MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **8**, Number **2**, April **2011**

doi:10.3934/mbe.2011.8.263

pp. **263–277**

MODELING AND SIMULATION OF SOME CELL DISPERSION PROBLEMS BY A NONPARAMETRIC METHOD

Christina Surulescu

ICAM, WWU Münster, Einsteinstr. 62
48149 Münster, Germany

Nicolae Surulescu

IMS, WWU Münster, Einsteinstr. 62
48149 Münster, Germany

Abstract. Starting from the classical descriptions of cell motion we propose some ways to enhance the realism of modeling and to account for interesting features like allowing for a random switching between biased and unbiased motion or avoiding a set of obstacles. For this complex behavior of the cell population we propose new models and also provide a way to numerically assess the macroscopic densities of interest upon using a nonparametric estimation technique. Up to our knowledge, this is the only method able to numerically handle the entire complexity of such settings.

1. **Introduction.** Simple biological organisms like bacteria use a random walk strategy to search for favorable signals in their environment; they appropriately biase their movement in order to adapt to such chemical cues. For instance, flagellated bacteria move through a so-called *velocity jump* process, i.e. a sequence of rather straight *runs* interrupted by reorientations allowing to choose a new velocity (*tumbles*). The changes are usually considered to be generated by a Poisson process whose intensity is given by the turning rate of the particle and the new velocity is dictated by some turning kernel providing the probability of the cell leaving the previous velocity regime.

In the context of bacterial movement one would like to understand e.g., the basic mechanisms underlying microbial behavior, since this could provide approaches for prospective innovations in biotechnology. Thus, a quantitative characterization of bacterial migration could be expedient for assessing microbial processes such as nitrogen fixation, infection spread or biofilm formation, e.g., on medical implants or submersed surfaces. Thinking of more evolved cells of particular relevance would be for instance the migration of endothelial cells in the process of wound healing or that of cancer cells penetrating the surrounding tissue. All these issues can be approached by assessing the evolution of the corresponding macroscopic cell density and this makes its characterization such an interesting problem. For this quantity an

2000 *Mathematics Subject Classification.* Primary: 92C17, 82C31, 60H10; Secondary: 60K40, 65C20, 62G07.

Key words and phrases. Cell movement, velocity jump processes, nonparametric density estimation.

The authors would like to thank two anonymous reviewers for their comments contributing to enhance the presentation of this paper. The first author was supported by the Eliteprogramme for Postdocs of the Baden-Württemberg Foundation.

integro-differential transport equation can be deduced, the diffusion limit of which has been shown to be under certain conditions the classical Patlak-Keller-Segel model for chemotaxis [17], [26], [5].

The transport model has been derived for a one-particle distribution function [30], however the use of transport equations for populations is not always unproblematic, as it was discussed e.g. in [15]. In the case of correlated movements of individuals in the population the deduction of transport equations from stochastic processes has still to be rigorously carried out. It has been attempted to handle the transport equation for the velocity jump model of bacterial movement with the aid of the method of moments, however the resulting system was only closed in some special cases where some rather restrictive assumptions on the turning kernel and on the higher order moments were satisfied [6], [12], [17], [25]. The diffusion based models have been polemized, too; bacterial motion normally obeys a velocity jump process and not a Brownian motion and the parameters of the model (like diffusion constants and chemotactic sensitivity) are not directly related to the individual movement pattern of the species [15].

This motivated us to look for other methods allowing for a description of bacterial motion without the aid of (reaction-diffusion-) transport equations. Our nonparametric approach proposed in [31] avoids the use of differential equations for the cell density; instead, we simulate bacteria trajectories which we employ to nonparametrically estimate the cell density upon applying the theory developed in [28]. For instance, we propose in the present work a model for bacterial motion through a heterogeneous medium, an issue which would be very difficult to address in the framework of partial differential equations. For a simplified setting we also assess the performance of the nonparametric method in comparison with solving a corresponding PDE by applying a method of moments.

In the next section we shortly recall the classical models for cell dispersal, along with some comments on the PDE approach. We briefly present in Section 3 the nonparametric method and illustrate it on the models in Section 4, in which we propose some equation-free descriptions of bacterial motion relying on velocity jump processes, along with some enhancements of the Ornstein-Uhlenbeck models for cell dispersal. We also apply the nonparametric method in order to appraise the behavior of the macroscopic cell density predicted by our models. Numerical simulations are performed in Section 5 via a method of moments for the forward Kolmogorov equation which has been shown to approximate under certain assumptions and in the macroscopic limit the corresponding transport partial integro-differential equation for the cell density. The comparison of the results with those obtained via the nonparametric procedure leads to the discussion in Section 6.

2. **Classical approaches to cell dispersal.**

The current models for describing cell dispersal rely on some classes of random processes either giving a geometrical description of the motion or applying stochastic increments to the cell velocity rather than to its position in space [25]. The former build the class of so-called velocity jump (VJ) models, while in the latter the particle velocity obeys a multivariate Ornstein-Uhlenbeck (OU) process.

The PDE approach to modeling cell population density is one of the most frequently used; it aims to deduce and analyze the PDE satisfied by the population

cell density and to use the computational power of well established numerical methods for solving the corresponding equations in order to assess the behavior of the population of interest.

Let us consider a cell population in a $2N$-dimensional phase space ($N = 1, 2, 3$). Further, let $f(t, \mathbf{x}, \mathbf{v})$ be the cell density function depending on $(\mathbf{x}, \mathbf{v})$, where $\mathbf{x} \in \mathbb{R}^N$ is the position of a cell and $\mathbf{v} \in \mathbb{R}^N$ is its velocity. Thus $f(t, \mathbf{x}, \mathbf{v})\, d\mathbf{x}d\mathbf{v}$ is the number of cells at time t with position between $\mathbf{x}$ and $\mathbf{x} + d\mathbf{x}$ and velocity between $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$. Let $V \subset \mathbb{R}^N$ denote the set of velocities. Then the macroscopic density of individuals at the position $\mathbf{x} \in \mathbb{R}^N$ and at the time t is given by

$$n(t, \mathbf{x}) = \int_V f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v}. \tag{1}$$

2.1. **OU type models.** The OU based models can be written in the general form

$$d\tilde{\mathbf{x}}_t = \mathbf{b}(t, \tilde{\mathbf{x}}_t)dt + \boldsymbol{\sigma}(t, \tilde{\mathbf{x}}_t)d\mathbb{B}_t, \ t \geq 0 \tag{2}$$

where $\tilde{\mathbf{x}}_t \in \mathbb{R}^{2N}$ is the multivariate stochastic process with components $\mathbf{x}_t$ and $\mathbf{v}_t$, whereas $\mathbb{B}_t$ is a multivariate Brownian motion in $\mathbb{R}^{2N}$. In this stochastic differential equation the drift $\mathbf{b} \in \mathbb{R}^{2N}$ and the diffusion matrix $\boldsymbol{\sigma} \in \mathbb{R}^{2N\times 2N}$ are most often of the form

$$\mathbf{b} = \begin{pmatrix} \mathbf{v}_t \\ \frac{\chi}{\rho}(\mathbf{m} - \mathbf{v}_t) \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\rho}\boldsymbol{\Sigma} \end{pmatrix}, \quad \text{with} \quad \boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}, \quad \mathbf{m} \in \mathbb{R}^N, \tag{3}$$

where χ/ρ denotes the rate of mean reverting for the velocity process $\mathbf{v}$ and is related to the chemotactic sensitivity. $\mathbf{m}$ and $\boldsymbol{\Sigma}$ are constant matrices or -in more realistic settings- they are themselves stochastic processes or involve some. If $\boldsymbol{\Sigma}$ is constant, then the involved variables are Gaussian distributed, which is often not realistic, see e.g. [7], [34], [35]. However, this problem can be overcome upon allowing $\boldsymbol{\Sigma}$ to be a stochastic process of the OU type comprising several stochastic effects in the biological system. Details will be provided in Section 4.

When the coefficients in (2) are sufficiently smooth it is well known that the evolution of the density is described by the forward Kolmogorov equation (FKE)

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{2N} \frac{\partial}{\partial \tilde{x}_i}(b_i(t, \tilde{\mathbf{x}})f) + \frac{1}{2}\sum_{i,j,k=1}^{2N} \frac{\partial^2}{\partial \tilde{x}_i \partial \tilde{x}_j}(\sigma^{ik}(t, \tilde{\mathbf{x}})\sigma^{jk}(t, \tilde{\mathbf{x}})f), \tag{4}$$

where $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{v})' \in \mathbb{R}^{2N}$, $\mathbf{b} = (b_i)_{1\leq i\leq 2N}$, $\boldsymbol{\sigma} = (\sigma^{ij})_{1\leq i,j\leq 2N}$ with b_i and σ^{ij} deterministic functions. For the case with nonsmooth coefficients as in some of our models (see Section 4) it would be technically quite elaborate to deal with this PDE approach.

2.2. **VJ type models.** In the simplest velocity jump case, i.e. in the absence of cell-cell interactions and when no external stimuli are present, the density f satisfies a Boltzmann like partial integro-differential equation [25] with the integral operator characterizing the turning events:

$$\partial_t f(t, \mathbf{x}, \mathbf{v}) + \mathbf{v} \cdot \nabla f(t, \mathbf{x}, \mathbf{v}) = -\lambda f(t, \mathbf{x}, \mathbf{v}) + \lambda \int_V K(\mathbf{v}, \mathbf{v}')f(t, \mathbf{x}, \mathbf{v}')d\mathbf{v}'. \tag{5}$$

$K(\mathbf{v}, \mathbf{v}')$ denotes the turning kernel and gives the probability of a velocity jump from the $\mathbf{v}'$ to the $\mathbf{v}$ regime. The velocity changes are modelled with a Poisson process with intensity λ and thus the mean running time is $\tau = 1/\lambda$.

A kinetic equation for the phase space cell density of the type of (5) has been introduced by Alt [1] and further studied in [25]. Handling the above integro-differential equations numerically is a very difficult task, especially when more complex turning kernels are involved. Therefore, macroscopic limits have to be deduced, eventually leading to parabolic or hyperbolic PDEs which can be treated more easily. In [5] such a Boltzmann like equation was coupled to a reaction-diffusion one for the external signal and several conditions have been imposed on the turning operator K in order to derive as a scaling limit the classical Patlak-Keller-Segel equations for chemotaxis. Another deduction, based on parabolic scaling, has been given in [17], [26]. However, for a rigorous deduction one needs to impose rather restrictive conditions on the turning kernels. For instance, an unrealistic assumption is

$$\int_V K(\mathbf{v}, \mathbf{v}')d\mathbf{v}' = 1, \tag{6}$$

which is not fulfilled by many reorientation kernels, among others the one based on the von Mises distribution proposed in [6] upon relying on experimental evidence. This is the case, too, for all mixture-based kernels like that to be used for one of our models in Section 4.

There is obviously a need to improve the classical models in order to allow capturing important experimentally observed features. One way of doing this is to consider more flexible turning kernels. However, this leads to a substantial increase in the complexity of the probabilistic settings, rendering the treatment of the resulting models very difficult when trying to use the classical PDE approach. Therefore, alternative methods are in demand, which should enable handling the new settings. The nonparametric approach presented in Section 3 is such a method: it is based on statistical estimation techniques and features a great versatility, since it actually only uses the primary description of the phenomena to be modelled.

3. **Nonparametric estimation of the population density function.**

The kernel density estimation technique is the most widely used method when estimating complex density functions, owing to its flexibility and the plethora of theoretical results establishing its consistency for various rates of convergence [28], [29]. This method uses computing power to allow a very effective handling of complicated structures. When the assumption of Gaussianity for the densities of interest is not appropriate, then a certain type of parametric density might be more suitable for describing the data. But other times it is more desirable to simply *let the data speak for themselves*, i.e. to look for an estimator of the population density, unconstrained (or as loosely as possible) by an apriori form. This is in fact the aim of nonparametric density estimators. In the following we apply this method for independent simulations, however it also works under fairly general conditions for dependent data, see e.g., [27], [36].

Starting from one of the models (velocity jump or OU processes) for cell movement, we simulate S independent bacterial trajectories on the interval of interest $[0, T]$ and we use the data sets obtained in this way to estimate the cell population density at an arbitrary moment of time $t \in [0, T]$. More details on how these simulations are performed are provided at the end of this section.

The nonparametric estimators for the cell population density n at some moment t are defined by (see [28]):

$$\widehat{n_{\mathbb{H}}}(t,\mathbf{x}) = \frac{1}{S\det\mathbb{H}}\sum_{i=1}^{S}\mathcal{K}(\mathbb{H}^{-1}(\mathbf{x}-\mathbf{X}_i)), \qquad \mathbf{x}\in\mathbb{R}^N, \tag{7}$$

or, in an analogous way, for other densities of interest like e.g. for f:

$$\widehat{f_{\tilde{\mathbb{H}}}}(t,\tilde{\mathbf{x}}) = \frac{1}{S\det\tilde{\mathbb{H}}}\sum_{i=1}^{S}\tilde{\mathcal{K}}(\tilde{\mathbb{H}}^{-1}(\tilde{\mathbf{x}}-\tilde{\mathbf{X}}_i)), \qquad \tilde{\mathbf{x}}\in\mathbb{R}^{2N}, \tag{8}$$

where $\mathcal{K}$ and $\tilde{\mathcal{K}}$ denote general kernel functions, $\tilde{\mathbf{x}} = (\mathbf{x},\mathbf{v})'$, $\tilde{\mathbf{X}}_i = (\mathbf{X}_i,\mathbf{V}_i)'$, $\mathbf{X}_i = (X_{i1},\dots,X_{iN})'$, $\mathbf{V}_i = (V_{i1},\dots,V_{iN})'$, $i = 1,\dots,S$ are the position, respectively the velocity at the moment t of the simulated ith trajectory (where the superscript $'$ denotes the transposition of a vector or a matrix) $\mathbb{H}$ and $\tilde{\mathbb{H}}$ are the corresponding bandwidth matrices, which are usually taken to be diagonal and invertible. For the concrete numerical applications in this paper they are taken to be of the form $\mathbb{H} = h\mathbb{I}$, with $\mathbb{I}$ denoting the identity matrix and $h > 0$ being the so-called bandwidth parameter.

One of the most frequently used kernels in the univariate case is the Gaussian kernel defined by $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$, $u\in\mathbb{R}$. Other classical choices are: Epanechnikov and its variants, triangular, rectangular etc [28]. In the multivariate case the easiest form to be chosen for the kernel $\mathcal{K}$ is the multiplicative one: $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^{N} K(x_j)$. Analogously for $\tilde{\mathcal{K}}$.

Thereby, the choice of the bandwidth matrix is important, whereas the choice of the kernel function is not crucial, since it is possible to rescale the kernel function such that the difference between two given density estimators using two different kernel functions is negligible [22].

The choice of the bandwidth matrix is one of the most difficult practical problems in connection with the above method. The bandwidths are chosen according to the available information about the density to be estimated. For example, if it is known that the density to be estimated is very close to a normal one, then the bandwidths can be optimally chosen with the so-called *rule of thumb* [28] giving an explicit expression for the bandwidth matrix. However, this is rarely the case, therefore a more adequate choice is to compute the bandwidth according to one of the data driven bandwidth selection criteria. One of the most popular ones is the *least squares cross validation* (LSCV), which has the goal to estimate the integral squared error (ISE), defined by

$$ISE(\mathbb{H}) = \int[\widehat{n_{\mathbb{H}}}(t,\mathbf{x}) - n(t,\mathbf{x})]^2 d\mathbf{x}, \tag{9}$$

where $\widehat{n_{\mathbb{H}}}(t,\mathbf{x})$ is the estimated density and $n(t,\mathbf{x})$ is the true density being estimated. The usual method for estimating ISE is the *leave-one-out* cross validation. The minimization of the estimated ISE leads to an optimal choice of the bandwidth matrix for a given kernel density function $\mathcal{K}$. It is this bandwidth selection criterion which we will use in the following. Alternatively, there are plenty of other bandwidth selectors in literature (see e.g., [10] and the references therein).

Several results on consistency of the kernel density estimators settling the theoretical foundations of the nonparametric method have been derived e.g. by Cacoullos [3], Deheuvels [8], and Devroye [9]. The issue of convergence speed has

been addressed a.o. by Devroye [9] and we refer for further, more specific convergence results and error estimates to Holmström and Klemelä [19] and the references therein. Similar results for the case with dependent data can be found e.g., in [36]. The computational cost of this method is tightly connected to the so-called curse of dimensionality: the amount of data necessary for an accurate estimation grows exponentially with the spatial dimension of the phenomena of interest. However, for the concrete biological problems handled in this work the spatial dimension is maximum three when estimating the macroscopic cell density and thus the size of the required data sets is a very reasonable one. Thus, the method performs well, which can be seen below upon comparing the kernel density estimation with the true density e.g., for a mixture of normals, as a representative of the class of non-Gaussian densities[1]. Here and for the rest of this paper the implementation has been done in Matlab using the built in Statistical Toolbox.

Experimental results. We assess the performance of the nonparametric method by validation against true analytical densities (in 2D). The following examples are dealing with two densities: a normal one, respectively a mixture of Gaussians.

In the first case we consider the cell population density to be normally distributed $\mathcal{N}(\mathbf{m}, \sigma^2 \cdot \mathbb{I}_2)$, where $\mathbf{m} = (0.9, 0.9)'$ and $\sigma = 2.05$. In the second case we estimate the following mixture of normals:

$$\frac{1}{2}\phi_{\mathcal{N}(\mathbf{m},\sigma^2\cdot\mathbb{I}_2)} + \frac{1}{2}\phi_{\mathcal{N}(-\mathbf{m},\sigma^2\cdot\mathbb{I}_2)}, \tag{10}$$

where $\mathbb{I}_p$ denotes the $p \times p$ identity matrix, $\phi_{\mathcal{N}(\mathbf{m},\Sigma)}$ means the density of the Gaussian distribution $\mathcal{N}(\mathbf{m}, \Sigma)$. We chose for this example $\mathbf{m} = (12, 12)'$ and $\sigma = 4$. In each of these cases we constructed the corresponding nonparametric estimator upon choosing the bandwidth with the aid of LSCV. The results are illustrated in Figure 1 of the Appendix.

In both cases we used up to 3000 simulations (1500 in the first case and 3000 for the second case). It can be seen that this provides accurate estimates of the corresponding densities. In the case with the simple Gaussian the absolute error between this density and the estimated one is at most 0.004, while in the case with the normal mixture the corresponding difference is less than 0.0004. This is due to the higher number of simulations used to estimate the mixture. We refer e.g. to [28], [29] for tables specifying the number of simulations needed w.r.t. the dimension of the problem.

We decided to perform all estimations to follow with the aid of $S = 10000$ simulations, though we have seen above that actually less is needed in order to ensure a good accuracy. The simulations of random variables and stochastic processes involved in our models have been done upon using classical algorithms described e.g., in [2], [20], where one can also find the theoretical results concerning the performance of these algorithms. Where applicable, the choice of parameters is consistent with the literature [25], [6], [12].

4. **Some new classes of models.**

In this section we present some ways to enhance the modeling power of classical settings for cell dispersal and apply the kernel density estimation method in the previous section to the new models to illustrate

[1]It is well known that any density can be arbitrarily well approximated with such normal mixtures. Therefore, the latter are with predilection recommended for testing the methods for estimating densities [23].

the behavior of the respective cell populations. The corresponding figures are postponed to the appendix in order not to complicate the exposition. We use throughout for the estimation a Gaussian type kernel and the bandwidths are computed in each time step with the aid of the least square cross validation (LCSV) method. The simulations of the corresponding SDE in the following Subsection 4.1 have been performed with an Euler-Maruyama scheme (see e.g., [20]) for which a time step $\Delta t = 0.01$ was used.

4.1. **OU type models with stochastic volatility.** One can distinguish between two moving regimes for cells: one influenced by a chemotactic signal (when $\mathbf{m}$ in Section 2.1 above is s.t. $|\mathbf{m}| \neq 0$) and an unbiased random motion when $\mathbf{m} = \mathbf{0}$. The classical OU model with $\mathbf{m}$ constant can only describe one of the behavioral paradigms (e.g., attractant or nonattractant), while obviously in practice both are encountered in a random alternation during the life cycle of a cell population.

The following model takes into account both aspects, upon choosing $\mathbf{m}$ to be a stochastic process of the form $\mathbf{m}_t = \mathbf{g}(t, U_t)$, with an appropriate function $\mathbf{g}$ and a stochastic process U_t, which can be seen e.g., as an individual indicator of the internal dynamics of the cell as a result of the influences of environmental factors like local abundance of nutrients. For example, one can model U_t as another OU process with dynamics independent on the Brownian motion $\mathbf{W}_t$. In a more detailed description it could also be the logarithm of some weighted mean over the outcomes of an intracellular signaling pathway initiated by some input signal. A possible parametrization for $\mathbf{g}$ to be used in the following is $\mathbf{g}(t, U_t) = e^{-\gamma t}\mathbf{1}_{\{U_t \leq 0\}}\mathbf{m}$, where $\gamma \geq 0$ is e.g. the decaying rate for the concentration of a stimulus [2]. Thus, it becomes clear how the stochastic process U_t discussed above influences the individual switch between the two behavioral movement regimes: with, respectively without bias (e.g., chemoattractant). Notice that for $\gamma > 0$ the population asymptotically passes to an unbiased regime.

In order to alleviate the assumption of Gaussianity for the involved random variables we choose $\mathbf{\Sigma}$ to be stochastic, e.g., of the form $\psi(t, U_t)\tilde{\mathbf{\Sigma}}$, with $\tilde{\mathbf{\Sigma}}$ a constant matrix and ψ an appropriate real function. For instance, one could make the choice $\psi(t, U_t) = ae^{-\gamma t} + (1 - ce^{-\gamma t})be^{U_t}$ $(a, b \geq 0,\ a^2 + b^2 > 0,\ 0 \leq c \leq 1)$, where $\gamma \geq 0$ and U_t have the same significance as above.

Introducing in this way the dependence on U_t in $\mathbf{\Sigma}$ also has the role to suggest the differentiation between biased and unbiased also at the level of velocity variance, which in the biased regime usually has to be smaller than in the unbiased one. The above considerations then lead to the following class of models:

$$d\mathbf{x}_t = \mathbf{v}_t\, dt \tag{11}$$

$$\rho d\mathbf{v}_t = \chi(\mathbf{g}(t, U_t) - \mathbf{v}_t)\, dt + \psi(t, U_t)\tilde{\mathbf{\Sigma}}\, d\mathbf{W}_t, \qquad t \geq 0 \tag{12}$$

$$dU_t = \alpha_U(m_U - U_t)\, dt + \beta_U\, dZ_t, \tag{13}$$

with $\mathbf{W}_t$ and Z_t independent Brownian motions, $\mathbf{g} : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}^N$, $\psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ some given functions, $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{v}(0) = \mathbf{v}_0$ and $U(0) = U_0$, with $\mathbf{x}_0$, $\mathbf{v}_0$, U_0 random variables independent of $\mathbf{W}_t$ and Z_t and with some given distributions. The meaning of the involved constants is the same as in Section 2.1, ρ is a measure of persistence, $\alpha_U, \beta_U > 0$, $m_U \in \mathbb{R}$.

[2] $\mathbf{1}_{[.]}$ denotes the characteristic function

The above choice of functions $\mathbf{g}$ and ψ is particularly suited to capture more features of the inter- and intracellular environments. This also improves the distributional properties of the models, which would better reflect the empirical observations.

For the simulation to follow we consider the above model to illustrate both the case with unbiased motions, as well as the one with a stochastic switcher between the attractant and nonattractant regimes. Since there are no experimental data available, we performed our simulations upon either taking the corresponding parameters from literature ([12], [25], [24]) or by choosing them inside biologically relevant ranges.

The parameters $\rho = 1$, $\chi = 11$ are common for both cases. For the unbiased case we take $\mathbf{g} \equiv 0$ and for the case involving the stochastic switcher we choose $\alpha_U = 4$, $\beta_U = 2.4$, $m_U = -0.15$, now with $\mathbf{g} = \mathbf{1}_{\{U_t \leq 0\}} \cdot k_m \cdot (1,1)'$, $\psi(t, U_t) = 4.5 + \exp(U_t)$, with the constant $k_m = 0.09$. The initial distributions of the involved processes are all of Gaussian type: for the process describing the cell position $\mathcal{N}(\mathbf{0}, \sigma_p^2 \cdot \mathbb{I}_2)$ with $\sigma_p = 0.03$, for the velocity $\mathcal{N}(k_m \cdot (1,1)', 0.02 \cdot \mathbb{I}_2)$, and for the stochastic switcher between the attractant and nonattractant regimes we take $\mathcal{N}(-0.15, 5.06)$.

Typical trajectories and the behavior of the stochastic switcher U_t are shown in Figure 2, while Figure 3 illustrates the initial population density[3] and the corresponding estimated densities at $t = 6$. Notice in the case of switch the displacement of the population mean, along with a faster spread.

4.2. **VJ type model preserving the features of an OU model with stochastic volatility.** In the previous Subsection 4.1 we have seen the power and versatility of modeling with OU processes, due to their nice mathematical properties. It would be interesting to translate these advantages to the VJ modeling framework. The new model will have similar properties with those in the previous Subsection 4.1 and has the advantage of the turning kernel being implicitly specified.

One natural idea would be to change equation (11) in the description of the OU model and put it e.g., in the form

$$d\mathbf{x}_t = \mathbf{v}_{\zeta(t, U_t, N_t)}\, dt, \quad t \geq 0, \tag{14}$$

where N_t denotes a Poisson process with intensity $\lambda > 0$ and ζ is a function depending on time and on the processes U_t and N_t, such that the resulting process on the right hand side of equation (14) has piecewise constant trajectories. For instance, we can choose for the subscript in (14)

$$\zeta(t, U_t, N_t) = \frac{1 + c_2^\zeta N_t + [c_3^\zeta t + c_4^\zeta U_t^2]}{c_1^\zeta}, \quad t \geq 0, \tag{15}$$

where $[\cdot]$ denotes the integer part and c_k^ζ, $k = 1, 2, 3, 4$ are some positive constants. The $\mathbf{v}_t$ and U_t processes are described and interpreted as before with the aid of equations (12),(13). The Poisson process N_t can be correlated to U_t and/or $\mathbf{v}_t$ (e.g., by letting λ depend on the velocity norm or on the norm of the process $\mathbf{g}_t$), however for the sake of simplicity we chose it to be independent of all other processes.

[3] which (unless otherwise specified) is the same for all simulation scenarios in this paper

4.3. **A VJ model for bacteria moving in a heterogeneous medium.** We consider a VJ model for a cell population which has to avoid $J \in \mathbb{N}^*$ hostile circular regions centered at $\mathbf{x}_k^O \neq \mathbf{m}_0$, with $k = 1, \dots, J$ and $\mathbf{m}_0$ denoting the mean of the initial macroscopic population density. In order to describe such a complex behavior we need to consider the turning kernel K depending not only on the velocity $\mathbf{v}'$, but also on the current position $\mathbf{x}$ at the moment of velocity change. One possible adequate choice is to describe such a kernel for a given position $\mathbf{x}$ and previous velocity $\mathbf{v}'$ to be the density of a random variable $\mathbf{V} = \Big(\sum_{j=1}^{J} w_j(\mathbf{x})\Big)^{-1} \sum_{j=1}^{J} w_j(\mathbf{x})\hat{\mathbf{V}}_j$ where $\hat{\mathbf{V}}_j$ are random variables with a density $K_j(\cdot, \mathbf{v}', \mathbf{x})$, $j = 1, \dots, J$, with

$$K_j(\mathbf{v}, \mathbf{v}', \mathbf{x}) = p_j(\mathbf{x}) \cdot K_{1j}(\mathbf{v}, \mathbf{v}') + q_j(\mathbf{x}) \cdot K_{2j}(\mathbf{v}, \mathbf{v}', \mathbf{x}), \tag{16}$$

$$p_j(\mathbf{x}) = \Big[1 - \exp(-\frac{1}{\sigma_j^O}\Big|\|\mathbf{x} - \mathbf{x}_j^O\|^2 - (r_j + d_j)^2\Big|)\Big] \cdot 1_{\mathbb{R}^n - \bar{\mathcal{B}}(\mathbf{x}_j^O, (r_j + d_j))}(\mathbf{x}), \tag{17}$$

$$q_j(\mathbf{x}) = 1 - p_j(\mathbf{x}), \quad r_j, d_j > 0, \quad w_j(\mathbf{x}) = \exp(\xi_j q_j(\mathbf{x})), \quad \xi_j > 0, \quad \mathbf{x} \in \mathbb{R}^N, \tag{18}$$

where r_j is the radius of the obstacle centred in $\mathbf{x}_j^O$, σ_j^O and d_j are some positive parameters allowing for a versatile control of the aversion with respect to the avoidance region. Large values of these parameters will force the cell to remain away from the obstacle. In relation (17) above, $\bar{\mathcal{B}}(\mathbf{x}_j^O, (r_j + d_j))$ is the closed ball centered in $\mathbf{x}_j^O$ and with radius $r_j + d_j$ and $\mathbf{1}_{[\cdot]}$ denotes as before the characteristic function. The positive space dependent weights $\{w_j\}_j$ quantify the influence of the obstacles on the cell motion: the nearer an obstacle is to the cell, the greater becomes its influence. The role of the parameters ξ_j is to provide a more flexible control of this influence.

The above turning kernels $\{K_j\}_j$ are mixtures with position depending weights $\{p_j(\mathbf{x}), q_j(\mathbf{x})\}$ of two types of kernels $\{K_{1j}\}_j$ and $\{K_{2j}\}_j$, the first one modeling a usual VJ movement and the other one being specialized in describing the motion around the avoidance region. According to the distance to the obstacles, the previous system of mixture weights has the role to induce switching between a motion type dictated by $\{K_{1j}\}_j$ or one following $\{K_{2j}\}_j$. For instance, the kernels in the former family can be some Gaussians with mean $\chi_{1j}(p_1^{K_{1j}} \frac{\mathbf{x}_j^O - \mathbf{m}_O}{\|\mathbf{x}_j^O - \mathbf{m}_O\|} + q_1^{K_{1j}} \frac{\mathbf{v}'}{\|\mathbf{v}'\|})$, $p_1^{K_{1j}} + q_1^{K_{1j}} = 1$ [4] [5] and given covariance Σ_{1j}, and where χ_{1j} are rescaling constants such that the mean speed is in some biologically reasonable interval (for instance 10-20 μm/s, see [11]). For the second kernel family $\{K_{2j}\}_j$, a possible choice is to take them as the corresponding densities of the random variables $\{\mathbf{V}_{K_{2j}}\}_{j \in J}$ given e.g., by $\mathbf{V}_{K_{2j}} = \text{sign}(< \frac{\mathbf{x} - \mathbf{x}_j^O}{\|\mathbf{x} - \mathbf{x}_j^O\|}, \tilde{\mathbf{V}}_j >)\tilde{\mathbf{V}}_j$, where [6] $\tilde{\mathbf{V}}_j$ is a random vector having any type of density which is suitable to describe a classical bacterial motion. For instance, it can be chosen to be a Gaussian vector with some variance Σ_{2j} and a mean $\chi_{2j}(p_1^{K_{2j}} \frac{\mathbf{x}_j^O - \mathbf{x}}{\|\mathbf{x}_j^O - \mathbf{x}\|} + q_1^{K_{2j}} \frac{-\mathbf{v}'}{\|\mathbf{v}'\|})$, $p_1^{K_{2j}} + q_1^{K_{2j}} = 1$, where χ_{2j} has a similar rescaling role as χ_{1j} and the minus in front of the velocity $\mathbf{v}'$ was chosen in order to allow for drastic direction changes. In this model, too, the velocity changes are modeled with a Poisson process with intensity λ.

[4] Choosing $p_1^{K_{1j}} > q_1^{K_{1j}}$ will induce a biased movement towards the obstacle j.
[5] we make the convention $\tilde{\mathbf{v}}/\|\tilde{\mathbf{v}}\| = \mathbf{0}$, whenever $\|\tilde{\mathbf{v}}\| = 0$ and for any $\tilde{\mathbf{v}} \in \mathbb{R}^N$
[6] $< \cdot, \cdot >$ denotes the scalar product in $\mathbb{R}^N$

A cell trajectory simulated with the aid of this model can be found in Figure 5a, while the estimated population density at several time moments is shown in the rest of Figure 5 and in Figure 6. One can see how the cells circumvent the unexpedient areas. Since we considered a situation where the movement was biased, the cell population will recollect after passing the avoidance regions and will migrate further as a cohesive entity in the direction of its bias. The simulations have been done with the following set of parameters: $J = 4$, $\mathbf{x}_1^O = (1.5, 4.6)'$, $\mathbf{x}_2^O = (3, 1.9)'$, $\mathbf{x}_3^O = (5, 3.6)'$, $\mathbf{x}_4^O = (1, 0.5)'$, $r_1 = 0.7$, $r_2 = 0.5$, $r_3 = 0.2$, $r_4 = 0.3$, $d_j = 0.3$, $\Sigma_{1j} = \Sigma_{2j} = 0.1 \cdot \mathbb{I}_2$, $\xi_j = 5$, $p_1^{K_{1j}} = 0.7$, $p_1^{K_{2j}} = 0.5$, $\chi_{1j} = \chi_{2j} = 15$ for all $j = 1, \ldots, J$, $\sigma_j^O = 2$ for $j = 1, 3$ and $\sigma_j^O = 3.2$ for $j = 2, 4$, $\mathbf{m}_O = (0, 0)'$, $\lambda = 2$. Following the same ideas, new adequate kernels can be constructed for further interesting aspects like finding food sources or tracking of a chemoattractant [32]. Moreover, this allows us to obtain a large and flexible class of models, for which the majority of the corresponding kernels no longer satisfy the rather unrealistic condition (6) imposed in [17], [26].

5. **The nonparametric approach as a numerical method for PDEs.** In this section we intend to offer a new perspective to the nonparametric approach and interpret it as a numerical method for solving the integro-differential PDEs of the type presented in Section 2. Thus, we consider in the following an example of PDE for the OU type of cell motion, which we solve numerically both with an appropriate classical method and with the nonparametric procedure, then illustrate graphically the difference between these results. It is quite difficult to directly solve numerically the Boltzmann type equation in the VJ case, however it also reduces under suitable assumptions and in the macroscopic limit (with an adequate scaling) to a Fokker-Planck equation. We refer to [4] for further details. Since we solve for the OU case an FKE we shall not do it again for the VJ case.

Let us consider the forward Kolmogorov equation (4) corresponding to the OU system (2), with the following parameters: $\rho = 1$, $\chi = 11$, $\mathbf{m} = 0.09 \cdot (1, 1)'$, and $\mathbf{\Sigma} = \mathrm{diag}(3.2, 4.5)$. In order to solve this PDE on an unbounded domain with a standard procedure we use the method of moments, since unlike other classical numerical methods it does not require to specify a bounded domain with the afferent conditions on the boundary. The method has already been applied e.g., in [14] in a fairly general context. With this technique finding the numerical solution of a PDE reduces to solving a system of ODEs, which can be done with standard Matlab procedures. However, it is well known that this method is only reliable for simple models and fails for more complex settings.

Here we translated the method of moments to our particular FKE case, while reducing the system upon using the fact that we deal here with normal population densities and accordingly expressing the higher order moments with respect to those up to the second order, from which the macroscopic densities of interest can be easily recovered. The errors between the macroscopic densities obtained with this method and with the nonparametric estimation are illustrated for the time moments $t = 0.2$, $t = 0.4$ and $t = 0.6$ in Figure 4. Notice that the maximum absolute error is for all time moments less than 0.009. Thus, the nonparametric method also behaves well as a numerical procedure for solving PDEs. Clearly, it can only be applied to the (rather large) class of PDEs emerging from a stochastic framework which enables one to perform simulations of the involved processes. More precisely, if the solution f (or any of its bijective transformations) of a PDE is a density of

some stochastic process X_t whose trajectories can be numerically simulated, then the above nonparametric technique can be used to numerically solve that PDE. If this is the case, then the advantages of this method become clear: unlikely most of the classical methods, it does not need the whole meshing instrumentary and does also not require imposing conditions on some artificial domain boundaries. This is particularly interesting when the shape of the domain is unknown or the domain is unbounded, as is both usually faced when handling biological problems. Moreover, the algorithm is relatively easy to implement and, as already mentioned in Section 3, there are well established results about its convergence properties.

6. **Conclusions and comments.** We proposed a class of new models for cellular dispersal which are able to enhance the description of relevant biological phenomena without complicating too much the corresponding settings.

Since the PDE approach for these models would lead to considerable technical difficulties we proposed an alternative way based on the kernel density estimation technique, which is very efficient for the problems handled in this paper, due to their small spatial dimension. Moreover, it works under less restrictive assumptions and it also allows handling some critical cases where the PDE approach becomes questionable. The great flexibility of this method relies on the fact that one only needs a model for cell movement from which trajectory simulations can be performed, so it allows to place the focus on modeling the cell movement rather than forcing intuitive biological facts into the compulsion of a PDE model for which a lot of assumptions have to be satisfied.

We gave here only a small selection of the features which can be modelled and handled with the aid of this method. For a larger variety of models, e.g., allowing for resting phases, tracking of a food source or a chemoattractant we refer to [31], [32].

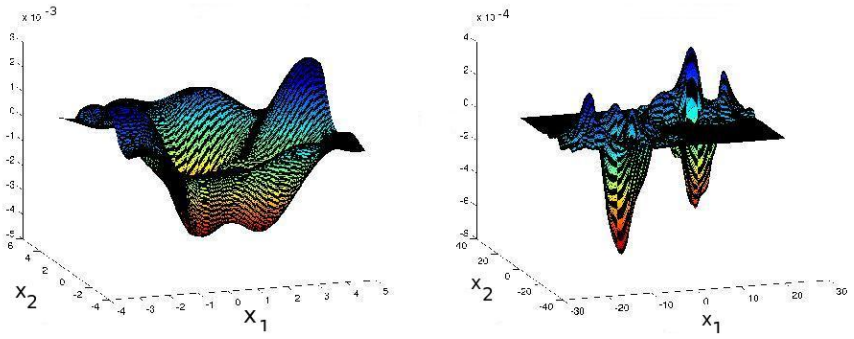
The examples presented in this work (an OU type model with an individual switcher between attractant and nonattractant, VJ type model for a population avoiding some hostile regions or a velocity jump movement described with the aid of an OU type model) seem to be numerically infeasible in the framework of a PDE approach, if such a PDE can be written at all. This clearly shows the advantage of using the nonparametric approach for a simple and direct treatment of many realistic cases. Most numerical methods for the PDEs are only able to solve the macroscopic limits, which however are known to provide a rather pale reflection of the actual biological problem.

Our approach provides an alternative to equation free numerical techniques like e.g. gap tooth methods (see e.g. [13]) and it seems to be more effective, since it needs a considerably smaller number of simulations to recover the dynamics of the cell population density. Thus, the gap tooth method applied in e.g. [11] needed 10^6 simulations, whereas the above nonparametric method led to very good results with up to 10000 simulations, as could be seen in Sections 3 and 5. This is mainly due to the good asymptotic properties of the nonparametric estimator and, again, to the low phase space dimension. Moreover, the technique can also be used as a numerical method for solving a large class of PDEs derived from a stochastic framework. This is particularly interesting -among others- for bacterial motion in the VJ case, where the corresponding PDE is a Boltzmann-like equation which -when coupled to the chemoattractant equation- has been handled so far only via passage to the macroscopic limits.

The method can be applied to any model from which one can simulate trajectories, thus to most models encountered in literature, including the multiscale ones involving internal dynamics and the influence of a chemoattractant, the latter being a topic which has recently attracted increasingly interest also in the context of tumor cell migration through tissue networks [18]. The use of the kernel density estimation method for a multiscale model involving intracellular dynamics and mixture based turning kernels explicitly depending on the chemotactic signal has been handled in [33].

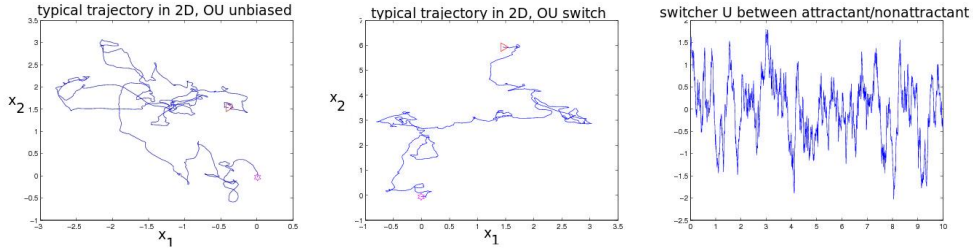
In the present work the nonparametric approach was applied in the context of independent cell trajectories. However, it can also be employed under fairly general conditions even in the case of dependent data (see e.g., [27], [36] for the mathematical framework), which opens the possibility of using this method also for self-organization models. This is ongoing work.

Appendix: Simulation and estimation results



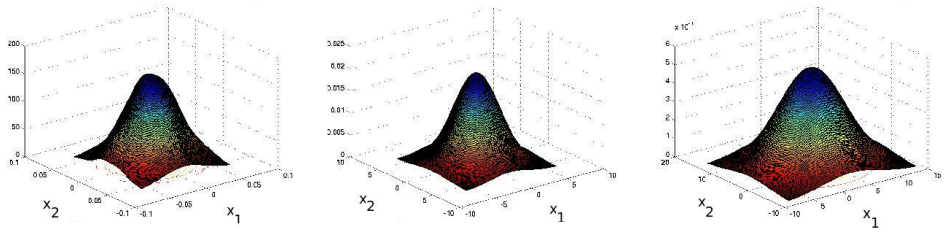
(A) (B)

FIGURE 1. Plot of differences between real and estimated densities for the experiment in Section 3 with the simple normal distribution (1a) and with 1500 simulations, respectively with the normal mixtures (1b) and with 3000 simulations.



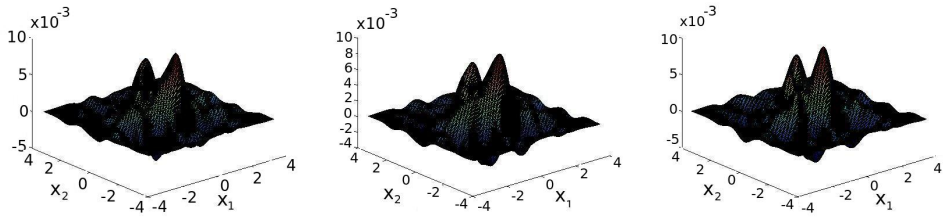
(A) Unbiased motion (B) Use of U_t (C) The process U_t

FIGURE 2. Simulation scenario 4.1. (2a),(2b): typical trajectories in the OU case. Start $*$, end $\triangleright$.



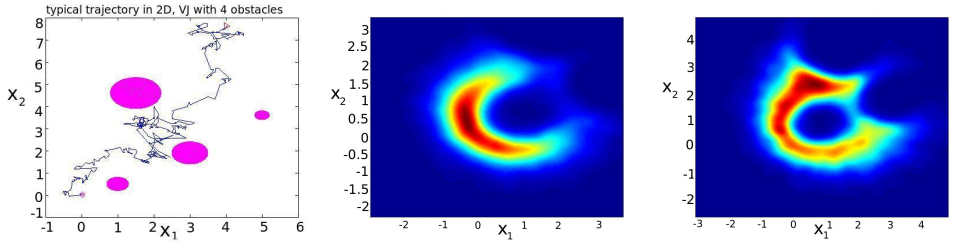
(A) Initial cell density (B) Unbiased (C) Use of U_t

FIGURE 3. Simulation scenario 4.1. Macroscopic cell density (initial and estimated at $t = 6$).



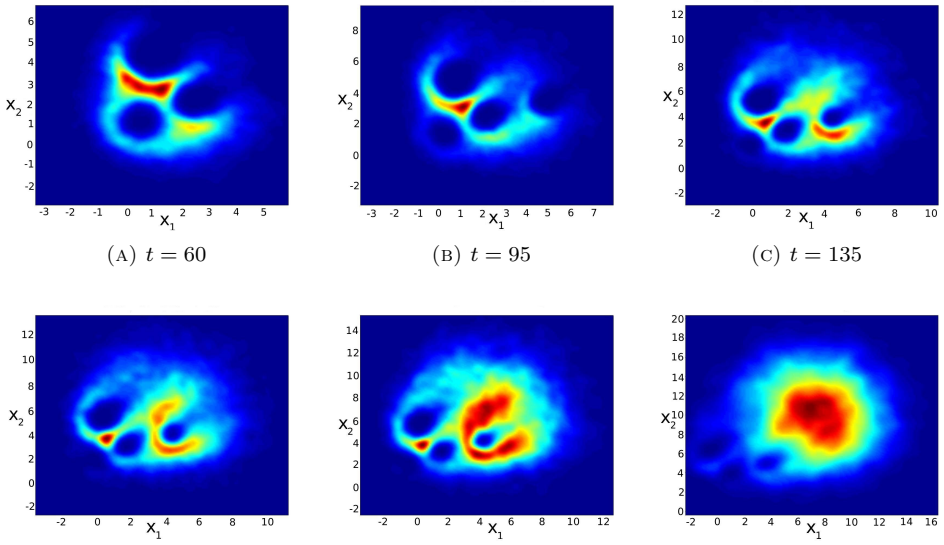
(A) $t = 0.2$ (B) $t = 0.4$ (C) $t = 0.6$

FIGURE 4. Difference between the macroscopic density obtained by using the method of moments and its nonparametric estimation.



(A) Typical bacterium trajectory (B) $t = 25$ (C) $t = 40$

FIGURE 5. Simulation scenario 4.3: estimated macroscopic density at several time moments.



(A) $t = 60$ (B) $t = 95$ (C) $t = 135$

(D) $t = 150$ (E) $t = 170$ (F) $t = 245$

FIGURE 6. Simulation scenario 4.3: avoiding 4 circular regions with different radii (continued from Figure 5).

REFERENCES

[1] W. Alt, *Biased random walk models for chemotaxis and related diffusion approximations*, Journal of Mathematical Biology, **9** (1980), 147–177.

[2] S. Asmussen and P. W. Glynn, "Stochastic Simulation. Algorithms and Analysis," Springer, 2007.

[3] T. Cacoullos, *Estimation of a multivariate density*, Annals of the Institute of Statistical Mathematics, **18** (1966), 179–189.

[4] F. A. C. C. Chalub, Y. Dolak-Struss, P. Markowich, D. Oelz, C. Schmeiser and A. Soreff, *Model hierarchies for cell aggregation by chemotaxis*, Mathematical Models and Methods in the Applied Sciences, **16** (2006), 1173–1197.

[5] F. A. C. C. Chalub, P. Markowich, B. Perthame and C. Schmeiser, *Kinetic models for chemotaxis and their drift-diffusion limits*, Monatshefte für Mathematik, **142** (2004), 123–141.

[6] E. A. Codling and N. A. Hill, *Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters*, Journal of Mathematical Biology, **51** (2005), 527–556.

[7] A. Czirók, K. Schlett, E. Madarász and T. Vicsek, *Exponential distribution of locomotion activity in cell cultures*, Physical Review Letters, **81** (1998), 3038–3041.

[8] P. Deheuvels, *Estimation non paramétrique de la densité par histogrames généralisés (II)*, Publications de l'Institut Statistique de l'Université de Paris, **22** (1977), 1–23.

[9] L. Devroye and L. Györfi, "Nonparametric Density Estimation: The L_1 View," John Wiley, New York 1985.

[10] L. Devroye, *Universal smoothing factor selection in density estimation: Theory and practice*, Test, **6** (1997), 223–320.

[11] R. Erban and H. G. Othmer, *From signal transduction to spatial pattern formation in E. Coli: A paradigm for multiscale modeling in biology*, Multiscale Modeling and Simulation, **3** (2005), 362–394.

[12] F. Filbet, P. Laurençot and B. Perthame, *Derivation of hyperbolic models for chemosensitive movement*, Journal of Mathematical Biology, **50** (2005), 189–207.

[13] C. W. Gear, J. Li and I. G. Kevrekidis, *The gap-tooth method in particle simulations*, Physics Letters A, **316** (2003), 190–195.

[14] T. Hillen, "Transport Equations and Chemosensitive Movement," Habilitation Thesis, University of Tübingen, 2001.
[15] T. Hillen, *Hyperbolic models for chemosensitive movement*, Mathematical Models and Methods in the Applied Sciences, **12** (2002), 1–28.
[16] T. Hillen, *Transport equations with resting phases*, European Journal of Applied Mathematics, **14** (2003), 613–636.
[17] T. Hillen and H. G. Othmer, *The diffusion limit of transport equations derived from velocity-jump processes*, SIAM Journal of Applied Mathematics, **61** (2000), 751–775.
[18] J. Kelkel and C. Surulescu, *A multiscale approach to cell migration in tissue networks*, preprint IANS, University of Stuttgart, 2010, submitted.
[19] L. Holmström and J. Klemelä, *Asymptotic bounds for the expected* L^1 *error of a multivariate kernel density estimator*, Journal of Multivariate Analysis, **42** (1992), 245–266.
[20] P. E. Kloeden and E. Platen, "Numerical Solution of Stochastic Differential Equations," Springer, 2000.
[21] K. V. Mardia, P. E. Jupp, "Directional Statistics," Wiley, 2000.
[22] J. S. Marron and D. Nolan, *Canonical kernels for density estimation*, Statistics and Probability Letters, **7** (1988), 195–199.
[23] J. S. Marron and M. P. Wand, *Exact mean integrated squared error*, Annals of Statistics, **20** (1992), 712–736.
[24] D. Ölz, C. Schmeiser and A. Soreff, *Multistep navigation of leukocytes: A stochastic model with memory effects*, preprint, TU Vienna, 2004.
[25] H. G. Othmer, S. R. Dunbar and W. Alt, *Models of dispersal in biological systems*, Journal of Mathematical Biology, **26** (1988), 263–298.
[26] H. G. Othmer and T. Hillen, *The diffusion limit of transport equations II: Chemotaxis equations*, SIAM Journal of Applied Mathematics, **62** (2002), 1222–1250.
[27] A. R. Pagan and A. Ullah, "Nonparametric Econometrics," Cambridge University Press, 1999.
[28] D. W. Scott, "Multivariate Density Estimation: Theory, Practice and Visualization," John Wiley & Sons, 1992.
[29] B. W. Silverman, "Density Estimation for Statistics and Data Analysis," Chapman & Hall, 1986.
[30] D. W. Stroock, *Some stochastic processes which arise from a model of the motion of a bacterium*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **28** (1974), 305–315.
[31] C. Surulescu and N. Surulescu, *A nonparametric approach to cell dispersal*, preprint IANS 14/2007, University of Stuttgart.
[32] C. Surulescu and N. Surulescu, *A nonparametric approach to cell dispersal*, International Journal of Biomathematics and Biostatistics, **1** (2010), 109–128.
[33] C. Surulescu and N. Surulescu, *On two approahes to a multiscale system modeling bacterial chemotaxis*, preprint IANS, University of Stuttgart, 2/2010.
[34] H. Takagi, M. J. Sato, T. Yanagida and M. Ueda, *Functional analysis of spontaneous cell movement under different physiological conditions*, PLoS ONE, **3** (2008), e2648.
[35] A. Upadhyaya, J. P. Rieu, J. A. Glazier and Y. Sawada, *Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates*, Physica A: Statistical Mechanics and its Applications, **293** (2001), 549–558.
[36] P. Vieu, *Quadratic errors for nonparametric estimates under dependence*, Journal of Multivariate Analysis, **39** (1991), 324–347.

Received February 22, 2010; Accepted December 3, 2010.

E-mail address: christina.surulescu@uni-muenster.de
E-mail address: nicolae.surulescu@uni-muenster.de