Journal of Mathematical Biology manuscript No.
(will be inserted by the editor)

Daniela Morale · Vincenzo Capasso · Karl Oelschläger

An interacting particle system modelling aggregation behavior: from individuals to populations.

Dedicated to Akira Okubo

Received: July 1998 / Revised version: March 2004 –

Abstract. In this paper we investigate the stochastic modelling of a spatially structured biological population subject to social interaction. The biological motivation comes from the analysis of field experiments on a species of ants which exhibits a clear tendency to aggregate, still avoiding overcrowding. The model we propose here provides an explanation of this experimental behavior in terms of "long-ranged" aggregation and "short-ranged" repulsion mechanisms among individuals, in addition to an individual random dispersal described by a Brownian motion. Further, based on a "law of large numbers", we discuss the convergence, for large N, of a system of stochastic differential equations describing the evolution of N individuals (Lagrangian approach) to a deterministic integro-differential equation describing the evolution of the mean-field spatial density of the population (Eulerian approach).

1. Introduction.

In biology and medicine there is a wide spectrum of examples which exhibit collective behavior, such as formation of patterns and clustering. This may happen at any scale: from the cellular scale of embryonic tissue formation, wound healing or tumor growth, and vasculogenesis, the microscopic scale of life cycles of bacteria or social amoebae, to the larger scale of animal grouping; indeed animals may form *swarms*, characterized by a cohesive but unorganized aggregation (e.g. midges), or *schools* with a cohesive and synchronized organization (e.g. in fish schooling, individuals are oriented so that distances are uniform), or *shoals* (e.g. for fish) and *flocks* (e.g. for birds) in which animals are gathered together for social aims, in a synchronized or asynchronized way, or *herds, congregation*, and so on.

Our interest is to understand the mechanism of formation of aggregates: what is the biological motivation for aggregation, and how is adhesion maintained? Does the environment influence aggregation?

The strong biological attention to aggregation phenomena has stimulated a lively mathematical interest. Indeed, the complexity of the biological clustering raises very interesting mathematical problems. Aggregation patterns are usually explained in terms of forces, external and/or internal, acting upon individuals. A remarkable aspect of these global organization is that individuals move altogether in a coordinated (though random) fashion even though interaction among them via relevant senses (sight, smell, hearing, etc) are typically limited to much shorter distances than the size of the group.

Over the past couple of decades, a large amount of literature has been devoted to the mathematical modelling of self-organizing populations, based on the concepts of short-range/long-range "social interaction" among different individuals of a biological population. The aim of the modelling is to catch the main features of the interaction at the lower scale of single individuals that are responsible, at a larger scale, for a more complex behavior that leads to the formation of aggregating patterns.

A classical widespread approach has been based on PDE's [21, 25, 26]. This is due, above all, to the wider spread knowledge on nonlinear PDE's. So grouping behavior has been described by relevant quantities such as scalar or vector fields. Such kind of models are often called *Eulerian models*; they describe the evolution of population densities, they are based on continuum equations, typically (deterministic) nonlinear partial differential equations of the advection-reaction-diffusion type

$$\rho_t + \nabla \cdot (\mathbf{v}\rho) = \nabla \cdot (D\nabla\rho) + \nu(\rho), \tag{1}$$

where ρ is the population density, $\mathbf{v}$ is the velocity field and $\nu(\rho)$ is a possible additive reaction term which may include birth and death processes. The advection term may describe the interaction mechanisms among

D. Morale, V. Capasso: MIRIAM & Department of Mathematics, University of Milan, Italy,
K.Oelschläger: SFB 359, Istitut für Angewandte Mathematik -Heidelberg, Germany
e-mail: morale@mat.unimi.it

Key words: Agent based models, aggregation, self-organization, stochastic differential equations, moderate limit, empirical measures

individuals (via the velocity $\mathbf{v}$), while the non-convective (diffusive) flux takes into account the spatial spread of the population.

The advantages of the continuum approach are those of ease of analysis and elimination of arbitrary spatial discretization; it is useful in the case of large and dense populations [6]. The continuum model may be constructed a priori, by coupling (1) with a dynamical description of the velocity field [21]. The disadvantages of this approach include especially the fact that the identity of individuals is compromised. In many situations it is more appropriate to use discrete individual-based models, in which a finite number of individuals is considered and only a finite sequence of decisions is made by individuals. As pointed out by Durrett and Levin [6] and by Grünbaum and Okubo [14], an individual-based approach is useful in deriving the correct limiting equation, also in the case when the use of a continuum model can be justified. Indeed, by an Eulerian approach some important features of the dynamics may be hidden. There is a lot of work done in the literature, in which the flux is given in terms of an averaged value of the individual movement [13–15,21]. For example, in [13] Okubo and Grünbaum have studied an Eulerian model for the density flux described by a nonlinear integral term, obtained via the assumption that the number of individuals observed in any spatial interval is a Poisson-distributed random variable.

As already mentioned, a fruitful approach suggested since long by various authors [6,30,29] is based on the modelling of the "movement" of each individual "particle" embedded in the total population of N similar particles (the so called individual based model - IBM). Each individual is treated as a discrete particle subject to simple rules of movement. This is called the *Lagrangian approach*: individuals are followed in their motion. Possible randomness may be included in the motion of a particle, so that the variation in time of the (random) location of the k-th individual in the group at time $t \geq 0$, $X_N^k(t) \in \mathbb{R}^d, k = 1, \ldots, N$ is described by a system of stochastic differential equations (SDEs). On the other hand particles are subject to specific forces of interaction which are included in the advection term. Interaction may also lead to density dependence in the diffusion terms. In other words, from a Lagrangian point of view, the state of a system of N particles may be described as a (stochastic) process $\{X_N^k(t)\}_{t\in\mathbb{R}_+}$ defined on a suitable probability space $(\Omega, \mathcal{F}, P)$ and valued in $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, where $\mathcal{B}_{\mathbb{R}^d}$ is the Borel σ-algebra generated by intervals in $\mathbb{R}^d$. The time evolution of each $X_N^k(t)$ will be subject to a stochastic differential equation (SDE) of the type (2). If the number of particles is constant over time, the system of SDEs could be of the type

$$\begin{aligned} dX_N^k(t) = &\left[f_N^k(t) + h_N^k(X_N^1(t), \ldots, X_N^N(t), t)\right] dt \\ &+ \sigma\left(X_N^1(t), \ldots, X_N^N(t), t\right) dW^k(t), \quad k = 1, \ldots, N, \end{aligned} \tag{2}$$

where we have considered an additive noise modelled via a family of independent Wiener processes. Let us remark that the choice of an additive noise, described by a Wiener process has a precise mathematical meaning: there exists a suitable theory that makes the analysis rigorous. This will be clarified to the reader later on.

The function $h_N^k : \mathbb{R}^{Nd} \times \mathbb{R}_+$ describes the interaction of the k-th particle with other particles in the system; the function $f_N^k : \mathbb{R}_+ \to \mathbb{R}$ models possible individual dynamics which may depend only on time or on the state of the particle itself.

Stochastic Lagrangian models offer the advantage of being directly related to experimental data on the behavior of individuals of a real population, especially when dealing with a relatively "small" number of individuals per unit space (Poisson like spatial processes) [16].

In another way, the state of the k-th particle may be modelled as a random Dirac-measure in $\mathcal{M}(\mathbb{R}^d)$

$$\epsilon_{X_N^k(t)}, \tag{3}$$

which is the degenerate measure defined as follows

$$\epsilon_{X_N^k(t)}(B) = \begin{cases} 1 & \text{if } X_N^k(t) \in B \\ 0 & \text{if } X_N^k(t) \notin B \end{cases} \qquad \forall B \in \mathcal{B}_{\mathbb{R}^d}. \tag{4}$$

Expression (4) means that its generalized density function is the Dirac-δ function $\delta_{X_N^k(t)}$, i.e. for any sufficiently smooth $f : \mathbb{R}^d \to \mathbb{R}$

$$\int_{\mathbb{R}^d} f(y)\epsilon_{X_N^k(t)}(dy) = \int_{\mathbb{R}^d} f(y)\delta_{X_N^k(t)}(y)dy = f\left(X_N^k(t)\right).$$

By an Eulerian approach, the collective behavior of the discrete (in the number of particles) system, may be given in terms of the spatial distribution of particles at time t, expressed in term of an empirical measure

$$X_N(t) = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(t)} \in \mathcal{M}(\mathbb{R}^d), \tag{5}$$

such that $\forall B \in \mathcal{B}(\mathbb{R}^d)$,

$$\left(X_N(t)\right)(B) = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(t)}(B) = \frac{\sharp\text{particles in } B \text{ at time } t}{N}.$$

This means that $X_N(t)$ measures the spatial relative frequency of the particles at time t. The stochastic measure valued process

$$X_N : t \in \mathbb{R}_+ \to X_N(t) \in \mathcal{M}(\mathbb{R}^d)$$

is known as the empirical process associated with our system.

In conclusion, the two different approaches (Lagrangian and Eulerian) describe the system at different scales: the finer scale description based on the (stochastic) behavior of individuals (microscale) (2) and the larger scale description based on the (continuum) behavior of population densities (macroscale) (1). "The central problem is to determine how information is transferred across scales, and what detail at fine scales is exactly necessary and sufficient for understanding patterns on averaged scales"[20].

Often a multiple scale approach is preferable: the global behavior of the population is described, at the macroscopic scale, by a continuum density whose evolution in terms of integro-differential equations is derived by a limiting process from the empirical distribution associated with a large number of particles.

The aim of our analysis is to provide a mathematically rigorous framework for bridging the gap between the two different scales: the microscale and the macroscale. In order to consider a rigorous limit, we introduce in the model the concept of "mesoscale" [27,28]. We look at the system at a scale which is much larger than the microscale and much smaller than the macroscale, so that when the number of individuals increases to infinity, at the "mesoscale" we have a sufficient number of particles so that a "law of large numbers" may still be applied. The mesoscale can be obtained by a suitable rescaling of the kernel modelling the interaction among particles. Indeed, the interaction among particles is mathematically modelled by an interaction kernel depending on the distance between two particles. The range of such kernel depends on the scale of the interaction.

A mathematical way to distinguish among different scales, in the above sense, is based on the choice of a "scaling" parameter in the kernel describing the interaction among individuals; if we consider a system of N particles located in $\mathbb{R}^d$, in the macroscopic space-time coordinates the typical distance between neighboring particles is is $O(N^{-1/d})$ and the order of the size of the whole space $O(1)$. In this respect, we may distinguish three main types of interactions:

a. *McKean-Vlasov interaction* (macroscale): any particle interacts with $O(N)$ other particles; collective long-range forces are predominant and the particles are weakly interacting; the range of interaction gets very large in comparison with the typical distance between neighboring particles, and its strength decreases fast, like $1/N$.
b. *hydrodynamic interaction* (microscale): any particle interacts with $O(1)$ other particles in a very small neighborhood with volume $O(1/N)$. The interaction gets short-ranged and rather strong for large N.
c. *moderate interaction* (mesoscale) [27]: any particle interacts with many $O\left(N/\alpha(N)\right)$ other particles in a small volume $O\left(1/\alpha(N)\right)$ where both $\alpha(N)$ and $(N/\alpha(N))$ tend to infinity as $N \to \infty$.

The three types of interaction may be obtained in terms of an appropriate rescaling of a given reference function V_1. Let V_1 be a sufficiently regular probability density; and assume that the interaction of two particles, out of N, located in x and y respectively is modelled by

$$\frac{1}{N} V_N(x-y), \tag{6}$$

where

$$V_N(z) = N^{\beta} V_1(N^{\beta/d} z), \tag{7}$$

which expresses the rescaling of V_1 with respect to the total number N of particles, in terms of a scaling coefficient $\beta \in [0,1]$. The force exerted on the k-th (out of N) single particle located at $X_N^k(t)$ due to the interaction of the single k-particle with all the others in the population is given by

$$\begin{aligned} I^k \equiv I^k(X_N^1(t), \dots, , X_N^N(t)) &= \sum_{i=1}^{N} \frac{1}{N} V_N\left(X_N^k(t) - X_N^i(t)\right) \\ &= \sum_{i=1}^{N} N^{\beta-1} V_1\left(N^{\beta/d}\left(X_N^k(t) - X_N^i(t)\right)\right) \end{aligned} \tag{8}$$

We easily recognize that the interaction term I^k can be expressed in terms of the empirical distribution (5) as follows

$$I^k = (X_N(t) * V_N)(X_N^k(t)) \tag{9}$$

We have a McKean-Vlasov (or weak) interaction if $\beta = 0$, a hydrodynamic interaction if $\beta = 1$, and a moderate interaction, if $\beta \in (0,1)$.

For a finite number N of individuals the empirical spatial distribution (5) may still suffer of significant stochastic fluctuations. But, being the empirical distribution a "relative frequency", a "law of large numbers" would suggest that, for N tending to infinity, stochastic fluctuations may disappear [29].

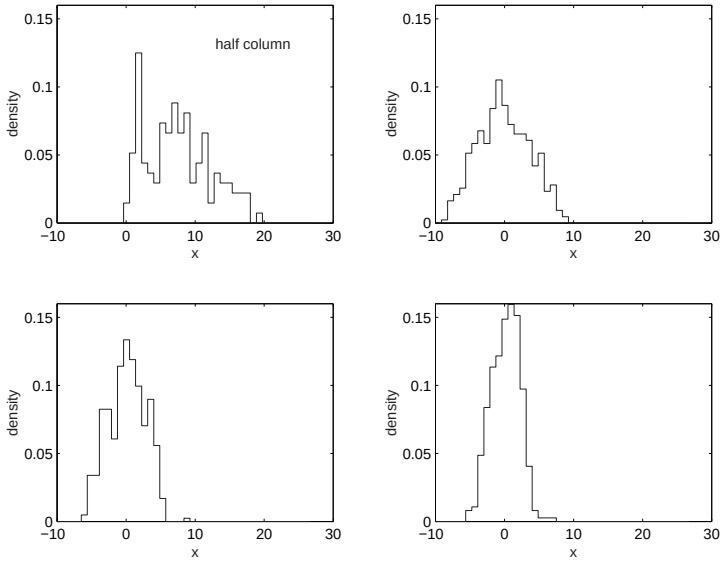

Fig. 1. "Field"-experiments [3]: the pictures show the density profile in a cross section of an army of Polyergus rufescens on different type of terrain: the number of obstacles is increasing, from a concrete ground at the upper left (where only one half of the army is represented) to a high grass field at the bottom right. The width of the distribution gets narrower, from about 40 cm to 10 cm.

Indeed a rigorous proof of the correctness of the above reasoning can be obtained based on the literature by Oelschläger [27,28] about "law of large numbers" for "moderately" interacting particles.

The biological motivation of the present paper arises from an example of animal swarming, investigated in [2, 3]. Field experiments have been performed in Piemonte, a northern region of Italy on a population of ants of the species Polyergus rufescens. The colonies of these species are characterized by the absence of polyethism in the worker cast which is composed only by soldiers, unable to attend any task (e.g. brood tending or nest maintenance) other than raiding activity [9]. These indispensable tasks are performed by individuals belonging to few specific species which have been kidnapped by *Polyergus* soldiers when they were newborn or pupae, and grew up in the *Polyergus* nest. To keep constant the slave's population in their nest, *Polyergus* ants periodically raid ant nests of the slave species. In these circumstances *Polyergus* soldiers aggregate in an army of 300-1000 individuals, 10-40 cm wide and some meters long. The army looks strikingly different than the foraging ant column we use to see in our houses. By visual inspection of the recorded videotapes, data about individual's distribution and density in the army have been gathered. Analyzing these data we found remarkable differences in the army structure in different environmental conditions such as regularity of the terrain, reciprocal visibility, etc. which may impose restrictions on the interaction range (sensitivity) among individuals [3]. In Figure 1 the density profile in a cross section of an army of Polyergus rufescens on different type of terrain: the number of obstacles is increasing, from a concrete ground at the upper left (where only one half of the army is represented) to a high grass field at the bottom right. The width of the distribution gets narrower, from about 40 cm to 10 cm.

Hence the ants during their raids seem to aggregate in a transversally organized army. Indeed, in the main direction of motion probably the dominant factor is the chemical trail produced by the scouts [3]. So we neglect this feature since in this phase we are interested only in the aggregation phenomenon.

We may consider our model as a modification of the Okubo-Grünbaum model [13,14]. They present a one-dimensional model for swarming. A key feature of their model is that individuals count their neighbors within a suitable range and estimate the local density gradient, moving towards higher density. Furthermore, in order to avoid overcrowding, at any time they compare the observed local density with a common target density μ (independent of space), which represents a desired number of ants in a neighborhood of fixed radius. So swarming is due to the aggregating behavior of individuals, and blow up is avoided by including in the dynamics the concept of a given common target density.

Actually, the density profiles of the field experiments presented in Figure 1 show that ants do not present overcrowding effects. In the model proposed here we capture this feature, without imposing a common target density as an external information. We achieve the same qualitative spatial pattern as the experimental one, by including short range repulsion in addition to the well accepted aggregation mechanism.

Now we are ready to list the distinctive features of the model studied here

a. individuals interact only directly, not via an underlying field (as in taxis); so in (2)

$$h_N^k(X_N^1(t),\dots,X_N^N(t),t) = F_N[X_N(t)](X_N^k(t)), \quad \forall k = 1,\dots,N, \tag{10}$$

where the functional F_N is defined on $\mathcal{M}(\mathbb{R}^d)$;

b. in order to avoid overcrowding, we do not consider a target density (which seems not to be realistic [2,3]); instead we introduce repulsion by means of a kernel as in (7);
c. aggregation and repulsion compete but act at different scales: repulsion at the mesoscale, aggregation at the macroscale; so both are modelled by components of the type (7) with $\beta \in (0,1)$ and $\beta = 0$, respectively;
d. the aggregation kernel may also model some environmental characteristics; in particular the dependence on the environment mentioned above is modelled by the introduction of a parameter, representing the "visibility range" of particles.
e. particles move randomly, with a variance which may depend on the population itself(e.g. via the size or via the distribution);
f. no individual drift is included, so in (2) we have $f_N^k(t) = 0, \forall t, \forall k$.

We wish to remark that both kind of interaction (attraction and repulsion) are described by non local operators; for aggregation, the kernel is not rescaled as a function of N, so that the limiting equation will keep non local integral terms. In contrast, repulsion, being described by a moderate kernel, will be represented by local operator, in the limit: the repulsive interaction gets more and more local so that we do not expect integral terms in the limiting PDE.

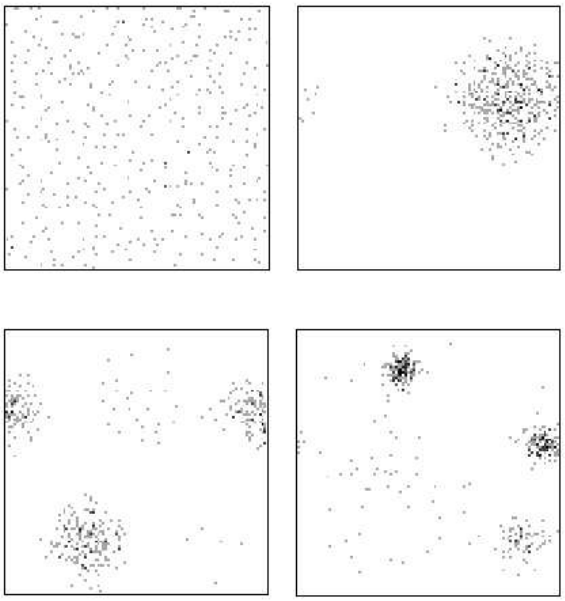

Fig. 2. Cellular automata simulations carried out on a 100 X 100 lattice, with uniform initial condition, N = 1000. The range of aggregation R_a is, 75%, 50%, 20% and 10% of the edge of the lattice.

By means of a corresponding cellular automaton model [23] we got qualitative results comparable to field experiments in [2]. In Fig. 2 simulations results are shown: at a fixed $N = 1000$, by reducing the range of the aggregation kernel, i.e. by assuming a reduction of the sensory range of single individuals, by means of our rules for the dynamics of aggregation, we get qualitative similarities with respect to the density profiles in Fig. 1: from the left top we decrease the support of the aggregation kernel; as a consequence the aggregation gets stronger, as we deduce from the formation of clusters of diameters smaller and smaller.

2. The Lagrangian description.

The Lagrangian description of the dynamics of our system of interacting particles is given via a system of stochastic differential equations as follows:

$$dX_N^k(t) = F_N[X_N(t)](X_N^k(t))dt + \sigma_N dW^k(t) \quad k = 1, \dots, N, \tag{11}$$

where $X_N^k(t) \in \mathbb{R}^d$ and $X_N(t) \in \mathcal{M}(\mathbb{R}^d)$ are defined by (3)-(5). We assume that the position of the k-th particle, $X_N^k(t)$, is subject to random dispersal described by the Brownian motion W^k. $\{W^k, k = 1, \dots, N\}$ is a family of independent standard Wiener processes. We take into account just the intrinsic stochasticity: each particle moves randomly with a mean free path σ_N, because of its need of socializing. When some other particle enter its interaction range, the predominant part of the dynamics is due to the drift term. This is the main reason why it might be allowed to consider σ_N reducing as N increases, and in particular to assume it vanishing. Indeed if the

number of particles is large, the mean free path of each particle may reduce up to a limiting value that may be zero

$$\lim_{N\to\infty} \sigma_N = \sigma_\infty \geq 0. \tag{12}$$

The drift term F describes the specific dynamics of the system of interacting particles, based on our modelling assumptions. We express our modelling assumptions by introducing in the drift term F_N two additive components [33]: F_1, responsible of aggregation, and F_2, responsible of repulsion, such that

$$F_N \;=\; F_1 \;+\; F_2.$$

Aggregation. Particles tend to aggregate subject to their interaction within a range of width $R_a > 0$ (finite or not). This corresponds to the assumption that each particle is capable of perceiving the others only within a suitable range; in other words each particle has a limited knowledge of the spatial distribution of the other particle. Aggregation is modelled by a McKean-Vlasov interaction kernel, by (9) with $\beta = 0$ and reference kernel

$$G_a : \mathbb{R}^d \longrightarrow \mathbb{R}_+$$

having a support confined to the ball centered at $0 \in \mathbb{R}^d$ and radius $R_a \in \bar{\mathbb{R}}_+$. R_a models the range of sensitivity for aggregation, independent of N.

A "generalized" gradient operator is obtained as follows. Given a measure $\mu \in \mathcal{M}(\mathbb{R}^d)$, we define the quantity

$$\nabla G_a * \mu$$

as the classical convolution of the gradient of the kernel G_a with the measure μ.

So the aggregation term F_1 is given by

$$F_1[X_N(t)]\left(X_N^k(t)\right) \;=\; \left[\nabla G_a * X_N(t)\right]\left(X_N^k(t)\right). \tag{13}$$

This means that each individual moves according to this generalized gradient of the measure $X_N(t)$ with respect to the kernel G_a; the positive sign for F_1 expresses a force of attraction of the particle in the direction of increasing concentration of individuals.

We emphasize the great generality included in this definition of generalized gradient of a measure μ on $\mathbb{R}^d$. By using particular shapes of G_a, one may include angular ranges of sensitivity, asymmetries, etc. at a finite distance [16].

Repulsion. Particles are subject to repulsion when they come "too close" to each other. Repulsion is modelled by the gradient of (7), with $\beta \in (0,1)$. The choice of β determines the range and the strength of the influence of neighboring particles; indeed, any particle interacts (repelling) with $O\left(N^{1-\beta}\right)$ other particles in a small volume $O\left(N^{-\beta}\right)$.

V_1 in (7) is a symmetric probability density on $\mathbb{R}^d$. It is clear that

$$\lim_{N\to+\infty} V_N = \delta_0, \tag{14}$$

where δ_0 is Dirac's delta function.

Finally we define

$$\begin{aligned} F_2[X_N(t)](X_N^k(t)) \;&=-\left(\nabla V_N * X_N(t)\right)\left(X_N^k(t)\right) \\ &= -\frac{1}{N}\sum_{m=1}^{N} \nabla V_N(X_N^k(t) - X_N^m(t)). \end{aligned} \tag{15}$$

The strength of the influence of neighboring particles is $N^{1-\beta}$; the negative sign for F_2 expresses a drift towards decreasing concentration of individuals. In this case the range of the repulsion kernel decreases to zero as the size N of the population increases to infinity.

Finally the system (11) reads

$$\begin{aligned} dX_N^k(t) = &\left(\left[\nabla G_a * X_N(t)\right]\left(X_N^k(t)\right) - \left[\nabla V_N * X_N(t)\right]\left(X_N^k(t)\right)\right) dt \\ &+ \sigma_N dW^k(t) \quad k = 1,\ldots,N, \end{aligned} \tag{16}$$

It is not difficult to prove that if G_a and V_N are sufficiently regular, i.e. for G_a and $V_N \in C_b^2(\mathbb{R}^d)$, this system of N (finite) stochastic differential equations is well posed, i.e. a unique solution of (16) exists [11].

In Fig. 3 and Fig. 4 we present some simulations of the empirical measure $X_N(t)$, whose atoms are solution of the system of stochastic differential equations (16).

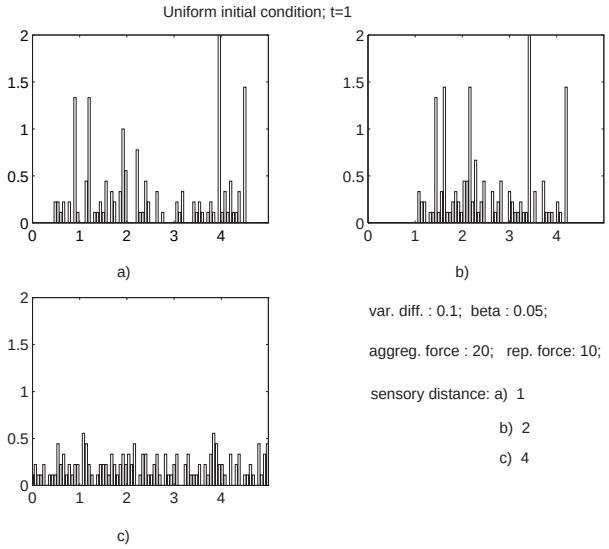

Fig. 3. Simulation results for the SDE model with a uniform initial distribution for different values of the range R_a.

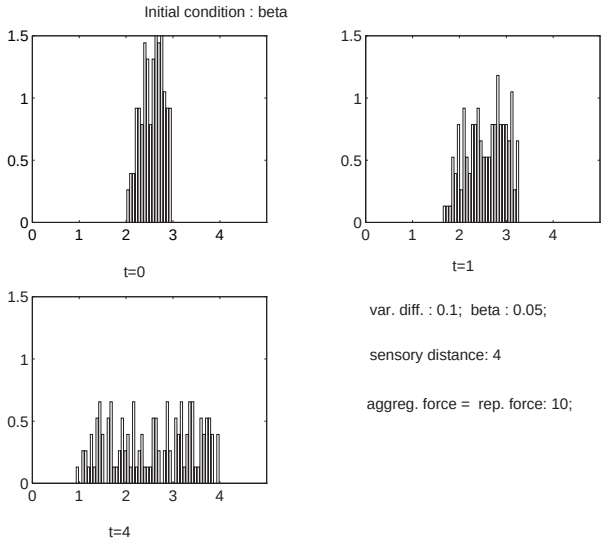

Fig. 4. Time evolution of the population density for the SDE model with a beta initial distribution.

3. The Eulerian description

In order to give an Eulerian description of the system of the N particles, we need to consider the time evolution of the empirical measure X_N.

A fundamental tool is Itô's formula [11] for the time evolution of any function $f(X_N^k(t), t)$, $f \in C_b^2(\mathbb{R}^d \times \mathbb{R}_+)$, of the trajectory $\{X_N^k(t), t \in \mathbb{R}_+\}$ of the individual particle subject to the SDE (11):

$$\begin{aligned} f(X_N^k(t),t) = f(X_N^k(0),0) \\ + \int_0^t F[X_N(s)](X_N^k(s)) \nabla f(X_N^k(s),s) ds \\ + \int_0^t \left[\frac{\partial}{\partial s} f(X_N^k(s),s) + \frac{\sigma_N^2}{2} \triangle f(X_N^k(s),s) \right] ds \\ + \sigma_N \int_0^t \nabla f(X_N^k(s),s) dW_s. \end{aligned} \tag{17}$$

Let us introduce the following notation

$$\langle \mu, f \rangle \;=\; \int f(x)\mu(dx),$$

for the integral of any (sufficiently smooth) function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ with respect to a measure μ on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$.

By taking into account expressions (13) and (15) for F_1 and F_2, we get, from (17), the following representation of the time evolution of $X_N(t)$:

$$\begin{aligned}
\langle X_N(t), f(\cdot,t)\rangle \;&=\; \frac{1}{N}\sum_{k=1}^{N} f(X_N^k(t),t) \\
&= \langle X_N(0), f(\cdot,0)\rangle \\
&\quad + \int_0^t \langle X_N(t), [(X_N(s) * \nabla G_a)(\cdot) - (X_N(s) * \nabla V_N)(\cdot)] \cdot \nabla f(\cdot,s)\rangle ds \qquad (18a) \\
&\quad + \int_0^t \left\langle X_N(t), \frac{\sigma_N^2}{2}\triangle f(\cdot,s) + \frac{\partial}{\partial s} f(\cdot,s)\right\rangle ds \qquad (18b) \\
&\quad + \frac{\sigma_N}{N}\int_0^t \sum_k \nabla f(X_N^k(s),s) dW^k(s), \qquad f \in C_b^{2,1}(\mathbb{R}^d \times [0,\infty)). \qquad (18c)
\end{aligned}$$

Equation (18) includes a nonlinear term (18a) which derives from the specific dynamical model (13) and (15); on the other hand a diffusion term appears in (18b) due to the additive Brownian motion in (11). The last term (18c)

$$M_N(f,t) \;=\; \frac{\sigma_N}{N}\int_0^t \sum_k \nabla f(X_N^k(s),s) dW^k(s);$$

is the only explicit source of stochasticity in the equation. It is a martingale with respect to the natural filtration of the process $\{X_N(t), t \in \mathbb{R}_+\}$ [11].

Simulation results of the empirical measure $X_N(t)$ are presented in Figure 5.

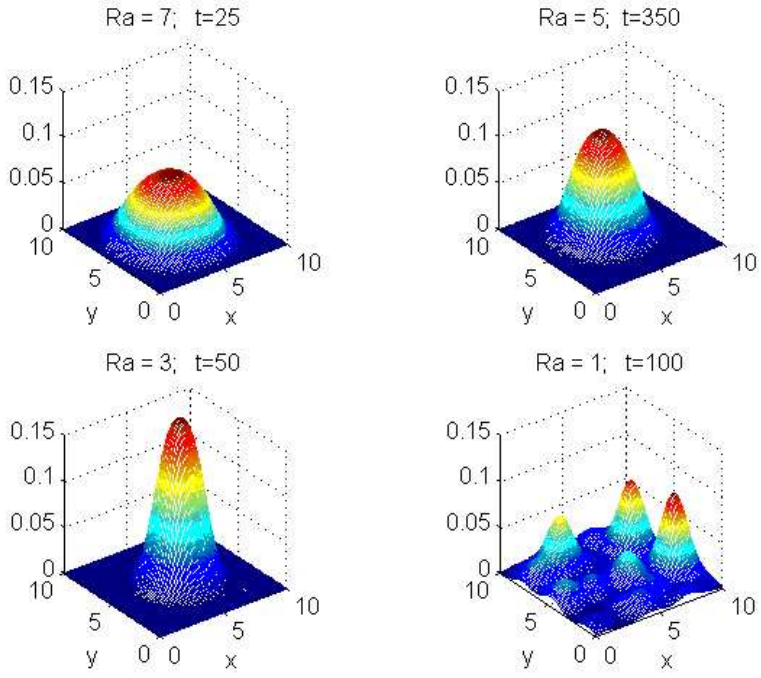

Fig. 5. Different aggregation profiles for the empirical distribution of particles.

4. Asymptotic behavior of the system for large populations

Equation (18) shows how, when the number of particles N is large but still finite, also from the Eulerian point of view the system keeps the stochasticity which characterizes each individual. This is not true anymore when the size of the system tends to infinity. The main reason is that the martingale term $M_N(f,t)$ vanishes in probability.

Indeed we may apply Doob's inequality [11] to obtain

$$\begin{aligned} E\left[\sup_{t\le T}|M_N(f,t)|\right]^2 &\le E\left[\sup_{t\le T}|M_N(f,t)|^2\right] \\ &\le 4E\left[|M_N(f,T)|^2\right] \\ &\le \frac{4\sigma_N^2}{N^2}\sum_{k=1}^{N} E\left[\int_0^T |\nabla f(X_N^k(s),s)|^2 ds\right] \\ &\le \frac{4\sigma_N^2\|\nabla f\|_\infty^2 T}{N}. \end{aligned} \tag{19}$$

Hence for the zero-mean martingale $M_N(f,t)$, the quadratic variation (19) vanishes in the limit $N\to\infty$. This implies convergence to zero in probability. This is the substantial reason of the deterministic limiting behavior of the process, as $N\to\infty$, since in this limit the evolution equation of the process will not contain the Brownian noise anymore.

Actually, various non trivial mathematical problems arise in connection with a rigorous proof of the existence of a unique deterministic limiting mean-valued process $\{X(t)\}_{t\in\mathbb{R}_+}$. This requires rather sophisticated mathematical techniques that we leave to a more technical paper [22] (see an anticipated discussion in Section 5). Here we restrict ourselves to a formal derivation of the mean-field model.

Suppose that indeed the empirical process $\{X_N(t),\ t\in\mathbb{R}_+\}$ tends, as $N\to\infty$, to a deterministic process $\{X(t),\ t\in\mathbb{R}_+\}$, and further that it admits for any $t\in\mathbb{R}_+$, a density $\rho(x,t)$ with respect to the Lebesgue measure on $\mathbb{R}^d$, so that

$$\begin{aligned} \lim_{N\to+\infty}\langle X_N(t), f(\cdot,t)\rangle &= \langle X(t), f(\cdot,t)\rangle \\ &= \int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx. \end{aligned}$$

As a formal consequence we get

$$\begin{aligned} \lim_{N\to+\infty}(X_N(t)*V_N)(x) &= \rho(x,t), \\ \lim_{N\to+\infty}(X_N(t)*\nabla V_N)(x)) &= \nabla\rho(x,t), \\ \lim_{N\to+\infty}(X_N(t)*\nabla G_a(x,t)) &= (X(t)*\nabla G_a(x,t)) \\ &= \int_{\mathbb{R}^d}\nabla G_a(x-y,t)\rho(y,t)dy. \end{aligned}$$

Hence, by applying the above limits, from (18) we get

$$\begin{aligned} \int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx &= \int_{\mathbb{R}^d} f(x,0)\rho(x,0)dx \\ &+ \int_0^t ds\int_{\mathbb{R}^d} dx[(\nabla G_a*\rho(\cdot,s))(x)-\nabla\rho(x,s)]\cdot\nabla f(x,s)\rho(x,s) \\ &+ \int_0^t ds\int_{\mathbb{R}^d} dx\left[\frac{\partial}{\partial s}f(x,s)\rho(x,s)+\frac{\sigma_\infty^2}{2}\triangle f(x,s)\rho(x,s)\right], \end{aligned} \tag{20}$$

where

$$\lim_{N\to\infty}\sigma_N = \sigma_\infty.$$

We recognize that (20) is the weak version of the following equation for the spatial density $\rho(x,t)$:

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\sigma_\infty^2}{2}\triangle\rho(x,t) + \nabla\cdot(\rho(x,t)\nabla\rho(x,t))$$

$$- \nabla\cdot[\rho(x,t)(\nabla G_a * \rho(\cdot,t))(x)], \quad x\in\mathbb{R}^d, t\geq 0, \tag{21}$$

$$\rho(x,0) \;=\; \rho_0(x), \quad x\in\mathbb{R}^d. \tag{22}$$

From (21) if $\sigma_\infty > 0$ the dynamics of the density is smoothed by the diffusive term. This is due to the memory of the fluctuations existing when the number of particle N is finite. This means also that the dynamics of a single particle is still stochastic (for a better discussion see [22]). When $\sigma_\infty = 0$ all stochasticity disappears. This brings to a degenerate equation

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla\cdot(\rho(x,t)\nabla\rho(x,t)) - \nabla\cdot[\rho(x,t)(\nabla G_a * \rho(\cdot,t))(x)], \quad x\in\mathbb{R}^d, t\geq 0. \tag{23}$$

5. Some mathematical remarks

A rigorous proof of the convergence of the stochastic evolution equation (18) for the mean-valued empirical process $\{X_N(t)\}_{t\in\mathbb{R}_+}$ to the evolution equation (21) or (23) for the spatial density of the deterministic mean-valued process $\{X(t)\}_{t\in\mathbb{R}_+}$ includes the following steps:

a) prove existence of a deterministic limiting measure values process $\{X_\infty(t),\, t\in\mathbb{R}_+\}$.
b) find the limit measure and prove its absolute continuity with respect to the usual Lebesgue measure on $\mathbb{R}^d$
c) provide existence and uniqueness for the solution of the deterministic density $\rho(x,t)$ of $X(t)$ satisfying the evolution equation (21) and (23).

The proof of point a) is based on compactness arguments of the sequence $\{X_N(t), t\in\mathbb{R}_+\}_{N\in\mathbb{N}}$. Relative compactness is a consequence of the relative compactness of the process $\{X_N(t)\}_{N\in\mathbb{N}}$ at a fixed time t and of small variations of $\{X_N(t), t\in\mathbb{R}_+\}_{N\in\mathbb{N}}$ in small time interval, uniformly in N (cf. [8], p. 137). The major mathematical difficulty in this framework comes from the non linear term

$$\Phi_{N,f}(t) = \int_0^t \langle X_N(s), [X_N(s) * \nabla V_N(\cdot)]\,\nabla f(\cdot,s)\rangle\, ds \tag{24}$$

in (18a). If we re-write (24) in an explicit form we get

$$\Phi_{N,f}(t) = \int_0^t \frac{1}{N^2}\sum_{k,m=1}^N \nabla V_N(X_N^k(s) - X_N^m(s))\nabla f(X_N^k(s),s)ds \tag{25}$$

$$= \int_0^t \frac{1}{N^2}\sum_{k,m=1}^N N^\beta\,\nabla V_1\left(N^{\beta/d}\left(X_N^k(s) - X_N^m(s)\right)\right)\nabla f(X_N^k(s),s)ds. \tag{26}$$

Since $\beta > 0$, the kernel $V_N \to \delta_0$, namely the Dirac delta function; this shows that, for N tending to infinity, even small changes of the relative positions of neighboring particles may have a considerable effect on $\Phi_{N,f}(t)$. This means that the systems of $X_N(t)$ needs to reach a sort of *equilibrium*. We consider the regularized distribution

$$h_N(x,t) = (X_N(t) * W_N)(x),$$

where

$$W_N(x) = N^\beta W_1(N^{\beta/d}x),$$

with W_1 a regular function such that

$$V_N(x) = (W_N * W_N)(x), \forall x\in\mathbb{R}^d.$$

The function $h_N(x,t)$ represents an averaging over a large number of particles, at the level of the mesoscale (given by the size of β); furthermore V_N (and so W_N) has to go towards a Dirac δ-function sufficiently slow ("moderate limit"), so that we may apply a "law of large numbers" in such a way that $h_N(x,t)\to\rho(x,t)$. Indeed,

$$h_N(x,t) = (X_N(t) * W_N)(x) = \frac{1}{N}\sum_{i=1}^N W_N(x - X_i(t)) = \frac{1}{N^{1-\beta}}\sum_{i=1}^N W_1(N^{\beta/d}(x - X_i(t))),$$

and we need to be sure that the sum is performed over a sufficiently large number of particles, that is equivalent to say that the range of W_N has not to shrink to zero too fast. Furthermore, in order to have the system in an equilibrium state, we need to control the variation of (6) with ∇V_N instead of V_N, at the origin; we have

$$\frac{1}{N}\left|\triangle V_N(0)\right| \leq N^{\frac{\beta(d+2)}{d}-1}. \tag{27}$$

This variation has not to go to infinity as the number of particles increases to infinity. As a consequence of (27) the following assumption is needed

$$\beta < \frac{d}{d+2}, \tag{28}$$

i.e. β needs to be small enough, so that the regularizing function W_N is able to smooth the irregularities of $X_N(t)$.

Actually we can also improve the (28) in the case $\sigma_N \to 0$. Indeed, in that case β is allowed to be greater than $\frac{d}{d+2}$, on condition that σ_N decreases to zero sufficiently fast, s.t.

$$\frac{d}{d+2} \leq \beta < 1, \qquad \lim_{N\to+\infty} \sigma_N N^{\beta(d+2)/d-1} = 0 \tag{29}$$

Here we avoid the technicalities needed to justify condition (29) (cf. [22]), but we want to highlight the possible meaning of (28) and (29): they are needed to control the variance of the stochastic fluctuations of the empirical process X_N, i.e. to ensure some relative compactness properties for the process; for $\beta < d/(d+2)$ the range of the kernel V_N is large enough to provide a smoothing of the possible fluctuations. For $\beta \geq d/(d+2)$ this is not sufficient, so that we have to control directly also the variance σ_N^2 of the fluctuations due to the Brownian movement.

As the point b) concerns, the form of the limit measure is obtained by some martingale arguments [8].

Finally, point c) i.e. the proof of the uniqueness may be based on the assumption that the limiting evolution equation admits a unique and sufficient regular solution.

Two cases are to be distinguished : the viscous case (21) and the non viscous case (23).

In the viscous case a large available literature shows existence and uniqueness of a solution satisfying sufficient regularity [4,7,19,32].

In the non viscous case uniqueness is not a trivial problem (see e.g. [5]). Recent literature [4] suggests that uniqueness may be recovered by selecting an entropy solution for (23). The problem of sufficient regularity is still an open problem in this case, which goes beyond the interest of the present paper. Partial answers may be found in classical literature [25,26,31].

As proven in [22], by assuming for some $T \in [0,\infty)$ equation (21) admits a unique, nonnegative solution $\rho \in C_b^{[(d+2)/2]+2,1}(\mathbb{R}^d \times [0,T])$ and that together with its partial derivatives of order $\leq [(d+2)/2]+2$, the solution ρ is integrable uniformly in $t \leq T$, we need the following assumption for the aggregation kernel:

$$G_a \in C_b^{[(d+2)/2]+2}(\mathbb{R}^d), \quad \nabla G_a \in L^1(\mathbb{R}^d). \tag{30}$$

Theorem 1. [22] *Under the above assumptions, if*

$$\lim_{N\to\infty} E\left[\|h_N(\cdot,0) - \rho_0(\cdot)\|_2^2\right] = 0,$$

then

$$\lim_{N\to\infty} E\left[\sup_{t\leq T} \|h_N(\cdot,t) - \rho(\cdot,t)\|_2^2\right] = 0 \tag{31}$$

where ρ is the unique solution of (21).

Equation (31) implies that

$$\lim_{N\to\infty} \left\langle X_N(t), f\right\rangle = \left\langle X(t), f\right\rangle = \int f(x)\rho(x,t)dx \tag{32}$$

uniformly in $t \in [0,T]$, for any $f \in C_b^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

Expression (31) states that if at time $t = 0$ the regularized function converges as N increases to ρ_0 in the L^2-norm, we have a "law of large numbers" (in the L^2-norm) at any time. Furthermore the convergence of the regularized function implies the weak convergence of the measure $X_N(t)$, to the measure $X(t) = \rho(x,t)dx$, uniformly in t.

Acknowledgments: It is a great pleasure to acknowledge fruitful discussions with Akira Okubo who inspired this research. Thanks are also due to Simon Levin, Andrea Di Liddo and Silvia Boi for relevant discussions.

During the revision process we have noticed the publication of a relevant paper [21], that might be of interest to the reader.

References

1. Billingsley P.,"Convergence of Probability Measures ", (Wiley series in probability and mathematical statistics. Tracts on probability and statistics). New York : John Wiley & Sons, 1968.
2. Boi, S., Capasso V. , Il comportamento aggregativo della specie *Polyergus* rufuscenslatr.: dalle osservazioni al modello matematico, Internal Report, Dipartimento di Matematica, Università di Milano (1997)
3. Boi, S., Capasso V., Morale, D. "Modeling the aggregative behavior of ants of the species *Polyergus rufescens*.", Nonlinear Analysis: Real World Applications, I, 2000, p. 163-176.
4. Carillo, J. *Entropy solutions for nonlinear degenerate problems*, Arch. Rat. Mech. Anal. **147** (1999), 269-361.
5. Diekmann O., Gyllenberg M., Thieme H. R., *Lack of uniqueness in transport equations with a nonlocal nonlinearity*, Math. Models and Meth. in Appl. Sciences **10** (2000), 581-592
6. Durrett, R., Levin, S.A., " The importance of being discrete (and spatial)." Theor. Pop. Biol., **46**, 1994, 363-394.
7. Evans L.C., *Partial Differential Equations* (AMS, Providence, 1998).
8. Ethier, S.N., Kurtz T.G.," Markov Processes: Characterization and Convergence." Wiley, New York, London, Sidney, Toronto, 1986.
9. Forel A. *Les Fourmis de la Suisse* 2nd edition Imprimerie Coopèrative, La Chaux-de-Fonds(1920).
10. Forel, A. *Le monde social des fourmis du globe*, **5** Libraire Kundig, Geneve(1923).
11. Friedman A., "Stochastic Differential Equations and Application", Vols. I and II, Academic Press, London, 1975.
12. Gärtner, J., "On the McKean-Vlasov limit for interacting particles." Math. Nachr., **139**, 1988, 197-248.
13. Grünbaum, D. "Translating stochastic density-dependent individual behaviour with sensory constraints to an Eulerian model of animal swarming." J. Math. Biology,**33**(1994), 139-161.
14. Grünbaum, D., Okubo, A. "Modelling social animal aggregations" In "Frontiers of Theoretical Biology" (S. Levin Ed.), Lectures Notes in Biomathematics, 100, Springer Verlag, New York, 1994, 296-325.
15. Gueron, S., Levin, S.A., "The dynamics of group formation. Math. Biosci., **128**, 1995, 243-264.
16. Gueron, S., Levin, S.A., Rubenstein, D.I., "The dynamics of herds: from individuals to aggregations, J. Theor. Biol., **182**, 1996, 85-98.
17. Harman, J.R. Some aspects of the ecology of the slave making ant, Polyergus lucidus. *Entomol. News*, 79, 217-223(1968).
18. Huth, A., Wissel C. "The movement of fish schools: a simulation model." In Biological motion. (W. Alt, G.Hoffman Eds.), Lectures Notes Biomath., Vol. 89, pp.577-590. Berlin, Heidelberg, New York Springer, 1990.
19. Ladyzenskaja O.A., Solonnikov V.A., Uralceva N.N., *Linear and Quasilinear Equations of Parabolic Type* (Translation of Mathematical Monographs, Volume 23, AMS, Providence, 1968).
20. Levin, S.A., "Conceptual and methodological isuues in the modeling of biological aggregation", in "A Consensus of Individuals: The study of 3D Animal Aggregation" (J.K.Parrish, W.M. Hamner, C.T. Prewitt, Eds),1995.
21. Mogilner A., L. Edelstein-Keshet, A non-local model for a swarm, J. Math. Bio. 38 (6) (1999), 534-549.
22. Morale, D., Capasso, V., Oelschläger K. "A rigorous derivation of a nonlinear integro-differential equation from a SDE for an aggregation model", Preprint 98-38 (SFB 359) - Reaktive Strömungen, Diffusion und Transport", IWR, Universität Heidelberg, Juni 1998.
23. Morale, D. "Cellular automata and many-particles systems modeling aggregation behaviour among populations.", 1998. In preparation.
24. Murray, J. "Mathematical Biology", Springer-Verlag, Berlin, Heidelberg, 1989.
25. Nagai, T., Mimura M. "Some nonlinear degenerate diffusion equations related to population dynamics", J. Math. Soc. Japan, **35** (1983), 539-561.
26. Nagai, T., Mimura M., "Asymptotic behaviour for a nonlinear degenerate diffusion equation in population dynamics, SIAM J. Appl. Math., **43**, 1983, 449-464.
27. Oelschläger K. "On the derivation of reaction-diffusion equations as lilit dynamics of systems of moderately interacting stochastic processes", Prob. Th. Rel. Fields, **82**,(1989), 565-586.
28. Oelschläger K. "Large systems of interacting particles and porous medium equation", J. Differential Equations, **88**(1990),294-346.
29. Okubo, A. "Dynamical aspects of animal grouping: swarms, school, flocks and herds." Adv. BioPhys., **22**, (1986), 1-94.
30. Okubo, A., S. Levin "Diffusion and ecological problems : Modern Perspectives", Springer-Verlag, 2002.
31. Satsuma J., " Animal aggregation and solitons: exact solutions for nonlinear diffusion equations, in Special Issue "Solitons, Mathematical Sciences, 1985, 108-113 (in Japanese).
32. R.E.Showalter, *Monotone Operators and Nonlinear Partial Differential Equations in Banach Space* (AMS Mathematical Surveys and Monographs, Providence, Rhode Island, 1997).
33. Warburton, K., Lazarus, J., "Tendency-distance models of social cohesion in animal groups , J. Theor. Biol., **150**, 1991, 473-488.
34. Wassman, E. Die Ameisen, die Termiten ind ihre Gaste. Regensburg Verl.- Anstat Vorm. G.I. Manz Ag (1934).