx)$  deposited in the infinitesimal surface  $x + dx$  centered around  $x$ ;

5. The variance of the number of particles deposited in the infinitesimal surface  $x + dx$  centered around  $x$ ;
6. The covariance between the numbers of particles deposited in the disjoint infinitesimal surfaces  $x_1 + dx$  and  $x_2 + dx$  centered around  $x_1$  and  $x_2$ .

The first three criteria concern the deposit location of a single particle. The last three criteria concern the number of propagules deposited in infinitesimal surfaces. The spatial covariance (criterion 6) is especially interesting to quantify the association or segregation between individuals (Bolker and Pacala, 1999). Table 1 gives the expressions of the criteria for the three models; see Appendix B for the proofs.

We can see that the GDM and the IDM1 which share the same dispersal function also share the same characteristics about the deposit location  $X$  of any single particle (first three criteria). Compared to the IDM2, the GDM and the IDM1 have an additional term in  $V(X)$  and  $E(\|X\|^2)$  because of the extra-dispersal caused by the Brownian motion (whatever the numbers of particles within the groups).

The characteristics about the number of particles deposited in infinitesimal surfaces are, however, different between the three models. The most noticeable differences are on criteria 5 and 6: assume that  $\mu = 1$  (so that the expected number of released particles is the same in the three models), the variance of the local number  $Q(x + dx)$  of deposited particles is greater in the GDM than in the two other models and the spatial covariance is positive under the GDM while it is zero in the two the models.

The extra-variance (criterion 5) as well as the positive spatial covariance (criterion 6) characterizing the GDM induce the occurrence of aggregates

Criterion	Model	Value
$E(X)$	GDM	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
	IDM1	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
	IDM2	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
$V(X)$	GDM	$V(X_j) + \nu E(\ X_j\ )I$
	IDM1	$V(X_j) + \nu E(\ X_j\ )I$
	IDM2	$V(X_j)$
$E(\ X\ ^2)$	GDM	$E(\ X_j\ ^2) + 2\nu E(\ X_j\ )$
	IDM1	$E(\ X_j\ ^2) + 2\nu E(\ X_j\ )$
	IDM2	$E(\ X_j\ ^2)$
$E\{Q(x+dx)\}$	GDM	$\lambda \mu f_{X_{j_n}}(x)dx$
	IDM1	$\lambda f_{X_{j_n}}(x)dx$
	IDM2	$\lambda f_{X_j}(x)dx$
$V\{Q(x+dx)\}$	GDM	$\lambda[\mu f_{X_{j_n}}(x)dx + (\sigma^2 + \mu^2 - \mu)E\{\phi_{\nu, X_j}(x)^2\}(dx)^2]$
	IDM1	$\lambda f_{X_{j_n}}(x)dx$
	IDM2	$\lambda f_{X_j}(x)dx$
$\text{cov}\{Q(x_1+dx), Q(x_2+dx)\}$	GDM	$\lambda(\sigma^2 + \mu^2 - \mu)E\{\phi_{\nu, X_j}(x_1)\phi_{\nu, X_j}(x_2)\}(dx)^2$
	IDM1	0
	IDM2	0

Table 1: Expressions of the six criteria based on moments and used to characterize the patterns generated by the GDM, IDM1 and IDM2. For  $E(X)$  and  $E(\|X\|^2)$   $f_{X_j}$  is assumed to be isotropic. For  $V\{Q(x+dx)\}$ ,  $E\{\phi_{\nu, X_j}(x)^2\} = \int_{\mathbb{R}^2} \phi_{\nu, y}(x)^2 f_{X_j}(y)dy$ . For  $\text{cov}\{Q(x_1+dx), Q(x_2+dx)\}$ ,  $E\{\phi_{\nu, X_j}(x_1)\phi_{\nu, X_j}(x_2)\} = \int_{\mathbb{R}^2} \phi_{\nu, y}(x_1)\phi_{\nu, y}(x_2)f_{X_j}(y)dy$ .

(or clusters) in space after a single generation, while such aggregates are not expected under the IDMs. These aggregates may be at the origin of secondary foci visible after several generations without resorting to a dispersal function with a heavy tail or to spatial heterogeneity.

It has to be noted that for a counting distribution over  $\mathbb{N}$  characterized by mean  $\mu > 0$  and variance  $\sigma^2$ , the quantity  $\sigma^2 + \mu^2 - \mu$  is positive; it is zero if and only if  $\mu = 0$  (which implies that  $\sigma^2 = 0$ ) or  $(\mu, \sigma) = (1, 0)$ . This implies that the covariance given above for the GDM is non-negative.

### 3.2. Concentration of particles

The previous section states that the GDM allows to generate aggregates in a single generation, contrary to the IDM1 (and the IDM2). Below, by investigating the properties of the furthest forward particle, we show that these patterns with aggregates occur with the GDM while the particles remain quite concentrated: under the GDM, the furthest forward particle stays in average closer to the point source than under the IDM1 (when the GDM and the IDM1 share the same marginal dispersal function  $f_{X_{jn}}$ ).

*Distribution of the distance to the furthest particle.* Consider the dispersal of particles from a point source. In what follows, we are interested in the properties of the distance from source to the furthest particle denoted by

$$R^{max} = \max\{R_{jn} : j \in \mathcal{J}, n \in \mathcal{N}_j\},$$

where  $R_{jn} = \|X_{jn}\|$  is the distance between the source (at the origin) and the  $n$ -th deposited particle of group  $j$ ,  $\mathcal{J} = \{1, \dots, J\}$  if  $J > 0$  and the empty set otherwise, and  $\mathcal{N}_j = \{1, \dots, N_j\}$  if  $N_j > 0$  and the empty set otherwise.

By convention, if no particle is dispersed ( $J = 0$  or  $N_j = 0$  for all  $j$ ), then  $R^{max} = 0$ .

Under the three dispersal models, the distribution of the distance between the origin and the furthest deposited particle is zero-inflated and satisfies (see Corollary 1 in Appendix C.1)

$$\begin{aligned} P(R^{max} = 0) &= \exp[\lambda\{p_{\mu,\sigma^2}(0) - 1\}] \\ f_{R^{max}}(r) &= \lambda f_{R_j^{max}}(r) \exp\{\lambda(F_{R_j^{max}}(r) - 1)\}, \quad \forall r > 0, \end{aligned} \quad (2)$$

where  $f_{R_j^{max}}$  is the p.d.f. of the distance  $R_j^{max} = \max\{R_{jn} : n \in \mathcal{N}_j\}$  between the origin and the furthest deposited particle of group  $j$ , and  $F_{R_j^{max}}$  is the corresponding cumulative distribution function ( $F_{R_j^{max}}(r) = P(R_j^{max} = 0) + \int_0^r f_{R_j^{max}}(u)du$ ).

Under the IDMs,  $N_j = 1$  for all  $j \in \mathcal{J}$  and, consequently,  $p_{\mu,\sigma^2}(0) = 0$  and  $f_{R_j^{max}}(r) = f_{R_{jn}}(r)$  in Eq. (2); see Appendix C.2 for details. Under the GDM, the distribution of  $R^{max}$  (Eq. (2)) takes a more complicated form which is provided in Appendix C.3.

*Analytic comparison.* The material provided in the previous paragraph and in the appendices is used in the following to compare the concentration of the particles in the three models (IDM1, IDM2 and GDM).

Here, we provide an analytic comparison of the probability that  $R^{max}$  is larger than a distance  $r > 0$ :

$$P(R^{max} \geq r) = \int_r^{+\infty} f_{R^{max}}(s)ds,$$

whose expression is derived in Appendix C.4 for the three models. We obtained the two following properties (see Appendix C.4 for the proofs):

*Property 1.* Consider a GDM and an IDM1 characterized by the same parameter values except that  $E(J) = \tilde{\lambda}$ ,  $E(N_j) = \tilde{\mu}$  and  $V(N_j) = \sigma^2$  for the GDM, and  $E(J) = \tilde{\lambda}\tilde{\mu}$ ,  $E(N_j) = 1$  and  $V(N_j) = 0$  for the IDM1. Then, for all  $r > 0$  the probability  $P(R^{max} \geq r)$  is lower for the GDM than for the IDM1.

*Property 2.* Consider an IDM1 and an IDM2 characterized by the same parameter values except that  $\nu > 0$  for the IDM1 and  $\nu = 0$  for the IDM2. Then, for all  $r > 0$  the probability  $P(R^{max} \geq r)$  is lower for the IDM2 than for the IDM1.

Property 1 means that for a GDM and an IDM1 characterized by the same dispersal kernel for the particles and the same expected number of dispersed particles  $\lambda\mu$ , the furthest particle under the GDM has less chance to be at a distance greater than any  $r > 0$  than the furthest particle under the IDM1. Therefore, the population of particles is expected to be more concentrated under the GDM than under the IDM1. In other words, the average expansion speed under the GDM is expected to be lower than the average expansion speed under the IDM1.

Property 2 means that the extra-diffusion term included in the IDM1 but not included in the IDM2 implies that the population of particles is expected to be more spread under the IDM1 than under the IDM2.

*Numerical comparison.* Here, we make a numerical comparison between the three models of the expectation of the distance to the furthest particle

$$E(R^{max}) = \int_0^{\infty} r f_{R^{max}}(r) dr.$$



To carry out this comparison, we used the negative binomial distribution for  $p_{\mu,\sigma^2}$  and the exponential kernel over  $\mathbb{R}^2$  parametrized by  $\beta$  for  $f_{X_j}$ . The expressions and the reasons motivating this choice are given in the next section; see equations (3) and (4). Parameter values are provided in Figure 2 which shows that

- $E(R^{max})$  is smaller in the case of the GDM than in the case of the IDM1 with the same marginal dispersal function  $f_{X_{j_n}}$  (this corroborates Property 1);
- $E(R^{max})$  is smaller in the case of the IDM2 than in the case of the IDM1 with additional diffusion of the particles (this corroborates Property 2);
- $E(R^{max})$  decreases with the variance  $\sigma^2$  of the number of particles per group for the GDM;
- Depending on the parameter values,  $E(R^{max})$  can be smaller or higher for the GDM than for the IDM2.

#### 4. Group dispersal model in the case of multiple generations

The analytic results provided above show that group dispersal from a single point source leads to several expected aggregates while independent dispersal does not, even with a fat-tailed dispersal kernel. Therefore, it is expected that group dispersal as defined in the GDM may generate patterns with secondary foci over multiple generations, that is to say when particles play the role of sources of particles once they are deposited.

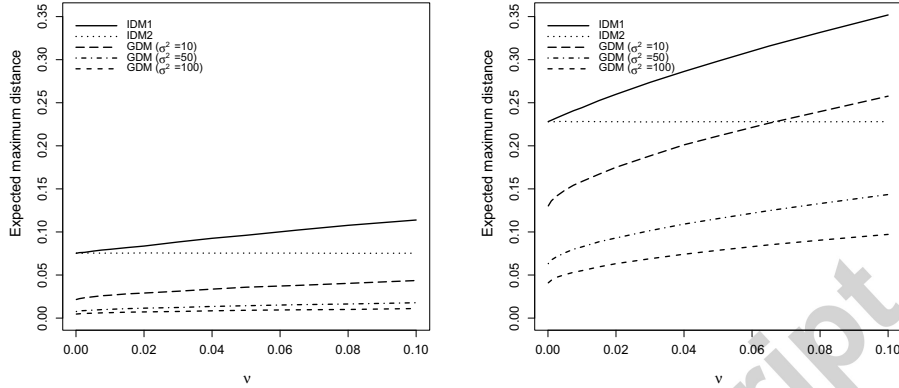


Figure 2: Expected distance to the furthest particle for the IDM1, the IDM2 and three GDMs characterized by different variances  $\sigma^2$  of the number of particles per group. The diffusion parameter ranges from 0 to 0.1 and the other parameters are equal to  $(\lambda, \mu, \beta) = (1, 1, 0.05)$  for the left panel and  $(\lambda, \mu, \beta) = (10, 1, 0.05)$  for the right panel. The curve is constant for the IDM2 because in this case there is no diffusion of the particles with respect to the group center.

To investigate this conjecture, we build in this section a spatiotemporal GDM and assess its ability to generate secondary foci. Then, we compare the GDM to an IDM with long distance dispersal. We also study the modification of the patterns when density-dependence is added to the GDM.

#### 4.1. Spatiotemporal GDM

In the spatiotemporal model, the spread occurs at a discrete time basis in a square domain  $S$  with bottom left corner  $(-1, -1)$  and top right corner  $(1, 1)$ .

The spread is initialized at time zero by a single point source located at the center  $(0, 0)$  of the square  $S$ . Particles are emitted and dispersed from this

source under the GDM. The particles deposited within the square  $S$  become sources of particles while those deposited outside  $S$  are lost.

At the following time step particles are emitted and dispersed under the GDM from each source, i.e. the initial source located at  $(0,0)$  as well as the particles deposited within  $S$ . For each source, the dispersal under the GDM is centered around the source location. The new particles deposited within the square  $S$  become new sources of particles while those deposited outside  $S$  are lost.

The process is then repeated again and again. At each time step the set of sources is the union of the set of sources at the previous time step and the set of particles emitted from these sources and deposited within  $S$ .

Two functions have to be specified to simulate this model: the counting distribution  $p_{\mu,\sigma^2}$  of the number of particles per group and the probability density function  $f_{X_j}$  of the final locations of the group centers. In the following,  $p_{\mu,\sigma^2}$  is the negative binomial distribution with mean parameter  $\mu$  and size parameter  $s = \mu^2/(\sigma^2 - \mu)$ :

$$p_{\mu,\sigma^2}(n) = \frac{\Gamma(n+s)}{\Gamma(s)n!} (\mu/\sigma^2)^s (1 - \mu/\sigma^2)^n. \quad (3)$$

When  $\mu = 1$  and  $\sigma \rightarrow 0^+$ , this distribution converges to the discrete Delta distribution with all the mass at  $n = 1$  and the GDM approximates the IDM1 (Indeed, using Taylor's expansions,  $p_{1,\sigma^2}(1) = 1 + (\sigma^2 \log \sigma^2)(1 + o(1))$  which tends to one when  $\sigma^2 \rightarrow 0^+$ ).

Besides,  $f_{X_j}$  is the exponential kernel defined over  $\mathbb{R}^2$ :

$$f_{X_j}(x) = \frac{1}{2\pi\beta^2} \exp(-\|x\|/\beta), \quad (4)$$

which is not heavy tailed (Klein et al., 2006a) and is not known as a dispersal kernel leading to patterns with multiple foci (Minogue, 1989).

Figure 3 shows simulations of the spatiotemporal versions of the GDM, the IDM1 and the IDM2. For the set of parameters  $(\lambda, \beta, \nu, \mu, \sigma^2)$ , we used  $(1, 0.05, 0.005, 1, 25)$  for the GDM,  $(1, 0.05, 0.005, 1, 0)$  for the IDM1, and  $(1, 0.05, 0, 1, 0)$  for the IDM2. We ran each model until the cumulated number of deposited particles within the square domain  $S$  was greater than 5,000.

#### 4.2. Definition of foci

Generally speaking, a focus is a place where there is a concentration of entities. In the context we consider here —the spread of a population— there are plenty of possibilities for defining a concentration of entities. The concentration has to be related to a threshold representing a standard number of entities but this standard number may be spatially dependent. In addition, the spatial supports used to compute the concentration or the threshold may vary. For this study, we chose to characterize the occurrence of foci under the spatiotemporal GDM, IDM1 or IDM2 with the following procedure.

- We ran the model until the cumulated number of deposited particles within the square domain  $S$  was greater than 5,000 —let  $\mathcal{X}$  denote the spatial point pattern giving the locations of the particles.
- We computed the kernel smoothed intensity of  $\mathcal{X}$  at the nodes of a regular square grid  $\mathcal{G}$  of size  $100 \times 100$  covering the square domain  $S$  with a Gaussian kernel of standard deviation 0.02 (using the `density.ppp()` function of the Spatstat  $R$  package; Baddeley and Turner, 2005).

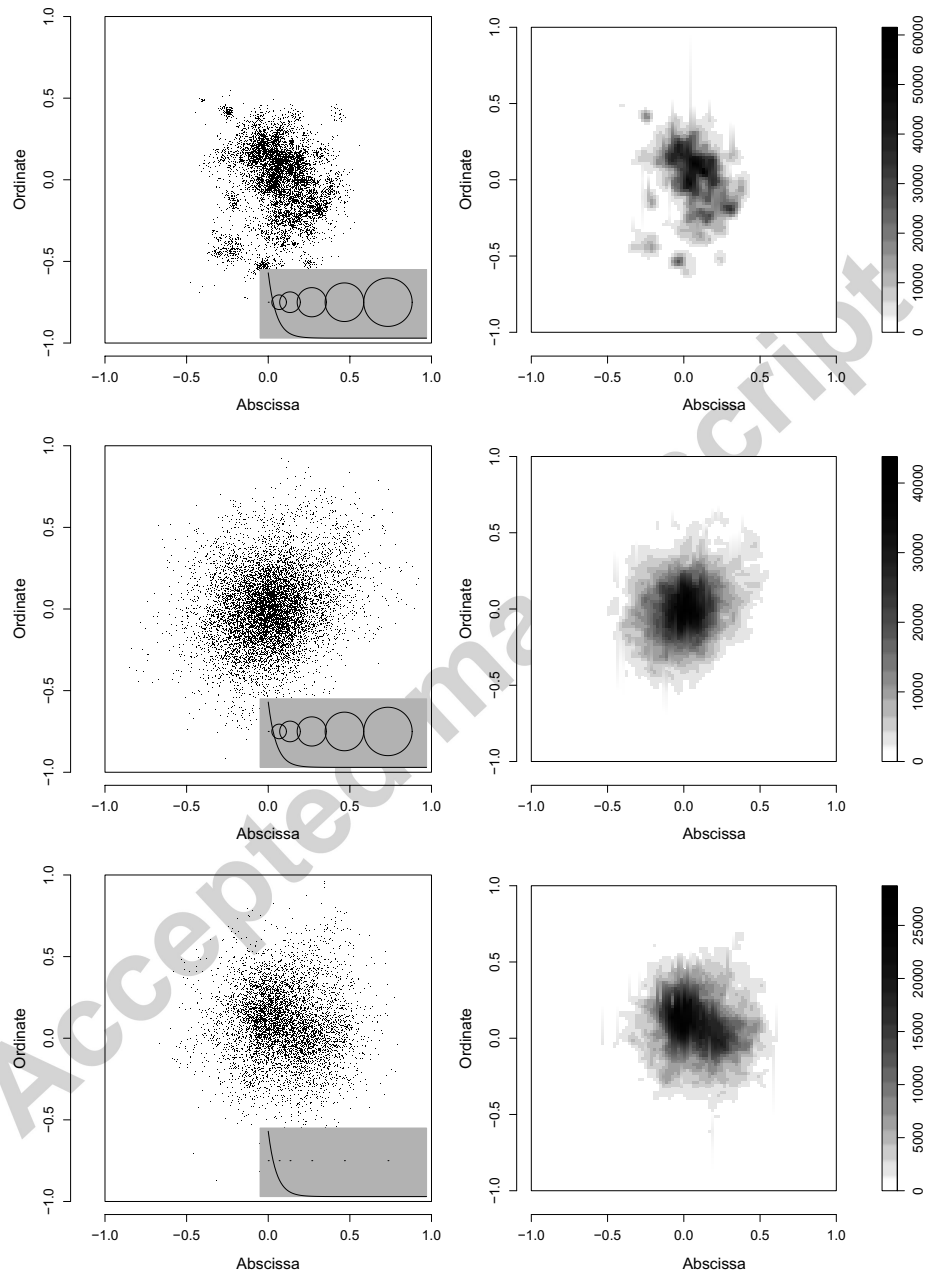


Figure 3: Spatiotemporal simulations of the GDM (first row), the IDM1 (2nd row) and the IDM2 (3rd row). Left: Pattern of particles; Right: Estimated point intensity obtained by kernel smoothing with a Gaussian kernel of standard deviation 0.02. In the grey rectangles of the left panels: shape of the dispersal function used in the simulation and deposit areas with probabilities  $0.95^2$  for the particles of groups deposited at various distances from the source. In the third row these areas are reduced to points (no diffusion).

- For a given threshold value  $\delta$ , we determined the  $\delta$ -nodes, i.e. the nodes of the grid  $\mathcal{G}$  for which the local intensities were greater than  $\delta$ , and determined the sets of connected  $\delta$ -nodes (two  $\delta$ -nodes are connected if they are neighbors in the abscissa or ordinate direction).
- To explore the ability of a model to generate foci of various concentrations, we counted the sets of connected  $\delta$ -nodes for different values of  $\delta$ . For the sake of shortness, sets of connected  $\delta$ -nodes are thereafter called  $\delta$ -foci.

#### 4.3. Numerical comparison of the number of foci

Figure 4 shows, for various parameter values, the average number of  $\delta$ -foci when the threshold  $\delta$  varies. The averages were computed from 1,000 simulations for each set of parameter values. In all the simulations of this paragraph, the dispersal parameter is  $\beta = 0.05$ , the expected number of groups per source and per time step is  $\lambda = 1$  and the expected number of particles per group is  $\mu = 1$ . Thus, the expected number of particles at each generation is the same for the IDM1, the IDM2 and the GDM. Remark: When all the simulations are merged together, the median number of generations required to achieve a population of size equal or greater than 5,000 is 15 (First quartile: 13, third quartile: 24).

Firstly, it has to be noted that the curves obtained under the IDM1 with  $(\nu, \sigma^2) = (0.005, 0)$  and the IDM2 with  $(\nu, \sigma^2) = (0, 0)$  are not significantly different. Thus, the over-dispersal of particles in the IDM1 is not significant when  $\nu = 0.005$  but it can be significant with other values of  $\nu$ .

Secondly, more  $\delta$ -foci are obtained under the GDMs with diffusion param-

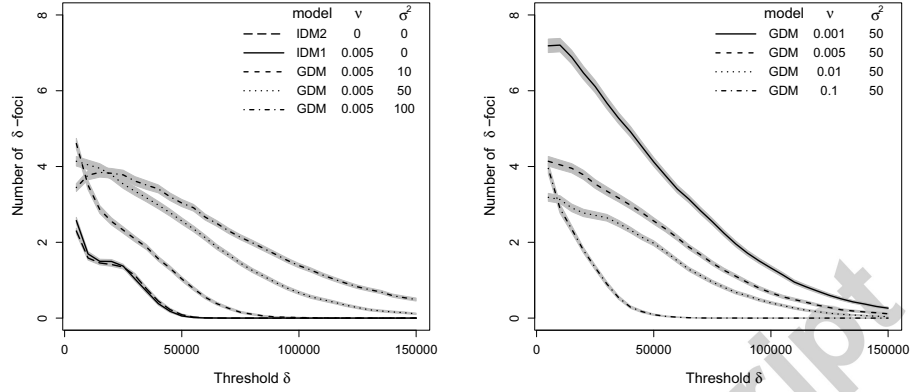


Figure 4: Variation in the average number of  $\delta$ -foci (sets of connected nodes with estimated intensities greater than  $\delta$ ) with respect to the threshold  $\delta$ . These variations are shown for various values of  $\nu$  and  $\sigma^2$  (see the top right corners of the panels), the other parameters being fixed at  $(\lambda, \mu, \beta) = (1, 1, 0.05)$ . The averages are computed from 1,000 simulations and the grey zones provide the 95%-confidence envelopes for the averages.

eter  $\nu = 0.005$  (curves corresponding to  $\sigma^2 > 0$  in the left panel) than under the IDM1 with the same diffusion parameter. In addition, more  $\delta$ -foci with high concentrations of particles are generated when  $\sigma^2$  is large (the emission of large groups of particles is more probable with large  $\sigma^2$  than with small  $\sigma^2$ ).

Thirdly, the increase of  $\nu$  (increase of the diffusion within the groups) has a negative effect on the generation of  $\delta$ -foci; see the right panel of figure 4. The reason is that the particles of a group whose barycenter is deposited at a given distance are less concentrated when  $\nu$  is large than when it is low.

#### 4.4. GDM versus IDM with long distance dispersal (IDM-LDD)

The main difference between the two models is that, in one generation, the GDM can generate particle clusters whereas the IDM-LDD cannot (see the analysis of moments in section 3). This is the reason why the comparison between the GDM and the IDM-LDD is proposed for the spatiotemporal versions of these models.

An exhaustive comparison of the GDM and the IDM-LDD would require much more place than a single paragraph because there are many GDMs (using various dispersal kernels) and IDMs with LDD (Austerlitz et al., 2004; Klein et al., 2006b; Tufto et al., 1997). Here, we only compare the GDM with the exponential kernel given by Equation (4) to the IDM2 with a logistic kernel parameterized by  $\beta_{LDD}$  (Klein et al., 2006b):

$$f_{X_j}(x) = \frac{2\beta_{LDD}}{\pi^2(1 + \beta_{LDD}^2 \|x\|^4)}.$$

The logistic kernel is a power-law kernel and is heavy-tailed. For the GDM, the parameter vector is set to  $(\lambda, \beta, \nu, \mu, \sigma^2) = (1, 0.05, 0.005, 1, 25)$ ; a realization is provided in Figure 3 (top). For the IDM2 the parameter vector is set to  $(\lambda, \beta_{LDD}, \nu, \mu, \sigma^2) = (1, 2300, 0, 1, 0)$ ; a realization is given in Figure 5.

With these sets of parameters, the two models have the same median dispersal distance in one generation ( $Med_{GDM}(\|X\|) = Med_{IDM2-LDD}(\|X\|) \approx 0.0099$ ). However, the mean dispersal distance in one generation is smaller for the GDM than for the IDM2-LDD ( $E_{GDM}(\|X\|) \approx 0.0091$  and  $E_{IDM2-LDD}(\|X\|) \approx 0.0110$ ); the first quartile of the dispersal distance in one generation is larger for the GDM than for the IDM2-LDD ( $Q_{GDM}^1(\|X\|) \approx 0.0073$  and  $Q_{IDM2-LDD}^1(\|X\|) \approx 0.0051$ ); the expectation of the distance to the furthest particle in one gener-



ation is lower for the GDM than for the IDM2–LDD ( $E_{GDM}(R^{max}) \approx 0.015$  and  $E_{IDM2--LDD}(R^{max}) \approx 0.023$ ).

This example allows us to see a noticeable difference in the clustered patterns generated by the two models: compared with the GDM, the IDM2–LDD induces, most of the time, a higher concentration of particles close to the primary source (lower first quartile) and foci further from the source (large expectation of  $R^{max}$ ). In Figure 5 for the IDM2–LDD, we can see a clear separation between the primary focus and the secondary foci whereas, in Figure 3 for the GDM, the secondary foci are in the slope of the primary focus.

#### 4.5. Density-dependence

In the previous simulations, there is no limit to the growth of the population except the domain boundaries: We used a simple birth process without death and interaction. In this paragraph, we introduce in the model a simple sort of interaction (a density-dependence) causing the immediate death of some particles and we investigate how the ability of the model to generate foci bears the density-dependence.

The GDM was modified as follows: We set a regular square grid made of cells with side  $\rho$  and covering the square domain  $S$  and, at each generation and for each occupied grid cell, we deleted all the particles except one among the old and new particles (this is a non-local density-dependence with maximum range  $\rho\sqrt{2}$ ). Thus, at each generation there is either zero or one particle per cell. This sort of density-dependence may occur in the case of resource limitation.

We carried out simulations under this spatiotemporal GDM incorporating

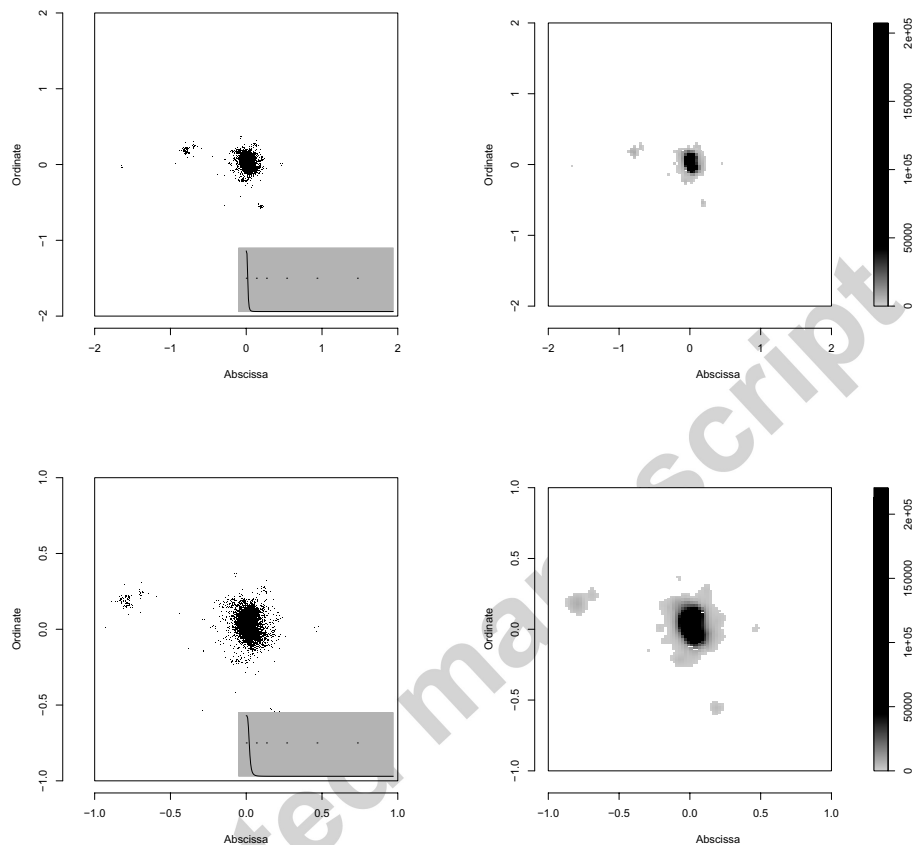


Figure 5: Spatiotemporal simulation of the IDM2-LDD. Top left: Pattern of particles; Top right: Estimated intensity of the point pattern. The plots were constructed as in Figure 3 except that the domain is  $[-2, 2] \times [-2, 2]$  to take into account long distance dispersal events, and that the grey scale for the intensity map is nonlinear to take into account the very high concentration of particles near the source point. Bottom panels: plots drawn at the top restricted to the domain  $[-1, 1] \times [-1, 1]$  to allow the comparison with plots of Figure 3.

density-dependence for various parameter values and applied the same procedure than above to investigate the ability of this model to generate foci. Figure 6 shows how the curve of the average number of  $\delta$ -foci against the threshold  $\delta$  is modified in the presence of density-dependence in the case of independent dispersal (left panel) and group dispersal (right panel). All the simulations were performed with  $(\lambda, \mu, \nu, \beta) = (1, 1, 0.005, 0.05)$  while  $\sigma^2$  and  $\rho$  were varied. Remark: The continuous lines in figure 6 used as benchmarks (no density dependence, i.e.  $\rho = 0$ ) were already given in the left panel of figure 4 but the range of the abscissa was modified.

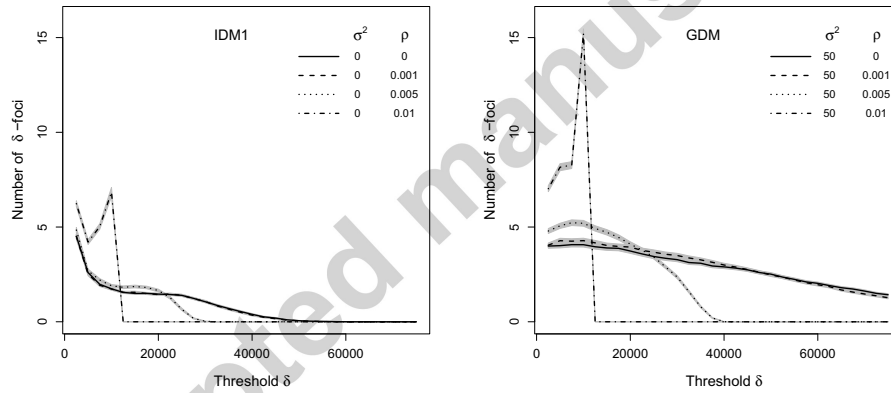


Figure 6: Variation in the average number of  $\delta$ -foci (sets of connected nodes with estimated intensities greater than  $\delta$ ) with respect to the threshold  $\delta$  in the presence of density-dependence ( $\rho > 0$ ) for the IDM1 (left panel,  $\sigma^2 = 0$ ) and the GDM (right panel,  $\sigma^2 = 50$ ). For the two models  $(\lambda, \mu, \nu, \beta) = (1, 1, 0.005, 0.05)$  and the other parameters are specified in the top right corners of the panels. The averages are computed from 1,000 simulations and the grey zones provide the 95%-confidence envelopes for the averages.

Qualitatively, the effect of density-dependence is the same under the

IDM1 and the GDM: the number of  $\delta$ -foci is increased for small  $\delta$  and decreased for large  $\delta$ . This observation is patent for large  $\rho$ . Let us explain these observations. Because of the density-dependence, there is less grid nodes with high intensities of particles and, consequently, less  $\delta$ -foci with large  $\delta$ . For small  $\delta$ , the  $\delta$ -foci which existed when  $\rho = 0$  remain almost unchanged when  $\rho > 0$ , since density-dependence has a small effect on low-density regions. Besides, some large- $\delta$ -foci which existed when  $\rho = 0$  are converted into small- $\delta$ -foci when  $\rho > 0$ ; this conversion explains the increase of  $\delta$ -foci for small  $\delta$  when  $\rho$  increases.

## 5. Discussion

### 5.1. Role of group dispersal in population dynamics

Determining the roles played by some components of population dynamics is of primary importance for species conservation and pest control. For example, there exist numerous studies highlighting the consequences of specific habitat spatial structures, environmental heterogeneities, genetic structures and dispersal modes on population spread and population persistence (see e.g. Filipe and Maule, 2004; Hamel et al., 2010; Kendall and Fox, 1998; Plotkin et al., 2006; Real and Biek, 2007; Roques and Stoica, 2007; Smith, 2006). The work presented in this article is in this vein since we considered a specific dispersal mode, namely group dispersal, and studied its consequences on the population spread. We showed that group dispersal, as it is defined in this article, can generate multiple foci without introducing long-distance dispersal, spatial heterogeneity or density-dependence. This property holds while the population can remain particularly concentrated. In addition, this

property can persist with competition between individuals.

### *5.2. Generation of patchy patterns*

Many dispersal or diffusion models can generate patchy patterns (see the introduction and the next paragraph). The GDM is one of these models but its interest depends on the situation. If the aim is to reproduce patterns with multiple foci in a case where we do not know enough about how the agents disperse, then more parsimonious models (e.g. IDMs with long distance dispersal) may fit the population data at least equally well and a model selection procedure taking the number of parameters into account could be carried out; see Section 5.4. If the aim is to reproduce patchy patterns generated by dynamics characterized by group dispersal, then the GDM may be a realistic quasi-mechanistic model adapted to the situation, in particular in the case of a single generation. Besides, in such a situation, the use of the GDM could lead to better interpretation of parameters and improved predictions.

### *5.3. Comparison with other approaches*

Pattern formation has been a core topic in mathematical biology (see e.g. Ben-Jacob et al., 2001; Cannas et al., 2006; Mimura et al., 2000; Mimura, 2004; Molofsky, 1994; Murray et al., 1998). This type of approaches is often based on systems of several reaction-diffusion equations and the pattern formation is for example due to species-nutrient interactions (Mimura, 2004), multispecies interactions (Morozov and Petrovskii, 2009) or chemotactic interactions between particles (Keller and Segel, 1970).

Nevertheless, patchy patterns can be achieved with single-species reaction-diffusion models such as the model of Murray and Spurb (1983). This model is based on a reaction-diffusion equation with a convection term which plays a role comparable to the dispersal kernel associated with group centers in the GDM; and a diffusion term which has a role comparable to the Brownian motion of the particles in the GDM. Despite these similarities, the way the foci are generated is very different. Indeed, a key ingredient in the model of Murray and Spurb (1983) is the boundary effect: in their case, complex pattern formation (i.e. with more than one focus) cannot appear with lethal boundaries. Conversely, in the GDM, boundaries are lethal and foci formation is therefore only driven by dispersal.

In this article we mentioned the IDMs with long distance dispersal (LDD) which allow the generation of patchy patterns. The models we had in mind are based on spatial point processes as in Section 4.4 and in Austerlitz et al. (2004), Klein et al. (2006a), Minogue (1989) and Tufto et al. (1997). There also exist models including LDD which are based on Gaussian puff or differential equations (Aylor, 1987, 2003; Ferrandino, 1993). In these models, which take the vertical dimension into account, the turbulence allows the particles to escape the canopy and to be transported over long distances.

The approach presented by Bolker and Pacala (1997, 1999) and Dieckmann and Law (2000), based on point processes, also allows the creation of patchy patterns. However, these patterns are achieved thanks to an interaction kernel included in their time-continuous model and governing density-dependent death and competition.

#### 5.4. *Statistical inference*

This study shows that group dispersal is one of the causes which can be invoked to explain patchy patterns. Thus, one may be interested in testing if the group dispersal hypothesis is not rejected based on observed patterns. Besides, if this hypothesis is not rejected, one may be interested in estimating the GDM parameters. The characteristics and properties that we provided in this article could be used to develop testing and estimation procedures. For example, a moment method could be developed using the material provided in Section 3. Alternatively, the inference tools developed for mixture models (Fraley and Raftery, 2002) and spatial point processes (Illian et al., 2008; Stoyan et al., 1995) could be exploited.

For the reasons explained in Section 5.2, issues of model selection can arise. Since the IDM is a special case of the GDM (setting  $(\mu, \sigma^2) = 0$  in the GDM leads to an IDM), procedures for nested models as the likelihood ratio test and the Akaike's criterion can be used, eventually associated with bootstrap. For nonnested models (e.g. the GDM and the model of Dieckmann and Law, 2000), a procedure based on a prediction criterion and cross validation can be used; see Burnham and Anderson (2002) for details about model selection.

#### 5.5. *Generalization of the model*

The model that we proposed benefits from the simplicity and flexibility of parametric dispersal kernels and may become a useful model for dispersal studies. Several generalizations can be proposed to extend the relevancy of the model. One can use various probability distributions for the number of

groups, the number of individuals per group and the locations of group centers. These distributions can depend on exogenous factors which may depend on time and space. Moreover, the stopping times of the Brownian motions can be changed (see Appendix A) and the Brownian motions themselves can be transformed into other diffusive processes. In the spatio-temporal GDM as it stands in this article, time is discrete and a new generation occurs at each time step. Time could be continuous by using spatio-temporal Poisson point processes (like in Soubeyrand et al., 2009) and overlapping generations could arise. Furthermore, the dispersal events occurring at the same time could be spatially correlated to take into account the similarity of environmental conditions affecting neighbor sources. Then, we could analyze the effects of these modifications on the properties of the GDM as we analyzed the effect of density-dependence. Both analytical and simulation procedures could be carried out for these analyses.

*Acknowledgements.* We thank the Editor and the reviewers for their comments and suggestions. This research was supported by a grant of the Plant Health and Environment division of INRA (Project N. 2010-546-1).

## References

- Austerlitz, F., Dick, C. W., Dutech, C., Klein, E. K., Oddou-Muratorio, S., Smouse, P. E., Sork, V. L., 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13, 937–954.
- Aylor, D. E., 1987. Deposition gradients of urediniospores of *Puccinia recondita* near a source. *Phytopathology* 77, 1442–1448.



- Aylor, D. E., 2003. Spread of plant disease on a continental scale: role of aerial dispersal of pathogens. *Ecology* 84, 1989–1997.
- Aylor, D. E., Ferrandino, F. J., 1986. Germination of urediniospore clusters of *Uromyces appendiculatus* and *Puccinia recondita*. *Transactions of the British Mycological Society* 86, 591–595.
- Aylor, D. E., Schultes, N. P., Shields, E. J., 2003. An aerobiological framework for assessing cross-pollination in maize. *Agricultural and Forest Meteorology* 119, 111–129.
- Baddeley, A., Turner, R., 2005. Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* 12, 1–42.
- Ben-Jacob, E., Cohen, I., Golding, I., Kozlovsky, Y., 2001. Modeling branching and chiral colonial patterning of lubricating bacteria. In: Maini, P. K., Othmer, H. G. (Eds.), *Mathematical Models for Biological Pattern Formation*, IMA Volumes. Springer-Verlag.
- Bianchini, M., Pacini, E., 1996. Explosive anther dehiscence in *ricinus communis* L. involves cell wall modifications and relative humidity. *International Journal of Plant Sciences* 157, 739–745.
- Bolker, B. M., Pacala, S. W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology* 52, 179–197.
- Bolker, B. M., Pacala, S. W., 1999. Spatial moment equations for plant competitions: Understanding spatial strategies and the advantage of short dispersal. *The American Naturalist* 153, 575–602.

- Branch, L. C., Villareal, D., Fowler, G. S., 1993. Recruitment, dispersal, and group fusion in a declining population of the plains vizcacha (*Lagostomus maximus*; Chinchillidae). *Journal of Mammalogy* 74, 9–20.
- Burnham, K. P., Anderson, D. R., 2002. *Model Selection and Multimodel Inference* (2nd Ed.). Springer-Verlag, New York.
- Cannas, S. A., Marcob, D. E., Montemurro, M. A., 2006. Long range dispersal and spatial pattern formation in biological invasions. *Mathematical Biosciences* 203, 155–170.
- David, H. A., Nagaraja, H. N., 2003. *Order Statistics* (3rd Ed.). Wiley, Hoboken.
- Dieckmann, U., Law, R., 2000. Relaxation projections and the method of moments. In: Dieckmann, U., Law, R., Metz, J. A. J. (Eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge University Press.
- Ferrandino, F. J., 1993. Dispersive epidemic waves: I. Focus expansion within a linear planting. *Phytopathology* 83, 795–802.
- Filipe, J. A. N., Maule, M. M., 2004. Effects of dispersal mechanisms on spatio-temporal development of epidemics. *Journal of Theoretical Biology* 226, 125–141.
- Fraley, C., Raftery, A. E., 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97, 611–631.

- Hamel, F., Fayard, J., Roques, L., 2010. Spreading speeds in slowly oscillating environments. *Bulletin of Mathematical Biology* 72, 1166–1191.
- Heesterbeek, J. A. P., Zadoks, J. C., 1987. Modelling pandemics of quarantine pests and diseases: problems and perspectives. *Crop Protection* 6, 211–221.
- Howe, H. F., 1989. Scatter- and clump-dispersal and seedling demography: hypothesis and implications. *Oecologia* 79, 417–426.
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley.
- Kauffman, M. J., Jules, E. S., 2006. Heterogeneity shapes invasion: host size and environment influence susceptibility to a nonnative pathogen. *Ecological Applications* 16, 166–175.
- Keller, E., Segel, L., 1970. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology* 26, 399–415.
- Kendall, B. E., Fox, G. A., 1998. Spatial structure, environmental heterogeneity, and population dynamics: Analysis of the coupled logistic map. *Theoretical Population Biology* 54, 11–37.
- Klein, E. K., Lavigne, C., Foueillassar, X., Gouyon, P.-H., Larédo, C., 2003. Corn pollen dispersal: quasi-mechanistic models and field experiments. *Ecological Monographs* 73, 131–150.
- Klein, E. K., Lavigne, C., Gouyon, P.-H., 2006a. Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology* 6, 3.

- Klein, E. K., Lavigne, C., Picault, H., Renard, M., Gouyon, P.-H., 2006b. Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology* 43, 141–151.
- Laine, A.-L., Hanski, I., 2006. Large-scale spatial dynamics of specialist plant pathogen. *Journal of Ecology* 94, 217–226.
- Marco, D. E., Cannas, S. A., Montemurro, M. A., Hu, B., Cheng, S.-H., 2009. Comparable ecological dynamics underlie early cancer invasion and species dispersal, involving self-organizing processes. *Journal of Theoretical Biology* 256, 65–75.
- Martin, M. D., Chamecki, M., Brush, G. S., Meneveau, C., Parlange, M. B., 2009. Pollen clumping and wind dispersal in an invasive angiosperm. *American Journal of Botany* 96, 1703–1711.
- Mimura, M., 2004. Pattern formation in consumer-finite resource reaction-diffusion systems. *Publications of the Research Institute for Mathematical Sciences* 40, 14131431.
- Mimura, M., Sakaguchi, H., Matsushita, M., 2000. Reaction-diffusion modelling of bacterial colony patterns. *Physica A* 282, 283–303.
- Minogue, K. P., 1989. Diffusion and spatial probability models for disease spread. In: Jeger, M. J. (Ed.), *Spatial Components of Plant Disease Epidemics*. Prentice Hall, pp. 127–143.
- Molofsky, J., 1994. Population dynamics and pattern formation in theoretical populations. *Ecology* 75, 30–39.

- Morozov, A., Petrovskii, S., 2009. Excitable population dynamics, biological control failure, and spatiotemporal pattern formation in a model ecosystem. *Bulletin of Mathematical Biology* 71, 863–887.
- Mörters, P., Peres, Y., 2010. *Brownian Motions*. Cambridge University Press, Cambridge.
- Murray, J. D., Cook, J., Tyson, R., Lubkin, S. R., 1998. Spatial pattern formation in biology: I. dermal wound healing. ii. bacterial patterns. *Journal of the Franklin Institute* 335, 303–332.
- Murray, J. D., Sperb, R. P., 1983. Minimum domains for spatial patterns in a class of reaction diffusion equations. *Journal of Mathematical Biology* 18, 169–184.
- Nash, D. R., Agassiz, D. J. L., Godfray, H. C. J., Lawton, J. H., 1995. The pattern of spread of invading species: two leaf-mining moths colonizing Great Britain. *Journal of Animal Ecology* 64, 225–233.
- Pizo, M. A., Simão, I., 2001. Seed deposition patterns and the survival of seeds and seedlings of the palm *Euterpe edulis*. *Acta Oecologica* 22, 229–233.
- Plotkin, J. B., Shave, J., Ashton, P. S., 2006. Cluster analysis of spatial patterns in malaysian tree species. *The American Naturalist* 160, 629–644.
- Real, L. A., Biek, R., 2007. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *Journal of the Royal Society Interface* 4, 935–948.

- Reichling, S. B., 2000. Group dispersal in juvenile *Brachypelma vagans* (Aranae, Theraphosidae). *The Journal of Arachnology* 28, 248–250.
- Roper, M., Seminara, A., Bandi, M. M., Cobb, A., Dillard, H. R., Pringle, A., 2010. Dispersal of fungal spores on a cooperatively generated wind. *PNAS* 107, 17474–17479.
- Roques, L., Stoica, R., 2007. Species persistence decreases with habitat fragmentation: an analysis in periodic stochastic environments. *Journal of Mathematical Biology* 55, 189–205.
- Sapoukhina, N., Tyutyunov, Y., Sache, I., Arditi, R., 2010. Spatially mixed crops to control the stratified dispersal of airborne fungal diseases. *Ecological Modelling* 221, 2793–2800.
- Shigesada, N., Kawasaki, K., 1997. *Biological Invasions: Theory and Practice*. Oxford University Press, Oxford.
- Smith, G. C., 2006. Persistence of disease in territorial animals: insights from spatial models of tb. *New Zealand Journal of Ecology* 30, 35–41.
- Soubeyrand, S., Enjalbert, J., Sache, I., 2008. Accounting for roughness of circular processes: Using gaussian random processes to model the anisotropic spread of airborne plant disease. *Theoretical Population Biology* 73, 92–103.
- Soubeyrand, S., Enjalbert, J., Sanchez, A., Sache, I., 2007. Anisotropy, in density and in distance, of the dispersal of yellow rust of wheat: Experiments in large field plots and estimation. *Phytopathology* 97, 1315–1324.

- Soubeyrand, S., Laine, A.-L., Hanski, I., Penttinen, A., 2009. Spatio-temporal structure of host-pathogen interactions in a metapopulation. *The American Naturalist* 174, 308–320.
- Stockmarr, A., 2002. The distribution of particles in the plane dispersed by a simple 3-dimensional diffusion process. *Journal of Mathematical Biology* 45, 461–469.
- Stoyan, D., Kendall, W. S., Mecke, J., 1995. *Stochastic Geometry and its Applications*, 2nd Ed. Wiley, Chichester.
- Sundberg, S., 2010. Size matters for violent discharge height and settling speed of *Sphagnum* spores: important attributes for dispersal potential. *Annals of Botany* 105, 291–300.
- Takahashi, K., Shiota, T., Tamatani, H., Koyoma, M., Washitani, I., 2008. Seasonal variation in fleshy fruit use and seed dispersal by the Japanese black bear (*Ursus thibetanus japonicus*). *Ecological Research* 23, 471–478.
- Trail, F., Gaffoor, I., Vogel, S., 2005. Ejection mechanisms and trajectory of the ascospores of *Gibberella zeae* (anamorph *Fuarium graminearum*). *Fungal genetics and biology* 42, 528–533.
- Tufto, J., Engen, S., Hindar, K., 1997. Stochastic dispersal processes in plant populations. *Theoretical Population Biology* 52, 16–26.
- Whitaker, D. L., Edwards, J., 2010. *Sphagnum* moss disperses spores with vortex rings. *Science* 329, 406.

Whitaker, D. L., Webster, L. A., Edwards, J., 2007. The biomechanics of *Corvus canadensis* stamens are ideal for catapulting pollen vertically. *Functional Ecology* 21, 219–225.

### A. Stopping times of the Brownian motions

In the deposit equation for particles defined by Eq. (1), i.e.  $X_{jn} = X_j + B_{jn}(\nu||X_j||)$ , there is a linear relationship between the deposit distance of the group center and the transport duration after which the Brownian motions are stopped: The transport duration of group  $j$ , say  $T_j$ , satisfies  $T_j = \nu||X_j||$ . Thus, the relationship between  $T_j$  and  $||X_j||$  is deterministic and linear. It can be noted that, if  $f_{X_j}$  is isotropic then the p.d.f. of  $T_j$  satisfies

$$\begin{aligned} f_{T_j}(t) &= (1/\nu)f_{||X_j||}(t/\nu) \\ &= (2\pi t/\nu^2)f_{X_j}\{t/\nu, 0\}, \end{aligned}$$

where  $f_{||X_j||}$  is the p.d.f. of  $||X_j||$ .

More generally, the deposit equation could be written:

$$X_{jn} = X_j + B_{jn}(T_j),$$

with  $X_j$  and  $T_j$  being related deterministically or not, linearly or not.

For example, consider the case where the group centers follow centered Brownian motions with independent components in the 3D space and where the stopping times, which determine the locations  $X_j$ , are defined with the vertical component: The particles are released at the height  $h > 0$  and are deposited once they hit the ground at the height 0. Then, in this case, the



relationship between  $T_j$  and  $\|X_j\|$  is not deterministic but stochastic and is nonlinear in expectation:

$$E(T_j \mid \|X_j\| = r) = h^2 + r^2. \quad (5)$$

Indeed,

$$\begin{aligned} E(T_j \mid \|X_j\| = r) &= \int_0^{+\infty} t f_{T_j \mid \|X_j\|}(t \mid r) dt \\ &= \int_0^{+\infty} t \frac{f_{\|X_j\| \mid T_j}(r \mid t) f_{T_j}(t)}{f_{\|X_j\|}(r)} dt, \end{aligned}$$

where the p.d.f. in the preceding line satisfy (see Mörters and Peres, 2010, especially chapter 2 including material about hitting times of one-dimensional Brownian motions):

$$\begin{aligned} f_{\|X_j\| \mid T_j}(r \mid t) &= (r/t) \exp(-r^2/2t) \\ f_{T_j}(t) &= \left( h/\sqrt{2\pi t^3} \right) \exp(-h^2/2t) \\ f_{\|X_j\|}(r) &= (r/h^2)(1 + r^2/h^2)^{-3/2}. \end{aligned}$$

Using these formulae, the above integral leads to Eq. (5). It has to be noted that this result is obtained without drift and gravity and without scaling parameters contrary to the results provided in Stockmarr (2002) and Klein et al. (2003). Here, the marginal p.d.f. of  $X_j$  is a bivariate Student's  $t$  ( $2Dt$ ) distribution:  $f_{X_j}(x) = (1 + \|x\|^2/h^2)^{-3/2}/2\pi h^2$ .

Now, assume that the stopping time  $T_j$  has the same marginal p.d.f. than above but that the conditional p.d.f. of  $\|X_j\|$  given  $T_j$  is exponentially distributed:

$$f_{\|X_j\| \mid T_j}(r \mid t) = (1/t) \exp(-r/t).$$

Then, the relationship between  $T_j$  and  $\|X_j\|$  is still stochastic but linear in expectation:

$$E(T_j \mid \|X_j\| = r) = h^2 + r/2.$$

In the simulations carried out for the preparation of this article, the results were qualitatively similar when we used a deterministic linear relationship between  $T_j$  and  $\|X_j\|$  and a deterministic quadratic relationship.

## B. Expressions of moments: Proofs

*Criterion 1.*  $E(X) = (0, 0)'$  holds if the marginal dispersal function  $f_{X_{j_n}}$  for the GDM and the IDM1 or  $f_{X_j}$  for the IDM2 is isotropic. This condition is satisfied for the three models when  $f_{X_j}$  and the Brownian motion are isotropic.

*Criterion 2.* For the IDM2,  $X = X_j$  and, therefore,  $V(X) = V(X_j)$ . For the IDM1 and the GDM,  $X = X_j + \sqrt{\nu\|X_j\|}Y$ , where  $Y$  is a centered bivariate normal vector with identity variance-covariance matrix  $I$ . Then, using the conditional variance formula leads to the result:

$$\begin{aligned} V(X) &= V(E(X|X_j)) + E(V(X|X_j)) \\ &= V(X_j) + E(\nu\|X_j\|I) \\ &= V(X_j) + \nu E(\|X_j\|)I. \end{aligned}$$

*Criterion 3.* For the IDM2,  $X = X_j$  and, therefore,  $E(\|X\|^2) = E(\|X_j\|^2)$ . For the IDM1 and the GDM, denote the bivariate vectors  $X$  by  $X = (X_{(a)}, X_{(b)})'$

and  $X_j$  by  $X = (X_{j,(a)}, X_{j,(b)})'$ . Then,

$$\begin{aligned}
 E(\|X\|^2) &= E(X_{(a)}^2 + X_{(b)}^2) \\
 &= V(X_{(a)}) + E(X_{(a)})^2 + V(X_{(b)}) + E(X_{(b)})^2 \\
 &= V(X_{(a)}) + V(X_{(b)}) \quad (\text{using criterion 1}) \\
 &= V(X_{j(a)}) + V(X_{j(b)}) + 2\nu E(\|X_j\|) \quad (\text{using criterion 2}) \\
 &= E(X_{j(a)}^2) + E(X_{j(b)}^2) + 2\nu E(\|X_j\|) \quad (\text{using criterion 1}) \\
 &= E(\|X_j\|^2) + 2\nu E(\|X_j\|)
 \end{aligned}$$

*Criteria 4, 5 and 6 for the IDM1 and the IDM2.* For the IDM1 and the IDM2, the particle locations form inhomogeneous Poisson point processes in  $\mathbb{R}^2$  with intensity functions  $\lambda f_{X_{j_n}}$  and  $\lambda f_{X_j}$ , respectively. Therefore, from the basic properties of such processes (Illian et al., 2008, sec. 3.4.1),  $E\{Q(x+dx)\} = V\{Q(x+dx)\} = \lambda f_{X_{j_n}}(x)dx$  for the IDM1,  $E\{Q(x+dx)\} = V\{Q(x+dx)\} = \lambda f_{X_j}(x)dx$  for the IDM2, and  $\text{cov}\{Q(x_1+dx), Q(x_2+dx)\} = 0$  for both models.

*Criterion 4 for the GDM.* Let  $Q_j(x+dx) = \sum_{n=1}^{N_j} \mathbf{1}(X_{j_n} \in x+dx)$  be the number of particles of group  $j$  deposited in the infinitesimal surface  $x+dx$ , with  $\mathbf{1}(\cdot)$  denoting the indicator function. Then,  $Q(x+dx) = \sum_{j=1}^J Q_j(x+dx)$

and

$$\begin{aligned}
E\{Q(x+dx)\} &= E[E\{\sum_{j=1}^J Q_j(x+dx) \mid J\}] \\
&= E[\sum_{j=1}^J E\{Q_j(x+dx)\}] \\
&= E[JE\{E(\sum_{n=1}^{N_j} \mathbf{1}(X_{jn} \in x+dx) \mid N_j)\}] \\
&= \lambda E[N_j E\{\mathbf{1}(X_{jn} \in x+dx)\}] \\
&= \lambda \mu f_{X_{j_n}}(x) dx.
\end{aligned}$$

*Criterion 5 for the GDM.* For the variance  $V\{Q(x+dx)\} = E\{Q(x+dx)^2\} - E\{Q(x+dx)\}^2$  we use the same tricks than above. Using the independence between  $Q_j(x+dx)$  and  $Q_{j'}(x+dx)$  when  $j \neq j'$  for line 2, the value  $E(J(J-1)) = \lambda^2$  for line 3 and the value  $E(N_j(N_j-1)) = \sigma^2 + \mu^2 - \mu$  for line 5,

$$\begin{aligned}
E\{Q(x+dx)^2\} &= E\{\sum_{j=1}^J Q_j(x+dx)^2 + \sum_{j=1}^J \sum_{j'=1, j' \neq j}^J Q_j(x+dx) Q_{j'}(x+dx)\} \\
&= E[JE\{Q_j(x+dx)^2\} + J(J-1)E\{Q_j(x+dx)\}E\{Q_{j'}(x+dx)\}] \\
&= \lambda E[\{\sum_{n=1}^{N_j} \mathbf{1}(X_{jn} \in x+dx)\}^2] + \lambda^2 (\mu f_{X_{j_n}}(x) dx)^2 \\
&= \lambda E[N_j E\{\mathbf{1}(X_{jn} \in x+dx)\} + N_j(N_j-1)E\{\mathbf{1}(X_{jn}, X_{j_n'} \in x+dx)\}] \\
&\quad + (\lambda \mu f_{X_{j_n}}(x) dx)^2 \quad \text{with } n \neq n' \\
&= \lambda \mu f_{X_{j_n}}(x) dx + \lambda(\sigma^2 + \mu^2 - \mu) f_{X_{j_n} X_{j_n'}}(x, x) (dx)^2 + (\lambda \mu f_{X_{j_n}}(x) dx)^2,
\end{aligned}$$

where, for  $n \neq n'$ ,

$$\begin{aligned} f_{X_{j_n} X_{j_{n'}}}(x, x)(dx)^2 &= E[E\{\mathbf{1}(X_{j_n}, X_{j_{n'}} \in x + dx) \mid X_j\}] \\ &= E[E\{\mathbf{1}(X_{j_n} \in x + dx)\}E\{\mathbf{1}(X_{j_{n'}} \in x + dx)\}] \\ &= E[\{\phi_{\nu, X_j}(x)dx\}^2] \\ &= \int_{\mathbb{R}^2} \phi_{\nu, y}(x)^2 f_{X_j}(y) dy (dx)^2, \end{aligned}$$

because, given  $X_j$ ,  $\mathbf{1}(X_{j_n} \in x + dx)$  and  $\mathbf{1}(X_{j_{n'}} \in x + dx)$  are i.i.d. with p.d.f.  $\phi_{\nu, X_j}(\cdot)$ . Consequently,

$$\begin{aligned} V\{Q(x + dx)\} &= E\{Q(x + dx)^2\} - E\{Q(x + dx)\}^2 \\ &= \lambda \mu f_{X_{j_n}}(x) dx + \lambda(\sigma^2 + \mu^2 - \mu) E[\{\phi_{\nu, X_j}(x) dx\}^2]. \end{aligned}$$

*Criterion 6 for the GDM.* For the covariance  $\text{cov}\{Q(x_1 + dx), Q(x_2 + dx)\} = E\{Q(x_1 + dx)Q(x_2 + dx)\} - E\{Q(x_1 + dx)\}E\{Q(x_2 + dx)\}$  where  $x_1 + dx \cap x_2 + dx = \emptyset$

$$\begin{aligned} &E\{Q(x_1 + dx)Q(x_2 + dx)\} \\ &= E\left\{\sum_{j=1}^J Q_j(x_1 + dx)Q_j(x_2 + dx)\right\} + E\left\{\sum_{j=1}^J \sum_{j'=1, j' \neq j}^J Q_j(x_1 + dx)Q_{j'}(x_2 + dx)\right\} \\ &= \lambda E\{Q_j(x_1 + dx)Q_j(x_2 + dx)\} + \lambda^2 E\{Q_j(x_1 + dx)\}E\{Q_{j'}(x_2 + dx)\} \\ &= \lambda E\left[\left\{\sum_{n=1}^{N_j} \mathbf{1}(X_{j_n} \in x_1 + dx)\right\}\left\{\sum_{n'=1}^{N_j} \mathbf{1}(X_{j_{n'}} \in x_2 + dx)\right\}\right] + \lambda^2 \mu^2 f_{X_{j_n}}(x_1) f_{X_{j_n}}(x_2) (dx)^2 \\ &= \lambda E\left\{\sum_{n=1}^{N_j} \sum_{n'=1, n' \neq n}^{N_j} \mathbf{1}(X_{j_n} \in x_1 + dx, X_{j_{n'}} \in x_2 + dx)\right\} + E\{Q(x_1 + dx)\}E\{Q(x_2 + dx)\} \\ &= \lambda E[N_j(N_j - 1)E\{\mathbf{1}(X_{j_n} \in x_1 + dx, X_{j_{n'}} \in x_2 + dx)\}] \\ &\quad + E\{Q(x_1 + dx)\}E\{Q(x_2 + dx)\} \quad \text{with } n \neq n' \\ &= \lambda(\sigma^2 + \mu^2 - \mu) f_{X_{j_n} X_{j_{n'}}}(x_1, x_2) (dx)^2 + E\{Q(x_1 + dx)\}E\{Q(x_2 + dx)\}, \end{aligned}$$

where

$$\begin{aligned} f_{X_{j_n} X_{j_n'}}(x_1, x_2) &= E\{\phi_{\nu, X_j}(x_1)\phi_{\nu, X_j}(x_2)\} \\ &= \int_{\mathbb{R}^2} \phi_{\nu, y}(x_1)\phi_{\nu, y}(x_2)f_{X_j}(y)dy. \end{aligned}$$

Therefore, for  $x_1 + dx \cap x_2 + dx = \emptyset$ ,

$$\begin{aligned} \text{cov}\{Q(x_1 + dx), Q(x_2 + dx)\} &= E\{Q(x_1 + dx)Q(x_2 + dx)\} - E\{Q(x_1 + dx)\}E\{Q(x_2 + dx)\} \\ &= \lambda(\sigma^2 + \mu^2 - \mu)E\{\phi_{\nu, X_j}(x_1)\phi_{\nu, X_j}(x_2)\}(dx)^2. \end{aligned}$$

## C. Distance to the furthest particle: Proofs

### C.1. Distribution of $R^{\max}$

The following lemmas and corollaries are used to provide the distribution of the distance between a point source and the furthest deposited particle dispersed under the GDM and the IDMs.

*Lemma 1. Consider  $J$  random variables  $R_1, \dots, R_J$  in  $\mathbb{R}_+$  independently drawn under the following zero-inflated distribution:*

$$\begin{aligned} P(R_j = 0) &= q \\ f_{R_j}(r) &= f(r), \quad \forall r > 0. \end{aligned}$$

*Let  $F$  be the cumulated distribution function of  $R_j$ : for all  $r \geq 0$ ,  $F(r) = q + \int_0^r f(u)du$ . Assume that  $J$  has distribution  $p$  over  $\mathbb{N}$ . Let  $\mathcal{J} = \{1, \dots, J\}$  if  $J \geq 0$  and  $\mathcal{J}$  be the empty set otherwise. Then, the cumulated distribution function of the maximum  $R^{\max} = \max\{R_j : j \in \mathcal{J}\}$ ,  $J$  being unknown, satisfies:*

$$P(R^{\max} \leq r) = \sum_{j=0}^J F(r)^j p(j), \quad \forall r \geq 0.$$

*Proof of Lemma 1.* The above equation is obtained by using the distribution of the maximum of  $J$  i.i.d. statistics (David and Nagaraja, 2003, chap. 2), namely  $R_1, \dots, R_J$ , and by integrating this distribution with respect to  $J$ : for all  $r \geq 0$

$$\begin{aligned} P(R^{max} \leq r) &= \sum_{j=0}^{\infty} P(R^{max} \leq r \mid J = j)p(j) \\ &= p(0) + \sum_{j=1}^{\infty} P(R_k \leq r : k = 1, \dots, J \mid J = j)p(j) \\ &= \sum_{j=0}^{\infty} F(r)^j p(j). \end{aligned}$$

*Corollary 1:* Case where  $J$  is Poisson distributed. Assume that the hypotheses of Lemma 1 are satisfied. In addition, assume that  $J$  is Poisson distributed with mean  $\lambda$ , i.e.  $p(j) = e^{-\lambda}\lambda^j/j!$  for all  $j \in \mathbb{N}$ . Then, the cumulated distribution function of  $R^{max}$  is:

$$P(R^{max} \leq r) = \exp[\lambda\{F(r) - 1\}], \quad \forall r \geq 0,$$

and its distribution is:

$$\begin{aligned} P(R^{max} = 0) &= \exp[\lambda\{F(0) - 1\}] \\ f_{R^{max}}(r) &= \lambda f(r) \exp[\lambda\{F(r) - 1\}], \quad \forall r > 0. \end{aligned}$$

Apply Lemma 1 to prove Corollary 1.

### C.2. Distribution of $R_j^{max}$ under the IDMs

Under the IDMs,  $N_j = 1$  for all  $j \in \mathcal{J}$ . Consequently, in Eq. (2),  $p_{\mu, \sigma^2}(0) = 0$  and

$$\begin{aligned} f_{R_j^{max}}(r) &= f_{R_{j_n}}(r) \\ &= \begin{cases} \int_0^{2\pi} r f_{X_{j_n}}((r \cos \theta, r \sin \theta)) d\theta & \text{for the IDM1} \\ \int_0^{2\pi} r f_{X_j}((r \cos \theta, r \sin \theta)) d\theta & \text{for the IDM2.} \end{cases} \end{aligned} \quad (6)$$

If in addition  $f_{X_j}$  is assumed to be isotropic, then  $f_{R_j^{max}}(r) = 2\pi r f_{X_{j_n}}((r, 0))$  for the IDM1 and  $f_{R_j^{max}}(r) = 2\pi r f_{X_j}((r, 0))$  for the IDM2 (the simplification holds also for the IDM1 because the isotropy of  $f_{X_j}$  and the isotropy of the Brownian motion imply that  $f_{X_{j_n}}$  is isotropic).

### C.3. Distribution of $R_j^{max}$ under the GDM

Under the GDM, the distribution of  $R_j^{max}$  is zero-inflated and satisfies:

$$\begin{aligned} P(R_j^{max} = 0) &= p_{\mu, \sigma^2}(0) \\ f_{R_j^{max}}(r) &= \int_{\mathbb{R}^2} f_{R_j^{max}|X_j}(r | x) f_{X_j}(x) dx \\ &= \sum_{q=1}^{+\infty} q p_{\mu, \sigma^2}(q) \int_{\mathbb{R}^2} f_{R_{j_n}|X_j}(r | x) F_{R_{j_n}|X_j}(r | x)^{q-1} f_{X_j}(x) dx, \quad \forall r > 0. \end{aligned} \quad (7)$$

where  $f_{R_{j_n}|X_j}$  is the conditional distribution of  $R_{j_n}$  given  $X_j$  satisfying

$$\begin{aligned} f_{R_{j_n}|X_j}(r | x) &= 2r \int_0^{r^2} h_1(u, x) h_2(r^2 - u, x) du, \\ h_i(u, x) &= \frac{f_i(\sqrt{u}, x) + f_i(-\sqrt{u}, x)}{2\sqrt{u}}, \quad \forall i \in \{1, 2\}, \\ f_i(v, x) &= \frac{1}{\sqrt{2\pi\nu\|x\|}} \exp\left(-\frac{(v - x^{(i)})^2}{2\nu\|x\|}\right), \quad \forall i \in \{1, 2\}, \end{aligned} \quad (8)$$



$x = (x^{(1)}, x^{(2)})$  and  $F_{R_{j_n}|X_j}(r | x) = \int_0^r f_{R_{j_n}|X_j}(s | x) ds$ . This result is obtained by applying Corollary 2 given below whose assumptions are satisfied by the sub-model which governs the dispersal of the particles of a given group  $j$ .

It has to be noted that Equation (7) reduces to  $f_{R_j^{max}}(r) = f_{R_{j_n}}(r)$  when  $p_{\mu, \sigma^2}(j) = 1$  if  $j = 1$  and zero otherwise, i.e. when the GDM is an IDM. Thus, we obtain the result given in Equation (6) for the IDMs.

Besides, Corollary 3 given below provides a closed form for Equation (7) when the distribution of  $N_j$  is negative binomial as proposed in Equation (3) for the simulation study.

*Lemma 2.* Consider  $N$  random vectors  $X_1, \dots, X_N$  in  $\mathbb{R}^2$  independently drawn under the p.d.f.  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ . Assume that  $N$  is drawn from a counting distribution  $p : \mathbb{N} \rightarrow [0, 1]$  and that  $f$  is multiplicatively separable, i.e.  $f(x) = f_1(x^{(1)})f_2(x^{(2)})$  for all  $x = (x^{(1)}, x^{(2)})$  in  $\mathbb{R}^2$ , where  $f_1$  and  $f_2$  are real functions over  $\mathbb{R}$ . Let  $\mathcal{N} = \{1, \dots, N\}$  if  $N \geq 0$  and  $\mathcal{N}$  be the empty set otherwise. Then, the distribution of the maximum Euclidean distance  $R^{max} = \max\{\|X_n\| : n \in \mathcal{N}\}$  between the origin and the  $N$  random vectors,  $N$  being unknown, satisfies:

$$\begin{aligned} P(R^{max} = 0) &= p(0) \\ f_{R^{max}}(r) &= \sum_{q=1}^{+\infty} q f_{\|X_n\|}(r) F_{\|X_n\|}(r)^{q-1} p(q), \quad \forall r > 0, \end{aligned} \quad (9)$$

where  $F_{\|X_n\|}(r) = \int_0^r f_{\|X_n\|}(u)du$  and  $f_{\|X_n\|}$  is the p.d.f. of  $\|X_n\|$  satisfying

$$\begin{aligned} f_{\|X_n\|}(r) &= 2r \int_0^{r^2} h_1(u)h_2(r^2 - u)du, \\ h_i(u) &= \frac{f_i(\sqrt{u}) + f_i(-\sqrt{u})}{2\sqrt{u}}, \quad \forall i \in \{1, 2\}. \end{aligned}$$

*Proof of Lemma 2.* If  $N = 0$ , then  $R^{max} = 0$ . If  $N = q > 0$ , then the p.d.f. of  $R^{max}$  is  $r \mapsto qf_{\|X_n\|}(r)F_{\|X_n\|}(r)^{q-1}$ ; see David and Nagaraja (2003, chap. 2). This yields Eq. (9). The expression of  $f_{\|X_n\|}$  is obtained as follows. Let  $r \geq 0$ , we have:

$$\begin{aligned} f_{\|X_n\|}(r) &= 2r f_{\|X_n\|^2}(r^2), \\ f_{\|X_n\|^2}(r) &= f_{X_{n,1}^2 + X_{n,2}^2}(r) \\ &= \int_0^r f_{X_{n,1}^2}(u)f_{X_{n,2}^2}(r - u)du, \end{aligned}$$

where  $X_n = (X_{n,1}, X_{n,2})$ . In addition, for  $i \in \{1, 2\}$  and  $u > 0$

$$\begin{aligned} f_{X_{n,i}^2}(u) &= f_{X_{n,i}^2|X_{n,i} \geq 0}(u)P(X_{n,i} \geq 0) + f_{X_{n,i}^2|X_{n,i} < 0}(u)P(X_{n,i} < 0) \\ &= \frac{1}{2\sqrt{u}} [f_{X_{n,i}|X_{n,i} \geq 0}(\sqrt{u})P(X_{n,i} \geq 0) + f_{X_{n,i}|X_{n,i} < 0}(-\sqrt{u})P(X_{n,i} < 0)]. \end{aligned}$$

The density of the  $i$ -th component of  $X_n$  is  $f_i$  (see the separability hypothesis in the lemma), therefore, for  $v \in \mathbb{R}$ ,

$$\begin{aligned} f_{X_{n,i}|X_{n,i} \geq 0}(v) &= f_i(v)\delta_{v \geq 0}/P(X_{n,i} \geq 0) \\ f_{X_{n,i}|X_{n,i} < 0}(v) &= f_i(v)\delta_{v < 0}/P(X_{n,i} < 0), \end{aligned}$$

and

$$f_{X_{n,i}^2}(u) = \frac{1}{2\sqrt{u}} [f_i(\sqrt{u}) + f_i(-\sqrt{u})\delta_{u \neq 0}],$$

where  $\delta_E = 1$  if event  $E$  occurs, zero otherwise. The previous equations and the fact that  $f_{X_{n,i}^2}$  and  $h_i : u \mapsto (1/2\sqrt{u})\{f_i(\sqrt{u}) + f_i(-\sqrt{u})\}$  are different at only one point lead to the expression of  $f_{\|X_n\|}$  provided in Lemma 1.

*Corollary 2: Application of Lemma 2 to obtain  $f_{R_j^{max}}$ . Here, we use the notation introduced for the spatial GDM. For group  $j$  and given  $X_j = x = (x^{(1)}, x^{(2)})$ , the family  $\{X_{j_n} : n \in \mathcal{N}_j\}$ , where  $\mathcal{N} = \{1, \dots, N_j\}$  if  $N_j > 0$  and the empty set otherwise, satisfy the assumption of Lemma 1, and the p.d.f. of the  $i$ -th component of  $X_{j_n}$ ,  $i \in \{1, 2\}$ , is*

$$v \mapsto f_i(v, x) = \frac{1}{\sqrt{2\pi\nu\|x\|}} \exp\left(-\frac{(v - x^{(i)})^2}{2\nu\|x\|}\right),$$

because  $X_{j_n} \sim \mathcal{N}(x, \nu\|x\|I)$ . This formula and Lemma 1 allows us to get the expression of  $f_{R_{j_n}|X_j}$  provided in Eq. (8) (we recall that  $R_{j_n} = \|X_{j_n}\|$ ).

Then, Eq. (9) can be written

$$P(R_j^{max} = 0 \mid X_j = x) = p_{\mu, \sigma^2}(0)$$

$$f_{R_j^{max}|X_j}(r \mid x) = \sum_{q=1}^{+\infty} q f_{R_{j_n}|X_j}(r \mid x) F_{R_{j_n}|X_j}(r \mid x)^{q-1} p_{\mu, \sigma^2}(q), \quad \forall r > 0.$$

It follows that

$$P(R_j^{max} = 0) = p_{\mu, \sigma^2}(0)$$

$$f_{R_j^{max}}(r) = \int_{\mathbb{R}^2} f_{R_j^{max}|X_j}(r \mid x) f_{X_j}(x) dx$$

$$= \sum_{q=1}^{+\infty} q p_{\mu, \sigma^2}(q) \int_{\mathbb{R}^2} f_{R_{j_n}|X_j}(r \mid x) F_{R_{j_n}|X_j}(r \mid x)^{q-1} f_{X_j}(x) dx, \quad \forall r > 0.$$

Corollary 3: Case where  $N_j$  follows a negative-binomial distribution and  $f_{X_j}$  is isotropic. Let us now give the expression of  $f_{R_j^{max}}$  under the GDM when the distribution of  $N_j$  is negative-binomial (see Eq. (3)) and  $f_{X_j}$  is isotropic,

$$P(R_j^{max} = 0) = a^s$$

$$f_{R_j^{max}}(r) = \int_0^{+\infty} \frac{2\pi a^s (1-a) s g(r, z) z}{\{1 - (1-a)G(r, z)\}^{s+1}} f_{X_j}(z, 0) dz, \quad \forall r > 0,$$

where  $a = \mu/\sigma^2$  and  $s = \mu^2/(\sigma^2 - \mu)$  ( $\mu$  and  $\sigma^2$  being the parameters of the distribution of  $N_j$ , see Eq. (3)),  $G(r, z) = \int_0^r g(u, z) du$  and

$$g(r, z) = 2r \int_0^{r^2} h_1(u, z) h_2(r^2 - u, z) du,$$

$$h_i(u, z) = \frac{f_i(\sqrt{u}, z) + f_i(-\sqrt{u}, z)}{2\sqrt{u}}, \quad \forall i \in \{1, 2\},$$

$$f_i(r, z) = \frac{1}{\sqrt{2\pi\nu z}} \exp\left(-\frac{(u - z\delta_1(i))^2}{2\nu z}\right), \quad \forall i \in \{1, 2\}.$$

#### C.4. Proofs of Properties 1 and 2

Properties 1 and 2 gives results about the probability that  $R^{max}$  is larger than a distance  $r > 0$  which satisfies:

$$P(R^{max} \geq r) = \int_r^{+\infty} f_{R^{max}}(s) ds$$

$$= 1 - \exp[\lambda\{F_{R_j^{max}}(r) - 1\}],$$

$$= \begin{cases} 1 - \exp[\lambda\{F_{R_{jn}}(r) - 1\}] & \text{for the IDMs} \\ 1 - \exp[\lambda\{E(F_{R_{jn}}(r)^{N_j}) - 1\}] & \text{for the GDM} \end{cases}$$

where  $E(F_{R_{j_n}}(r)^{N_j}) = \sum_{q=0}^{+\infty} F_{R_{j_n}}(r)^q p_{\mu, \sigma^2}(q)$  and

$$F_{R_{j_n}}(r) = \int_0^r \int_0^{2\pi} u f_{X_{j_n}}((u \cos \theta, u \sin \theta)) du \quad \text{for the IDM1 and the GDM,}$$

$$F_{R_{j_n}}(r) = \int_0^r \int_0^{2\pi} u f_{X_j}((u \cos \theta, u \sin \theta)) du \quad \text{for the IDM2.}$$

*Proof of Property 1.* Let  $P_{IDM1}(r) = P(R^{max} \geq r)$  under the IDM1 and  $P_{GDM}(r) = P(R^{max} \geq r)$  under the GDM. With the notation of Property 1,

$$P_{GDM}(r) - P_{IDM1}(r) = \exp[\tilde{\lambda} \tilde{\mu} \{F_{R_{j_n}}(r) - 1\}] - \exp[\tilde{\lambda} \{E(F_{R_{j_n}}(r)^{N_j}) - 1\}],$$

where the distribution of  $N_j$  is  $p_{\tilde{\mu}, \sigma^2}(\cdot)$ .  $P_{GDM}(r) - P_{IDM1}(r) \leq 0$  if and only if  $\Delta(F_{R_{j_n}}(r)) \leq 0$  where

$$\Delta(\gamma) = \tilde{\mu}(\gamma - 1) - \{E(\gamma^{N_j}) - 1\}.$$

Let us study the sign of  $\Delta$  for  $\gamma \in ]0, 1]$  since  $F_{R_{j_n}}(r) \in ]0, 1]$  for all  $r > 0$ . As  $\tilde{\mu} = E(N_j)$ ,  $\Delta$  is equal to

$$\Delta(\gamma) = \sum_{n=0}^{+\infty} p_{\tilde{\mu}, \sigma^2}(n) \{n(\gamma - 1) - \gamma^n + 1\}.$$

For all  $\gamma \in ]0, 1]$  and  $n \in \mathbb{N}$ ,  $n(\gamma - 1) - \gamma^n + 1 \leq 0$  and, therefore,  $\Delta(\gamma) \leq 0$ . This leads to the conclusion of Property 1.

*Proof of Property 2.* Consider one particle whose location  $X_{j_1}$  is distributed under the IDM1:  $X_{j_1} = X_j + B_{j_1}(\nu ||X_j||)$ . The location  $X_j$  is distributed under the IDM2. Assume without lost of generality that  $X_j$  is on the  $x$ -axis and has a positive abscissa. Because the Brownian motion is isotropic, the probability that  $X_{j_1}$  is in the half-plane at the right of  $y = X_j^{(1)}$ , where  $X_j^{(1)}$ ,

is the first coordinate of  $X_j$  is  $1/2$ . The probability that  $X_{j1}$  is in the half-plane at the left of  $y = X_j^{(1)}$  and in the disk with radius  $X_j^{(1)}$  is less than  $1/2$ . Consequently,

$$P(\|X_{j1}\| \geq \|X_j\|) \geq 1/2$$

and  $F_{R_{j1}}(r) = P(\|X_{j1}\| \leq r)$  under the IDM1 is lower than  $F_{R_j}(r) = P(\|X_j\| \leq r)$  under the IDM2. Thus, Property 2 holds.

Accepted manuscript