

# A Multi-agent approach for *in silico* simulations of micro-biological systems

Proverbio D., Gallo L., Passalacqua B., Pellegrino J., Maggiora M.

## Abstract

Using a Multi-agent systems paradigm, the present project develops, validates and exploits a computational *testbed* that simulates micro-biological complex systems, namely the aggregation patterns of the social amoeba *Dyctiostelium discoideum*. We propose a new design and implementation for managing discrete simulations with autonomous agents on a microscopic scale, thus focusing on their social behavior and mutual interactions. Then, the dependence on the main physical variables is tested, namely density and number of amoebas; in addition, we analyze the robustness of the dynamics against various noise sources. Along with these results, we suggest a methodology for further studies that make use of our validated model.

## 1 Introduction

The behavior of the social amoeba *D. discoideum* [1,2] represents an archetype of swarm intelligence. It has inspired models of distributed control [3] and a bio-inspired model for problems of decentralized gathering [4], cell signaling and chemotaxis [5–7]. In fact, if food sources (bacteria or agar) are available, each cell lives and proliferates individually; once starving, though, a multitude (up to 100,000 cells) of single amoebas starts sending pulses of 3',5'-Cyclic Adenosine Monophosphate (cAMP) molecules [8] to its surroundings [9–12] which, after being received and transduced by membrane receptors [13], let *D. discoideum* migrate via chemotaxis [14–16]. The final stage of the gathering process is the formation of a multicellular slug organism [17,18].



Figure 1: A gathering of *D. discoideum* [19]. Cells can be seen migrating (some individually, and some in streams) toward a central body [20].

To inquire *D. discoideum* dynamics, the most common modeling approaches tackle such systems via continuous diffusion equation [21–23], eventually solved with Monte Carlo methods [24,25] or, more recently, with in-lattice discrete simulations [26,27]. Other than that, the last years IT developments made it possible to analyze the microscopic processes underlying self organization of a colony while relying on a hybrid continuous-discrete modeling approach (still on a macroscopic scale) [28,29].

Besides them, the main contribution of the present project is to follow a *Multi-Agent Systems* computational approach to a) build an agent architecture that is consistent with biological knowledge; b) define metrics to quantitatively study the gathering process; c) validate the model by means of comparison with well-established results and by generation of predictive data.

## 2 Methodology

The change of modeling paradigm was justified as the aforementioned methods from literature mainly describe the dynamics from a macroscopic perspective [30]. However, they are not capable of inquiring the emergence of complex patterns from microscopic behaviors [23]; at the same time, it is impossible to gauge the impact of individual stochasticity or failure on the overall evolution of the system, thus resulting in poor resolution on the system control side [31]. Therefore, rather than following said approaches, we exploited a bottom-up paradigm, namely that of *Multi-Agent Systems* (MAS).

An *agent* is an individual “computer system that is capable of independent action on behalf of its user or owner [...] A *multi-agent system* is one that consists of a number of agents, which interact with one another, typically by exchanging messages through some computer network infrastructure” [32]. As a special feature, these models allow to account for intracellular decision processes which are triggered by biochemical cell-cell or cell-matrix interactions [33]. In addition, they provide natural candidates for modeling the evolution and pattern formation of large multi-cellular systems [34] since they tie cellular properties to macroscopic behavior on the population level [35–37].

Other than that, the agent-based paradigm allows for a natural management of communication issues among individuals: specifically, we were able to model the cAMP pulses without using an instantaneous *point to point* messaging [38] nor diffusing waves [39] (as it was done in most of continuous reaction-diffusion models [22, 26, 28]). On the contrary, modeling the signals as discrete traveling packages reduces the computational complexity. Given  $N$  the number of agents, we obtain  $\mathcal{O}(N)$  instead of state-of-the-art  $\mathcal{O}(N \log^2 N)$  for fractional diffusion equations [40], thus allowing better scaling for large populations. Validity of such vectorial approach was checked by comparing generated data (qualitative and quantitative ones) with those in literature that make use of diffusing waves or geometrical approaches [22, 24, 28, 41, 42]. Finally, since the dynamics is generated during each simulation run, such models may be used as testbeds to control hypothesis and to generate predictive data [43].

As a result, we obtain a finite set of (logical) behavioral rules that lead to an emergence of collective synchronization, and we verify if and under which conditions the global self-aggregation is obtained. In particular, thanks to the chosen scale, both the microscopic and macroscopic variables that define the model, namely the number and the density of cells, were studied and their role assessed. To end with, predictions about the system behavior are made, discussed and used to assess its robustness.

## 3 Design and Implementation

Our model consists of three main agent species: a rigid, non-toroidal, dynamically evolving 2-D [44] grid (the environment), amoeba agents (simulating single cells) and vectorial cAMP packages (carrying signal).

After a thorough literature search (see Sec. 1 as well as [45–48]), Amoeba agents are implemented with individual behaviors embedded in a complex architecture, as they are not purely reactive, yet keep information about the environment and their past history (see Fig. 2). Note that all Amoeba agents are alike (no pace-maker cells were introduced at a first approximation [49]). They are given two specific states: wandering  $\mathbb{W}$  (there is food, so the amoeba is eating) and starving  $\mathbb{S}$  (the amoeba is *not* finding food, so it begins the aggregation). As we are only interested in looking for gathering patterns, we didn’t add “aggregated” or “slug” states [50]. Amoebas are capable of the following: while their inner state is  $\mathbb{W}$ , they wander and eat the bacteria; as soon as their state becomes  $\mathbb{S}$ , they isotropically shoot signals (with a  $\pm 20^\circ$  random smearing  $\delta_1$ ) and, if they are reached by a traveling cAMP agent, they orientate according to the direction of the just-absorbed signal and move on (a smearing  $\delta_2 = 10\%$  on the trajectory is also considered) [48]. Since the main focus is on the aggregation patterns, we neglect the complex inner synchronization mechanisms within each cell [51]; on the contrary, we rely on the demonstrated robustness of said mechanisms [52]. Inertial effects are also neglected, as cells and micro-organisms typically exist in low-Reynold’s number environments [53].

After setting the dimension  $\dim$ , the environment is represented by a square domain  $\text{Env} = \langle x_k; y_k \rangle$  (where  $k = 0 \dots \dim \in \mathbb{N}$ ), each cell of it being assigned a scalar value (whose evolution might be

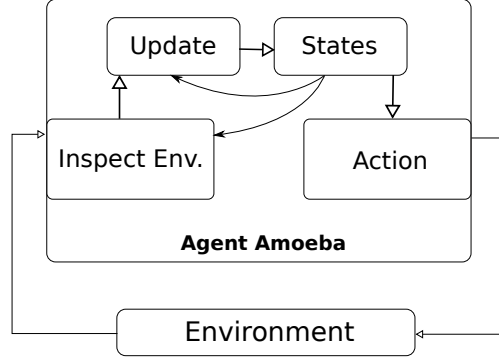


Figure 2: Schematic representation of the Amoeba agent architecture and its environment. Given the set of configurations  $\{E_g\}$  of the environment  $Env$ , the agent is a tuple  $A_m = \langle inspect, internal, action \rangle$  where “Internal” is a set of features  $\{update, states\}$  that are not directly interfacing with the environment (those that distinguish the agent from a purely reactive one). In particular,  $States = \{S_j\}$  is the set that defines the internal peculiarities of each agent and that selects the other methods;  $Inspect = \{I_i^j: E_g \mapsto per\}$  sees  $E_g$  and generates an internal perception  $per$ ;  $Update = \{U_i^j: per \mapsto S_j\}$  may update the states according to the perceptions;  $Action = \{A_k^j: S_j \mapsto E_{g'}\}$  is the set of feasible actions on  $Env$  according to the state the agent is in.

ruled by appropriate parameters) that represents the food sources, the presence of which makes the *in-silico* experiment more realistic and better adherent to wet-lab biological observations, where several factors may affect amoebas behavior. Formally, at a given time  $t \in \mathbb{N}$ , a food source ( $m$ -th bacteria) is  $b[t] : (\{x_m, y_m\} \in Env, t) \mapsto b \in \mathbb{N}$ .

cAMP agents are sent vectorial signals, representing discrete packages of chemicals: they are shot radially by amoebas and they travel across the environment, with a speed  $v_{cAMP}$  whose module is equal to that of the field mean diffusion speed, and are eventually absorbed by other amoebas or the borders. Consequently, what varies is not precisely the diffusive gradient, but the probability associated to finding a cAMP agent with defined direction: from Agent A perspective, absorbing an Agent B’s signal is equal to  $P^A(\{I_{B-cAMP}^{starving} : E \mapsto per\}) \propto \frac{1}{r^2}$  where  $r = v_{cAMP} \cdot \hat{t}$  is the length of the cAMP trajectory ( $\hat{t}$  corresponds to the elapsed time between emission from A and absorption by B).

To implement the code, we chose the MAS platform GAMA [54], that lets us a) handle the agents consistently with the desired design; b) manage the parameters easily; c) have a user-friendly interface showing the whole environment that allowed us to inspect what was happening in real time and to obtain informative animations.

## 4 Launching an experiment

An experiment is defined by the choice of the agents parameters, the run of repeated simulations and the measurements of various quantitative variables.

### 4.1 Parameters and variables

The discussed model contains different free parameters, whose value is to be set. These were chosen to be biologically consistent [18, 22, 45–47], although scaled (note that a simulation cycle equals 1 min). Therefore the values of reference are: diameter of amoeba  $D = 1.8$  units (while the real typical dimension is at order of  $18\mu m$ ); speed of amoeba  $v_A = 0.3 \frac{units}{cycle}$  [55], which corresponds to the average speed leading to the “visual” starting of an aggregate after six hours [18]; diffusion speed of cAMP molecules  $v_c = 0.9 \frac{units}{cycle}$ ; cAMP shooting time  $t_s = 10$  cycle, this latter indicating the mean period between two cAMP shootings. The reference value for dim (edge of the environment) is 100 units =  $55.5 \times D = 1$  mm. Note that, from now on, the measure unit *units* will be implied.

The model manages “thermal noise” parameters as well:  $P_G$  represents the growth probability  $P(b[t + 1] \leftarrow b[t] + 1)$  of bacteria in the environment, namely the probability that a bacteria cell increases its

scalar value by one at the given time;  $P_A$  is the “agitation probability”  $P_A = P(\{I_i^j : E_g \rightarrow \text{per}\})$ , that represents a biological source of detection failure, namely an asymmetric binding receptor occupancy that, during chemotaxis, leads to a noisy input signal or signal detection failure [56]. When interpreting the whole system from a statistical mechanics perspective, the latest parameters represent the contribution on the thermal  $\beta$ , therefore the term “noise”. During the process of validating the model these values were set arbitrarily; their impact on the dynamics is studied shortly afterwards.

As stated previously, we are interested in understanding the dependence of the model on the variables density ( $\rho$ ) and number ( $N$ ) of amoebas (macroscopic and microscopic variables, respectively), which are mutually related by

$$\rho = \frac{N \cdot A}{\text{dim}^2} \quad (1)$$

with  $A$  being the area of a single amoeba. The reference density ranges from  $\frac{80 \cdot A}{100^2}$  up to  $\frac{600 \cdot A}{100^2}$  and greater, the same magnitude as the (typically biological) range of  $(\frac{10000 \cdot A}{\text{cm}^2} \div \frac{100000 \cdot A}{\text{cm}^2})$  as reported in [18, 22]. Therefore, in order to tune  $\rho$ , we need to act over  $\text{dim}$  alone, holding  $N$  fixed, in order to separate the contribution of the microscopic variable from the macroscopic one. After that, we adequately rescale the “spatial” parameters, namely  $v_A, v_C, P_G$ . On the contrary, as  $D$  is already involved in  $A$ , we do not scale it. From now on, “ $\cdot A$ ” will be implied.

## 4.2 Simulations

Given an initial random cell configuration and a certain bacteria distribution in the domain, a run is characterized by the steps depicted in Fig. 3, 4, 5.

### Example: evolution of a single simulation

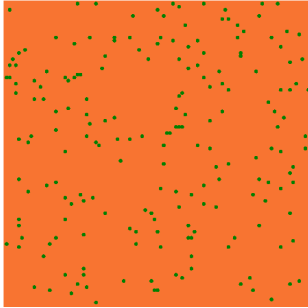


Fig. 3: Initial setting of an average simulation.

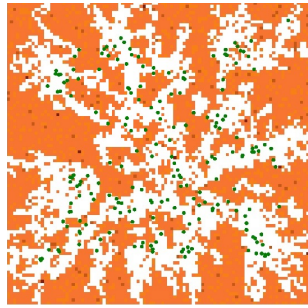


Fig. 4: Evolution and movement.

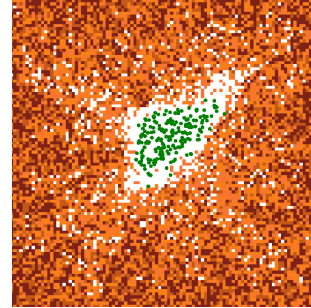


Fig. 5: Aggregation and simulation end.

Three steps of a simulation with a relatively small number of amoebas (for clarity reasons). Setting: green dots are scattered amoeba cells, the orange landscape depicts a uniform bacteria distribution (Fig. 3). Run: amoeba agents move dynamically across the environment, while bacteria growth (brown pixels) is ruled by  $P_G$ . White areas are without bacteria, as amoebas ate them in previous cycles. At each cycle, the new position of an agent is calculated according to its direction and its speed. cAMP agents are not shown, although traveling across the environment at the same time as amoebas (Fig. 4). Aggregation: amoeba cells form an aggregate. As we didn’t implement an “aggregated” state, the slug is hold tight solely by mutual movement along signals and not by a potential field representing cellular adhesion (Fig. 5).

When it comes to qualitative features of a certain system such as pattern formation, the first an experimentalist does is to look for visual similarities [57]. As a consequence, the behaviors observed in animations have been confirmed by means of comparison with biological snapshots, videos (Fig. 1, [58]) and results from previous models [22, 24, 26, 28, 29, 41, 59]. Finally, repeated simulations for each set of parameters and variables are run in order to obtain statistics.

### 4.3 Measurements

Being the first step to validate the present model, we focused on the aggregation rate. The order parameter is the time  $t$  defined by simulation cycles. The chosen quantitative variable to measure the rate is the marginal posterior variance along the two axis  $x$  and  $y$ . Being  $\{x_n, y_n\}$  the coordinates of the  $n$ -th amoeba in  $Env$ , for each  $i$ -th simulation we measure

$$\text{var}^i(t) = \frac{1}{\text{dim}^2} \frac{1}{N} \sum_{n=1}^N \|E(x_n, y_n)(t) - (x_n, y_n)(t)\| \quad (2)$$

were  $\bar{x}(t)$  is the mean value at each cycle  $t$ . This way we get information about the global behavior towards aggregation around a specific center as the *dispersion* of the colony. Note that, thanks to the rotational invariance of  $Env$ ,  $\frac{1}{\sqrt{2}} \text{var}^i(t) = \text{var}_X^i(t)$ , so that we can directly use the latter for further analysis (it is lighter to be directly calculated and saved by GAMA software). In Fig. 6 is shown the evolution of the dispersions  $\text{var}_X^i$  and  $\text{var}_Y^i$  during time, for one simulation, as depicted by GAMA interface.

Having repeated measures, it is natural to evaluate the mean value and the associated statistical error as the standard deviation of  $\text{var}_X^i(t)$ :  $(\text{var}_X(t) \pm \sigma_{\text{var}_X}(t))$ . This method will be repeated for each experiment.

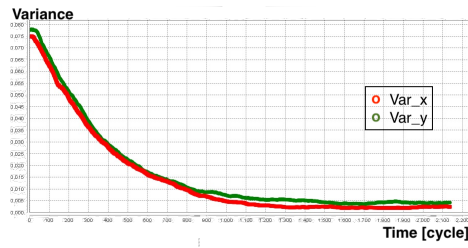


Figure 6: Evolution of the marginal dispersions  $\text{var}_X^i$  and  $\text{var}_Y^i$  during time, as shown by GAMA graphic interface.

## 5 Analysis

After having obtained  $(\text{var}_X^k(t) \pm \sigma_{\text{var}_X}^k(t))$  for each  $k$ -th experiment (that is, each choice of physical variables), they were compared by focusing on the very first stages of the aggregation process. Since  $\text{var}_X$  vs  $t$  follows an exponential rule:

$$\text{var}_X(t) = C + A \cdot e^{-B \cdot t} \quad (3)$$

it was informative to study how the transient behavior towards the aggregate differs quantitatively from experiment to experiment. Therefore, the interval of interest is that in which  $|\frac{d\text{var}_X}{dt}| > 1$ , namely where the exponential trend is stronger. As stated previously, when *in vitro* experiments are concerned, biologists begin to see a “visual aggregate” after approximately 6h [18], corresponding to circa 400 cycles in our simulation. Hence, such non-informative lag was ignored; additionally, by doing so, it was possible to neglect such a transient whose characteristics are mainly due to the initial conditions. Thus we focused on the interval  $\mathbb{I} = \{t : 400 < t < \tilde{t}\}$ , where  $\tilde{t}$  is such that  $|\frac{d\text{var}_X}{dt}|_{t=\tilde{t}} \leq 1$ . The value  $t = 1000$  was then conventionally chosen as the representative for  $\mathbb{I}$ ; consequently, what is going to be studied to estimate the quantitative differences between each  $k$ -th experiment is  $\text{var}_X^k(t=1000)$ . An example of the process of defining  $\text{var}_X^k(t=1000)$  is seen in Fig. 7.

## 6 Results

To begin with, it was assessed the impact of the variables that define the model scale on the dynamics, namely  $\rho$  (macroscopic) and  $N$  (microscopic). In all the experiments we chose both the cell food source

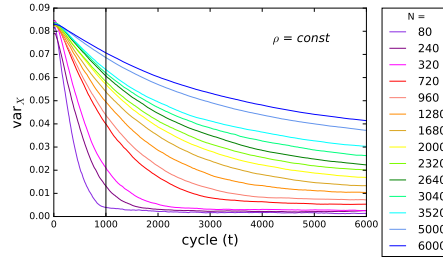


Figure 7: Several experiments at  $\rho = \text{const}$ , thus varying  $N$ . Each colored curve depicts  $\text{var}_X^k$  vs  $t$ , whose trend is exponential. The vertical line corresponds to  $t = 1000$  so that, for each experiment  $k$ ,  $\text{var}_X^k(t=1000)$  is given by the intersection between such line and the  $k$ -th curve. Note that, for clarity reasons,  $\sigma_{\text{var}_X^k}$  is not shown.

and the amoebas to be uniformly distributed over the environment.

### 6.1 $N = \text{const}$

The first one to be inquired is the macroscopic variable  $\rho$ , namely the density of amoebas. In order to do so, three different  $N$  (respectively 250, 600, 1200) were selected and kept constant, thus defining three experimental sets; then the density for each  $k$ -th experiment was varied. As stated before,  $\text{var}_X^k(t=1000)$  was calculated as the quantitative estimator. Global results are shown in 8.

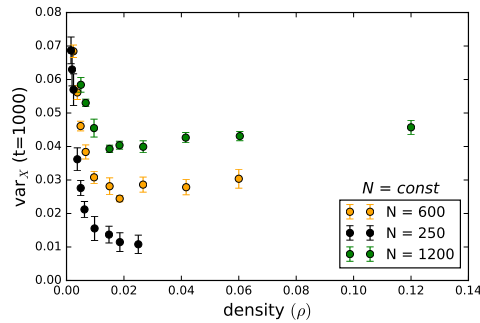


Figure 8: Three set of experiments, each of them characterized by a different  $N$  value, are shown. Note that they show the same qualitative behavior. The statistical associated errors are evaluated as explained before.

As it can be seen from the chart, we can heuristically say that there are two distinct dependences of  $\text{var}_X$  on  $\rho$ . For small values of  $\rho$ ,  $\text{var}_X$  decreases as the density increases, while for bigger values  $\text{var}_X$  seems to "saturate" reaching a "plateau". Considering the chart, it is important to stress that: a) the critical density at which the functional dependence changes is not independent of  $N$ ; b) the value of the "plateaus" of  $\text{var}_X$  increases with  $N$ . Not only that: what other studies called the "typical density range" [22] seems to correspond to the "saturation" region only; we believe that our analysis, at least in the present model, have succeeded in highlighting the presence of a lower region in which the system shows a different behavior.

Finally, while in continuous models it was generally accepted that the most important variable was  $\rho$  alone, it is evident that the dynamics of our MAS discrete system depends on the number of amoebas as well; hence, it is necessary to study the numerosity, looking for a scale-free range in which the dependence on it can be neglected.

## 6.2 $\rho = \text{const}$

In order to evaluate the impact of  $N$ , three different set of experiments were chosen, each of them with a specific constant density (respectively  $\rho = \frac{80}{100^2}, \frac{120}{100^2}, \frac{200}{100^2}$ ). After that,  $N$  was varied for each  $k$ -th experiment. The results are shown in Fig. 9.

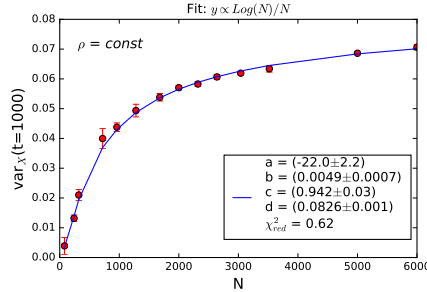


Figure 9: The quantitative estimator  $\text{var}_X(t=1000)$ , as well as its associated statistical error, was calculated as described previously. As every experiment set showed the very same dependency, thus the same trend, only the one identified by the value  $\rho = \frac{80}{100^2}$  is depicted.

As the  $\chi^2$  guarantees, the fit follows

$$\text{var}_X(t = 1000) = a \cdot \frac{\log(b \cdot N + c)}{N} + d \quad (4)$$

therefore resulting in a  $\text{var}_X(t = 1000) \propto \text{Log}(N)/N$  trend. It was verified that all sets of experiments follow the very same rule, albeit with different values for the fit parameters. As a consequence, hereafter we focus on one particular experiments set, that is  $\rho = \frac{80}{100^2}$ .

The function (4) tends to an asymptote, so in principle it is possible to identify a scale-free range for the variable  $N$ ; however, the value of  $N$  for which  $\text{var}_X(t = 1000)$  differs by 5% from the asymptotic value 25,000, which is a biological magnitude of numerosity. Despite its relevance for evolutionary studies, said scale was found not to be handleable with ease since it exceeds our own computational limits; future simulations with greater computation capabilities may exploit this result to set their variables in a purely scale-free region. Moreover, it may be interesting, for the purpose of scaling the model to lower values of  $N$ , to understand and quantify the error that comes from such scaling choice. Considering  $\text{var}_X(t = 1000) = f(N)$  a proxy to identify an interval

$$\mathbb{L} = \{\tilde{N} \text{ s.t. } 0 < \frac{\partial f}{\partial N} \Big|_{\tilde{t}} < 1 \text{ and } \tilde{N} \text{ is computationally treatable}\}$$

we chose the value  $\tilde{N} = 2640$  to represent  $\mathbb{L}$ . Starting from that, it was possible to estimate the induced systematic error associated with MAS models that are placed in  $\mathbb{L}$  instead of in a scale-free regime.

## 6.3 Systematics

The aforementioned systematic error was not determined on the estimator  $\text{var}_X(t = 1000)$  but directly on the order parameter  $t$  instead. The reason is the following: in a real biological experiment, one cannot control the evolution of the amoeba dispersion, which necessitates an extensive control over every single cell, but can easily measure the time needed to reach an aggregation state. Knowing the dependences given by eq. 3 and 4, it is straightforward to obtain, in  $\mathbb{L}$ :

$$t(N) = -\frac{1}{B} \cdot \log\left\{\frac{a \cdot \text{Log}[b \cdot N + c]}{N} + d - C\right\} \quad (5)$$

(recall that it was verified that  $t = 1000$  is a valid proxy for every  $t \in \mathbb{L}$ ).

After that, once the desired aggregation state (defined by a value  $\text{var}_X$ ) is chosen, one can measure  $t_{thr}$ ,

namely the time that is needed to reach it (Fig. 10).

After that, the systematic error is given by  $\delta t(N, \Delta N)$ . To evaluate it, we assumed a range of uncertainty on the number of amoebas  $\Delta N$ ; then we hypothesized a linear dependence of  $\delta t$  on  $\Delta N$  as it was an error propagation. Such hypothesis was verified *a posteriori* by evaluating the values of  $\delta t_{exp}$  coming from interpolation of experimental data; to do so, we selected several values of  $\hat{N}$  around  $N = 2640$  to define a  $\Delta N$ ; for each of them we evaluated the corresponding  $\text{var}_{\hat{X}}^{\hat{N}}(t = 1000)$  using equation (4); finally, we estimated  $\hat{t}$  such that  $\text{var}_{\hat{X}}^{\hat{N}}(t = \hat{t}) = \text{var}_{\hat{X}}^{2640}(t = 1000)$  by using equation (3). Thus

$$\delta t(N, \Delta N) = \frac{\partial}{\partial N} \left( -\frac{1}{B} \cdot \text{Log} \left\{ \frac{a \cdot \text{Log}[b \cdot N + c]}{N} + d - C \right\} \right) \Delta N \quad (6)$$

In order to contrast the systematics against the sensitivity of the proposed model (so that one may estimate how much the magnitude of  $\delta t$  impacts on the statistical analysis), three more experiments were performed, each of them being assigned a different value of density. This way we consider the sensitivity as the capability of discriminating different regimes depending on the macroscopic variable  $\rho$ . With reference to the previous results, the three values were chosen around the critical density which leads to “saturation”:  $\rho = \left\{ \frac{80}{100^2}, \frac{104}{100^2}, \frac{120}{100^2} \right\}$  (of course, as  $N$  of interest is  $N = 2640$ , they were scaled accordingly). Thanks to such decision, it was possible to compare the sensitivity in and out the scale-free region of  $\rho$ . Results are shown in Fig. 10.

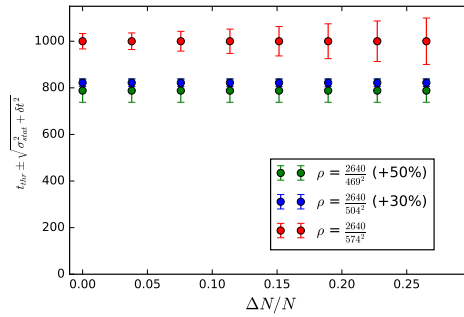


Figure 10:  $t_{thr}$  vs  $\Delta N/N$ . Three sets of experiments are shown and compared. The error bars are a superimposition of statistical and systematic error, so that it is possible to estimate how big the total error is in comparison with the sensitivity. Calling the critical density  $\tilde{\rho}$ , note that the two experiments with  $\rho > \tilde{\rho}$  (namely the blue and green one) are already indistinguishable, as they are in a density-scale-free region, while the third one ( $\rho < \tilde{\rho}$ ) can be statistically separated if and only if  $\Delta N/N$  is “low”, that is, if the associated systematics does not exceed a confidential value. Otherwise, a normal test Z cannot tell between the three sets.

There are three main identified features that emerge from such analysis.

First, it was confirmed that, in a density-scale-free region, the experimental sets defined only by the choice of  $\rho$  are not distinguishable - even without considering the systematics; as a consequence, we can identify an *intrinsic*  $\rho$ -scale-free region that does not depend on measurement performances.

Second, the experimental mutual choice of  $N$  and  $\Delta N$  has a significant impact on  $\delta t$ , that is, if we work with a number of amoebas which is lower than the biological ones, we ought to pay particular attention to its uncertainty. It is also true the opposite: when setting the system to biological numbers,  $\Delta N$  may be tuned less accurately, as it doesn't impact too much on the aggregation trend; this idea may lead to biological suggestions like “how the evolution tuned the colonies to large numbers that are minimally dependent on stochastic fluctuation”.

Third, it is clear that, when modeling a computational discrete simulation, one must pay attention not only to the macroscopic variable  $\rho$ , but to the microscopic  $N$  as well. If not, in the case that one notices indistinguishable results, he couldn't say for sure if they were intrinsically so or, on the contrary, a systematic error shaded them. Therefore, when setting a computational model, there is not such a thing like a “ $\rho$ -scale-free region” or a “ $N$ -scale-free region” alone, but there is a “ $\rho$ - $N$ -scale-free surface” where



the experimentalists should set their variables, knowing that a certain error is associated to sets that are outside such surface (but not too far, as we stressed during the whole analysis).

## 7 System robustness

In biological systems, robustness is defined as the ability to maintain functionality in the presence of internal and external perturbations [60, 61]; uncovering its mechanisms is a key issue in system biology [62]. In the present context, it is of great importance to study the “thermal-noise” parameters, namely  $P_G$  and  $P_A$  (Sec. 4.1).

In consistency with the analysis conducted so far, we set  $N = 2640 \in \mathbb{L}$  and  $\rho = \frac{2640}{469^2}$  (in the “density-scale-free” region). For the reasons stated before, it was estimated ( $t_{thr} \pm \sigma_{stat}$ ) as well as the associated systematic error  $\delta t$  for different values of  $\Delta N$  (conventionally, a big a medium and a small one:  $\Delta N = \{100, 300, 600\}$ ). Results of ( $t_{thr} \pm \sqrt{\sigma_{stat}^2 + \delta t^2}$ ) vs  $P_G$  and vs  $P_A$  are shown in Fig. 11.

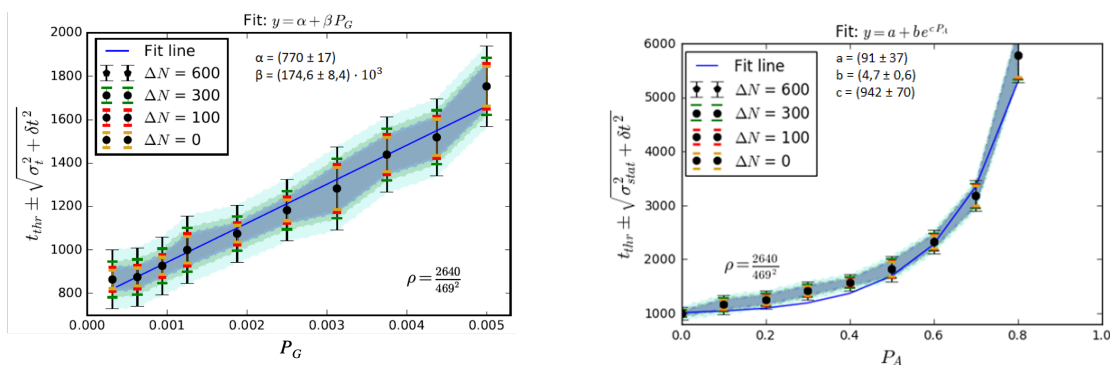


Figure 11: Experimental results for the dependence of the aggregation time on  $P_G$  (left) and  $P_A$  (right). Along with the mean values, four different error bars are shown: the statistical alone (with  $\Delta N = 0$ ), and three superimposed systematic, each of them depending on a different  $\Delta N$ . With regard to  $P_A$ , the linear fit is guaranteed by  $\chi_{red}^2 < 1$  for every associated error. As for  $P_A$ , despite the clear exponential trend,  $\chi_{red}^2 < 1$  only with a  $\delta t$  ( $\Delta N = 600$ ) associated error. Once again, we see that the associated systematics is not a formal, *a priori* accident, but it is a necessary and informative quantity that must be taken into account when working within the considered parameter surface.

Regarding the role of  $P_G$ , big values of  $\Delta P_G$  have to be chosen in order to magnify the linear trend (in fact, if  $P_G = 0.00125$  is the reference value, the study interval is  $\mathbb{P} = \{-75\%; +300\%\}$ ). Therefore, large perturbations are necessary to significantly increase the aggregation time. On the other hand,  $P_A$  leads to a possible critical slowing down of the system [63] towards the loss of attractor when values are too high. However, the aggregation rate is nonetheless consistent to biological values up to  $P_A \simeq 0.5$ , meaning that the system as a whole is robust even in adverse cases when individuals can fail half of the times to correctly process the chemical signal (inner noise). In other words, non-linearity of the problem balances different sources of noise, leading to controlled mean collective behaviors.

## 8 Validating model predictions

The model makes two important preliminary predictions about the colony evolution: (i) a finite set of individual behavioral rules, coupled with social interactions, are sufficient to elicit self-organization and decentralized gathering; (ii) thanks to mutual interactions, the system is robust against numerous sources of stochasticity. The first prediction has already been shown during the whole text, as it is the main verified hypothesis that led to the model design and implementation; however, we believe that, by expanding the class of candidate models, we may be able to produce an even better model. As for the second prediction, not only it has already been shown how the system collective behavior is robust against biological-like smearing of trajectories ( $\delta_1, \delta_2$ ), but “thermal-noise” sources has been studied, too. As a result, we showed how the system is able to cope with internal and external perturbations,

evolving in its original basin of attraction until said perturbations became greater than those found in nature.

To summarize, all the results (comparison with literature results after having set the same initial conditions, comparison with biological knowledge, assessment of a scale-free region for simulation repeatability and system robustness, validation of predictions) concur in the model validation, as per guidelines followed by other simulations [64,65].

## 9 Conclusions

A *Multi-agent system* approach could represent an innovative methodology to tackle complex biological systems such as the *D. discoideum* aggregation process. In the present study we firstly validated our model. To begin with, it was identified a “scale-free surface” in which setting repeatable and reliable simulations; likewise previous research, which speaks of a “normal density region”, we highlighted the presence of a  $\rho$ -scale-free-region towards which the system tends when values are biologically compatible. Not only that: by studying the number of individual cells, we demonstrated that, at least in a MAS approach, the macroscopic variable  $\rho$  alone is not sufficient to fully set the model, but that a researcher needs to care about the microscopic variable  $N$  as well. As a consequence, we highlighted a  $\rho$ - $N$ -scale-free surface in which the experimentalist should place his variables during simulations. However, since such region is identified by values that are computationally hard to manage, we analyzed its neighborhood for  $N$  slightly off-scaled ( $\tilde{N}$ , see Sec.6.2); after that, we suggested how to consider the induced systematic error. This analysis states that it is important to handle an eventual uncertainty over  $N$  with care, as it may lead to ambiguity during the comparison of different set of variables and/or hypothesis. On the contrary, if  $\Delta N$  is treated adequately, the present analysis shows how to obtain quantitative results for a rescaled model to smaller, non-biological values of numerosity, easier managed by an *in silico* simulation. Finally, the validated model was used to perform predictions about the system robustness against internal and external perturbations, finding interesting results about the system control.

The present model focuses on the first stages of *D. discoideum* aggregation patterns; however, an interesting future perspective may be to extend the class of possible models by going beyond the present hypothesis. Other than that, this model can represent a suitable testbed to study dynamical and topological properties of the system. Moreover, we believe that the present model could have interesting practical applications in the field of system control and self-synchronization of robotic populations [66–69].

To end with, we suggest that our project could represent a link between an abstract and theoretical mathematical modeling and a biological and more qualitative description because, on one hand, the replica of the physical space (“wet lab-alike simulation”) guarantees a high level of interpretability; on the other hand, the extensive control of the parameters and the possibility to realize quantitative measures and repeated tests allows going beyond the experimental limits of the *in vivo* or *in vitro* observations.

## Acknowledgments

The authors would like prof. P. Terna and dr. J. Markdahl for valuable feedbacks.

## References

- [1] Kessin R. H. *Dictyostelium: evolution, cell biology, and the development of multicellularity*, volume 38. Cambridge University Press, 2001.
- [2] Romeralo M., Baldauf S., and Escalante R. *Dictyostelids: evolution, genomics and cell biology*. Springer Science & Business Media, 2013.
- [3] Bonabeau E., Dorigo M., and Theraulaz G. *Swarm intelligence: from natural to artificial systems*. Number 1. Oxford university press, 1999.

- [4] Lana Strmecki, David M Greene, and Catherine J Pears. Developmental decisions in dictyostelium discoideum. *Developmental biology*, 284(1):25–36, 2005.
- [5] Bonner J.T. Induction of stalk cell differentiation by cyclic amp in the cellular slime mold dictyostelium discoideum. *Proceedings of the National Academy of Sciences*, 65(1):110–113, 1970.
- [6] Devreotes P. Dictyostelium discoideum: a model system for cell-cell interactions in development. *Science*, 245(4922):1054–1058, 1989.
- [7] Jeffrey G Williams. Dictyostelium finds new roles to model. *Genetics*, 185(3):717–726, 2010.
- [8] Morris H.R., Taylor G. W., Masento M. S., Jermyn K. A., and Kay R. R. Chemical structure of the morphogen differentiation inducing factor from dictyostelium discoideum. *Nature*, 328(6133):811, 1987.
- [9] Duan W., Ding H., Zhu W.G., Srinivasan K., Otterson G.A., and Villalona-Calero M.A. camp binding protein assay for widespread use in cell signaling studies. *Biotechniques*, 33(1):66–72, 2002.
- [10] Robertson A., Drage D. J., and Cohen M. H. Control of aggregation in dictyostelium discoideum by an external periodic pulse of cyclic adenosine monophosphate. *Science*, 175(4019):333–335, 1972.
- [11] Van Haastert P.J., Van der Meer R.C., and Konijn T. M. Evidence that the rate of association of adenosine 3', 5'-monophosphate to its chemotactic receptor induces phosphodiesterase activity in dictyostelium discoideum. *Journal of bacteriology*, 147(1):170–175, 1981.
- [12] O'day D. H. and Keszei A. Signalling and sex in the social amoebozoans. *Biological Reviews*, 87(2):313–329, 2012.
- [13] Hans G Othmer and Pauline Schaap. Oscillatory camp signaling in the development of dictyostelium discoideum. *Comments on Theoretical Biology*, 5:175–282, 1998.
- [14] Durston A.J. Dislocation is a developmental mechanism in dictyostelium and vertebrates. *Proceedings of the National Academy of Sciences*, 110(49):19826–19831, 2013.
- [15] Futrelle R.P., Traut J., and McKee W.G. Cell behavior in dictyostelium discoideum: preaggregation response to localized cyclic amp pulses. *The Journal of Cell Biology*, 92(3):807–821, 1982.
- [16] Bonner J.T. and Savage L.J. Evidence for the formation of cell aggregates by chemotaxis in the development of the slime mold dictyostelium discoideum. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 106(1):1–26, 1947.
- [17] Darmon M., Brachet P., and Da Silva L.H. Chemotactic signals induce cell differentiation in dictyostelium discoideum. *Proceedings of the National Academy of Sciences*, 72(8):3163–3166, 1975.
- [18] Fey P., Kowal A.S., Gaudet P., Pilcher K.E., and Chisholm R.L. Protocols for growth and development of dictyostelium discoideum. *Nature protocols*, 2(6):1307, 2007.
- [19] In Columbus B. Ficheiro:dictyostelium aggregation.jpg, 2008.
- [20] Kessin R. H. Cell motility: Making streams. *Nature*, 422(6931):481, 2003.
- [21] Gierer A. and Meinhardt H. A theory of biological pattern formation. *Kybernetik*, 12(1):30–39, 1972.
- [22] Dallon J.C. and Othmer H.G. A discrete cell model with adaptive signalling for aggregation of dictyostelium discoideum. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1351):391–417, 1997.
- [23] Baker R.E., Gaffney E.A., and Maini P.K. Pdes for self-organization in cellular and developmental biology. *Nonlinearity*, 21(11), 2008.
- [24] Palsson E. and Othmer H. G. A model for individual and collective cell movement in dictyostelium discoideum. *Proceedings of the National Academy of Sciences*, 97(19):10448–10453, 2000.

- [25] C Van Oss, Alexandre V Panfilov, Pauline Hogeweg, Florian Siegert, and Cornelis J Weijer. Spatial pattern formation during aggregation of the slime mould dictyostelium discoideum. *Journal of theoretical biology*, 181:203–213, 1996.
- [26] Nagano S. Modeling the model organism dictyostelium discoideum. *Development, growth & differentiation*, 42(6):541–550, 2000.
- [27] Pineda M.I, Weijer C. J., and Eftimie R. Modelling cell movement, cell differentiation, cell sorting and proportion regulation in dictyostelium discoideum aggregations. *Journal of theoretical biology*, 370:135–150, 2015.
- [28] Fates N. Gathering agents on a lattice by coupling reaction-diffusion and chemotaxis. 2008.
- [29] Nagano S. Diffusion-assisted aggregation and synchronization in dictyostelium discoideum. *Physical review letters*, 80(21):4826, 1998.
- [30] Grüne-Yanoff T. and Weirich P. The philosophy and epistemology of simulation: a review. *Simulation & Gaming*, 41(1):20–50, 2010.
- [31] Pablo A Iglesias and Brian P Ingalls. *Control theory and systems biology*. MIT Press, 2010.
- [32] Wooldridge M. *An introduction to multiagent systems*. John Wiley & Sons, 2002.
- [33] Galle J., Hoffmann M., and Aust G. From single cells to tissue architecture—a bottom-up approach to modelling the spatio-temporal organisation of complex multi-cellular systems. *Journal of mathematical biology*, 58(1-2):261, 2009.
- [34] Mohammad Parhizkar and Giovanna Di Marzo Serugendo. Agent-based models for first-and second-order emergent collective behaviours of social amoeba dictyostelium discoideum aggregation and migration phases. *Artificial Life and Robotics*, pages 1–10, 2018.
- [35] Drasdo D., Hoehme S., and Block M. On the role of physics in the growth and pattern formation of multi-cellular systems: what can we learn from individual-cell based models? *Journal of Statistical Physics*, 128(1-2):287, 2007.
- [36] Kennedy R. C., Ropella G. E.P., and Hunt C. A. A cell-centered, agent-based framework that enables flexible environment granularities. *Theoretical Biology and Medical Modelling*, 13(1):4, 2016.
- [37] Djamel Zeghida, Djamel Meslati, and Nora Bounour. Bio-ir-m: A multi-paradigm modelling for bio-inspired multi-agent systems. *Informatica*, 42(3), 2018.
- [38] Singhal A., Eunice E. S., and James D. A.r. *An Effective Communication Framework for Inter-Agent Communication In a Complex Adaptive System Simulation With Application To Biology*. PhD thesis, Virginia Polytechnic Institute and State University, 2006.
- [39] Robert Stephen Cantrell and Chris Cosner. *Spatial ecology via reaction-diffusion equations*. John Wiley & Sons, 2004.
- [40] Wang H., Wang K., and Sircar T. A direct  $O(n \log^2 n)$  finite difference method for fractional diffusion equations. *Journal of Computational Physics*, 229(21):8095–8104, 2010.
- [41] Elliott C. M., Stinner B., and Venkataraman C. Modelling cell motility and chemotaxis with evolving surface finite elements. *Journal of The Royal Society Interface*, page rsif20120276, 2012.
- [42] Estefania Vidal-Henriquez and Azam Gholami. Spontaneous center formation in dictyostelium discoideum. *arXiv preprint arXiv:1812.08061*, 2018.
- [43] Saetzler K., Sonnenschein C., and Soto A. M. Systems biology beyond networks: generating order from disorder through self-organization. In *Seminars in cancer biology*, volume 21, pages 165–174. Elsevier, 2011.

- [44] Bonner J. T. A way of following individual cells in the migrating slugs of dictyostelium discoideum. *Proceedings of the National Academy of Sciences*, 95(16):9355–9359, 1998.
- [45] Alcantara F. and Monk M. Signal propagation during aggregation in the slime mould dictyostelium discoideum. *Microbiology*, 85(2):321–334, 1974.
- [46] Nanjundiah V. Periodic stimuli are more successful than randomly spaced ones for inducing development in dictyostelium discoideum. *Bioscience reports*, 8(6):571–577, 1988.
- [47] Willard S. S. and Devreotes P. N. Signaling pathways mediating chemotaxis in the social amoeba, dictyostelium discoideum. *European journal of cell biology*, 85(9-10):897–904, 2006.
- [48] Jowhar D. and Janetopoulos C. The chemotactic compass. In *Dictyostelids*, pages 71–87. Springer, 2013.
- [49] Gross J.D., Peacey M.J., and Trevan D.J. Signal emission and signal propagation during early aggregation in dictyostelium discoideum. *Journal of cell science*, 22(3):645–656, 1976.
- [50] Juliet C Coates and Adrian J Harwood. Cell-cell adhesion and signal transduction during dictyostelium development. *Journal of cell science*, 114(24):4349–4358, 2001.
- [51] Jongrae Kim and Declan G Bates. Robustness of oscillations in biological systems. 2009.
- [52] Jongrae Kim, Pat Heslop-Harrison, Ian Postlethwaite, and Declan G Bates. Stochastic noise and synchronisation during dictyostelium aggregation make camp oscillations robust. *PLoS computational biology*, 3(11):e218, 2007.
- [53] Edward M Purcell. Life at low reynolds number. *American journal of physics*, 45(1):3–11, 1977.
- [54] Grignard A., Taillandier P., Gaudou B., Vo D.A., Huynh N.Q., and Drogoul A. Gama 1.6: Advancing the art of complex agent-based modeling and simulation. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 117–131. Springer, 2013.
- [55] McCann C. P., Kriebel P. W., Parent C. A., and Losert W. Cell speed, persistence and information transmission during signal relay and collective migration. *J Cell Sci*, 123(10):1724–1731, 2010.
- [56] Van Haastert P. J.M. and Postma M. Biased random walk by stochastic fluctuations of chemoattractant-receptor interactions at the lower limit of detection. *Biophysical journal*, 93(5):1787–1796, 2007.
- [57] Steven A MacKay. Computer simulation of aggregation in dictyostelium discoideum. *Journal of cell science*, 33(1):1–16, 1978.
- [58] Di Criscio S. A simple mechanism for complex social behavior, 18/12/2011.
- [59] John C Dallon, Brittany Dalton, and Chelsea Malani. Understanding streaming in dictyostelium discoideum: theory versus experiments. *Bulletin of mathematical biology*, 73(7):1603–1626, 2011.
- [60] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826, 2004.
- [61] Jörg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J Doyle III, and John Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, 2004.
- [62] Hiroaki Kitano. Towards a theory of biological robustness. *Molecular systems biology*, 3(1):137, 2007.
- [63] Marten Scheffer. *Critical transitions in nature and society*, volume 16. Princeton University Press, 2009.
- [64] Bhalla U. S. and Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, 1999.

- [65] Khan S., Makkena R., McGeary F., Decker K., Gillis W., and Schmidt C. A multi-agent system for the quantitative simulation of biological networks. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 385–392. ACM, 2003.
- [66] Vlassopoulos N. and Fates N. Clustering behavior of a bio-inspired decentralized aggregation scheme. In *European Conference on Artificial Life-ECAL 2011*, pages 836–837. MIT Press, 2011.
- [67] Sugihara K. and Suzuki I. Distributed motion coordination of multiple mobile robots. In *Intelligent Control, 1990. Proceedings., 5th IEEE International Symposium on*, pages 138–143. IEEE, 1990.
- [68] Dorigo M., Birattari M., Blum C., Clerc M., Stützle T., and Winfield A. *Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings*, volume 5217. Springer, 2008.
- [69] Timmis J. Mitchell P., Hart E. and Dabiri T. N. F. Bio-inspired model of network, information, and computing systems. 2011.