Detection and characterisation of chemotaxis without cell tracking

Jack D. Hywood[1]*, Gregory Rice[2], Sophie V. Pageon [3], Mark N. Read[4], Maté Biro[3].

1 Sydney Medical School, The University of Sydney, Sydney, Australia
2 Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada
3 EMBL Australia, Single Molecule Science node, School of Medical Sciences, University of New South Wales, Sydney, Australia
4 School of Computer Science & Charles Perkins Centre, University of Sydney, Sydney, Australia

* jhyw1620@uni.sydney.edu.au

Abstract

Swarming has been observed in various biological systems from collective animal movements to immune cells. In the cellular context, swarming is driven by the secretion of chemotactic factors. Despite the critical role of chemotactic swarming few methods to robustly identify and quantify this phenomenon exist. Here we present a novel method for the analysis of time series of positional data generated from realisations of agent-based processes. We convert the positional data for each individual time point to a function measuring agent aggregation around a given area of interest, hence generating a functional time series. The functional time series, and a more easily visualised *swarming metric* of agent aggregation derived from these functions, provide useful information regarding the evolution of the underlying process over time. We extend our method to build upon the modelling of collective motility using drift-diffusion partial differential equations (PDEs). Using a functional linear model we are able to use the functional time series to estimate the drift and diffusivity terms associated with the underlying PDE. By producing an accurate estimate for the drift coefficient, we can infer the strength and range of attraction or repulsion exerted on agents, as in chemotaxis. Our approach relies solely on using agent positional data. The spatial distribution of diffusing chemokines is not required, nor do individual agents need to be tracked over time. We demonstrate our approach using random walk simulations of chemotaxis and experiments investigating cytotoxic T cells interacting with tumouroids.

1 Introduction

Chemotaxis refers to directed cell movement driven by individual cellular responses to a biochemical gradient. Chemotaxis is a critical driver of an array of physiological processes; it plays a crucial role in embryonic development, directing migratory behaviour of cells in growing tissues [1], formation of new capillaries to sites of ischaemia via angiogenesis [2], and orchestrates cell dynamics in wound healing [3, 4]. In addition, the self-amplifying clustering of immune cells driven by chemotaxis, which we term chemotactic swarming, has been observed during responses to injury where neutrophils swarm to sites of inflammation and recently in cytotoxic T cells interacting with tumour cells [5–7]. These concepts are emerging as highly relevant in the engineering field of swarm robotics, where the spatial coordination of a collective of autonomous vehicles is concerned.

Developments in live-cell imaging techniques increasingly allow for precise localisation of individual cells over time [8–11]. As such, a significant amount of imaging data capturing cell motility is becoming available. Probing this data to detect the presence of chemotactic swarming presents a unique challenge, due to the fact that visualising chemoattractant gradients in conjunction with cell movements is very difficult, particularly without disturbing the underlying dynamics. Whilst individual cell tracking can be used to analyse chemotaxis, tracking individual cells requires a high sampling rate (with associated photo-toxicity), a smaller field of view than would otherwise be available, a low cell density so that consecutive positions of cells are non-ambiguous, and imaging capacity that is capable of imaging in 3D if the cells are not confined to a 2D substrate [10, 12–14]. New statistical approaches need to be developed to detect and characterise chemotactic swarming using cell position data alone.

In this paper we set out a method for detecting and characterising the chemotaxis of agents towards a known area of interest, e.g. a tumour. We first present a general framework for the analysis of aggregation of agents for any type of agent-based processes, independent of the underlying rules governing individual agent movements. To do this, we adapt existing spatial statistical approaches to analyse the positional data generated by an agent-based process. Our methodology involves adapting Ripley's K function from spatial statistics, [15–18], to analyse the distribution of agents around the given area of interest. We refer to these adapted functions as *focal K functions*. Similarly to the analysis of static spatial data, these focal K functions can be used to analyse the spatial aggregation of agents around an area of interest. We then use these functions to produce a scalar metric of agent aggregation, the time series of which provides an easily visualised measure of agent aggregation over time.

We next provide a statistical method for the analysis of the drift and diffusivity governing agent movement under the assumption that the evolution of the agent population is well approximated by the Fokker-Planck equation. The Fokker-Planck equation is an advection-diffusion type partial differential equation (PDE) that allows for spatial and temporal dependence with respect to drift and diffusivity terms [19–21]. Our statistical approach utilises a novel combination of spatial statistics and functional data analysis. The empirical estimates for focal K functions generated from an agent-based process form a functional time series. By modelling the evolution of focal K function using functional linear models we are able to estimate the drift and diffusivity of the approximating Fokker-Planck equation. Our approach is generalised to deal with both time-invariant and time-dependent drift and diffusivity. Our approach, using a simple functional linear model and providing the capacity to produce estimates for drift and diffusion dependent on both space and time, is in contrast to other techniques for fitting microscopic dynamics to PDEs [22–30].

The Fokker-Planck equation is commonly used to model collective cell movement and chemotaxis. Keller-Segel models of chemotaxis, popular and well studied continuum models for chemotaxis, utilise coupled Fokker-Planck type equations for cell position and chemokine concentration [21, 31–33]. Developing techniques to fit Keller-Segel models to experimental data is an active area of research [34]. Fokker-Planck type equations have been shown to provide close approximations to the evolution of cell density among agent-based models of cell movement [35–44].

We develop our methodology for motile agents in two dimensions, but we expect that our methodology can be easily extended to 1D and 3D processes. Our formulation of this approach for 2D analysis is motivated by T cell-tumour experiments detailed in a later section.

We provide an open source implementation of our tool in R [45], along with code for producing and analysing the presented simulations and experiments, available at https://github.com/JackHywood/Chemotacticswarming.

2 Agent-based processes with clustering

We consider systems of agents (e.g. cells, animals, robots) that change their position over time and analyse their arrangements within some observation region, denoted by R. With a particular focus on chemotactic swarming we are interested in establishing whether agents are attracted or repulsed from some central region of interest, B, which lies in R. For any given time t, the realisation of the dynamic agent-based process results in a spatial point pattern within R. Let $\boldsymbol{X}(t)$ be the spatial co-ordinate data for agents at a given time t, such that $\boldsymbol{X}(t) = \{x_1, x_2, ..., x_{n(t)}\}$, with x_j being the spatial co-ordinates of the j^{th} agent, and $n(t)$ being the total number of agents observed within the observation region R at time t. Likewise, let the time points at which observations of an agent-based process occur be given as $\boldsymbol{T} = \{t_1, t_2, ..., t_N\}$, with N giving the total number of time points observed.

For simplicity we here consider the case where the number of agents within R does not change with time, i.e. $n(t) = n$ is constant. In particular, agents cannot cross the boundary of R. We consider cases in which $n(t)$ varies with time in later sections.

We continue by considering the following idealised spatial geometry for such agent-based processes. For simplicity we continue by working in 2D with the assumption that R is circular with radius r_R, and also that B is a circle with radius r_B, with B occurring at the centre of R. We also assume that the distribution of agents in R is isotropic, i.e. that agents are distributed uniformly with respect to direction from the centre of R. We note that this framework allows for the case in which B is a point, i.e. setting $r_B = 0$.

Using polar coordinates and assuming distribution of agents within R is isotropic, we define the density of agents as

$$\lambda(r,t) = \lim_{|ds|\to 0} \frac{\mathbb{E}[Y(ds,t)]}{|ds|}, \qquad 0 \le r \le r_R \tag{1}$$

with ds defining a small annulus with inner radius r from the centre of R, $|ds|$ giving the volume of ds, $Y(ds,t)$ giving the number of agents within ds at time t, and $\mathbb{E}[\cdot]$ denoting expectation with respect to the distribution of a given process evolving stochastically from a given set of initial conditions.

Using a common continuum model for cell movement [20,21,31–34], we assume that the evolution of $\lambda(r,t)$ is governed by the Fokker-Planck equation,

$$\frac{\partial \lambda}{\partial t} = -\nabla \cdot (\mathbf{f}\lambda) + \nabla \cdot (\mathbf{D}\nabla\lambda). \tag{2}$$

It is well established that under certain conditions the evolution of agent-density for discrete agent-based models can be well approximated by PDEs [20,40,43,46].

Working in polar coordinates, and using the assumption of isotropy, we have

$$\frac{\partial \lambda}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\Big(r\Big(-f(r,t)\lambda(r,t) + D(r,t)\lambda'(r,t)\Big)\Big) \tag{3}$$

where $f(r,t)$ is the drift term, $D(r,t)$ the diffusivity term, and $\lambda'(r,t) = \frac{\partial \lambda}{\partial r}$.

For a given time t, if the movement of agents positioned at a distance r from the origin is unbiased then $f(r,t) = 0$. If there is bias towards the centre of R then $f(r,t) < 0$. If there is bias away from the centre of R then $f(r,t) > 0$. The diffusivity term D gives a measure of the component of movement that can be modelled as simple diffusion.

We note that agent-based processes may evolve such that agent behaviour within B is of no interest, i.e. we are only interested in the behaviour of agents outside of B. For example, an agent-based process might evolve such that agents that reach or enter B stop moving and as such their behaviour and distribution within B is of no interest. In such circumstances the range of r values analysed can be truncated accordingly, for example, $r \in [r_B, r_R]$, for the procedures demonstrated in the following sections.

2.1 Focal K functions for agent-based processes

K functions are a well known summary spatial statistic used for detecting and characterising the deviation of spatial point patterns from complete spatial randomness [15, 17, 18]. A crucial fact that we make use of is that certain spatial K functions can be both easily estimated based on sequential observations of agent positions alone, and then related to the function f defined in Eq (3) allowing one to estimate f. We here adapt K functions for the analysis of agent-based processes, introducing what we refer to as *focal K functions*.

We define the overall density of agents within R as

$$\lambda_R = \frac{n}{|R|}, \tag{4}$$

with n giving the number of agents within R, and $|R|$ giving the volume of R.

We define focal K functions as follows:

$$\begin{aligned} K(r,t) &= \frac{1}{\lambda_R}\mathbb{E}[\text{Number of agents within } r \text{ of the origin at time } t] \\ &= \frac{2\pi}{\lambda_R}\int_0^r s\lambda(s,t)ds, \quad 0 \le r \le r_R. \end{aligned} \tag{5}$$

For each time point t_i we can use the associated spatial point pattern $\boldsymbol{X}(t_i)$ to produce an empirical $\hat{K}(r,t_i)$ function as an estimate of $K(r,t_i)$. We assume observation times are regularly spaced, so that for a positive constant τ, $t_{i+1} - t_i = \tau$ for all $i \ge 1$, and hence we express $\hat{K}(r,t_i)$ as $\hat{K}_i(r)$. We define $\hat{K}_i(r)$ as

$$\hat{K}_i(r) = \frac{1}{\lambda_R}\sum_{k=1}^{n} \mathbb{1}(\delta(k) \le r), \quad 0 \le r \le r_R, \tag{6}$$

where $\delta(k)$ is the distance from the origin for the k^{th} agent in R, and $\mathbb{1}(\cdot)$ is the indicator function. Since our analysis of agents is limited to the behaviour of agents within R we do not require an edge correction term in Eq (6) as is typical when producing K function estimates in spatial statistics, though we expect that one could easily be introduced if required; see [18].

2.2 Focal L functions and the swarming metric

Analogous to spatial statistics we can transform focal K functions to produce focal L functions that are linear with respect to r and visually easier to interpret, especially for smaller r values. Such focal L functions can then be used to produce what we refer to as a swarming metric, giving a single value that measures the level of agent aggregation.

As noted previously we may only be interested in the distribution of agents outside of B, i.e. their distribution within B is not of interest. In this circumstance we can compare a given distribution of agents against a uniform distribution of agents within the annulus $R - B$, with the agent density within B equal to 0. As such, the agent density associated with this distribution is

$$\lambda(r,t) = \begin{cases} 0, & 0 \le r < r_B \\ \frac{\lambda_R r_R^2}{r_R^2 - r_B^2}, & r_B \le r < r_R. \end{cases}$$

It then follows from Eq (5) that in this case the focal K function takes the form

$$K(r,t) = \begin{cases} 0, & 0 \le r < r_B \\ \frac{\pi r_R^2 (r^2 - r_B^2)}{r_R^2 - r_B^2}, & r_B \le r < r_R. \end{cases} \tag{7}$$

Given Eq (7), we adjust the usual definition of the L function found in spatial statistics accordingly, defining focal L functions as

$$L(r,t) = \sqrt{\frac{\frac{r_R^2 - r_B^2}{r_R^2} K(r,t) + \pi r_B^2}{\pi}} - r_B, \quad r_B \le r \le r_R. \tag{8}$$

As such, for a uniform distribution of agents within the given annulus, with the associated focal K function in Eq (7), we have $L(r,t) = r - r_B$ for $r \in [r_B, r_R]$. Eq (8) reverts to the normal L function of spatial statistics by setting $r_B = 0$.

Focal L functions can be used to further simplify the measurement of agent aggregation by producing a simple scalar metric of aggregation as detailed below. By producing a time series of this metric we are able to demonstrate how the distribution of agents changes with time in a simple and easily visualisable manner. We define this *swarming metric* as $M(t)$, and, in order to give a formal definition, we proceed by defining a perfectly aggregated distribution of agents and a perfectly dispersed distribution of agents.

Under this formulation a given number of agents is perfectly aggregated within R if they are all within B. The focal L function associated with this distribution of agents is

$$L(r) = r_R - r_B, \quad r_B \le r \le r_R. \tag{9}$$

The perfectly dispersed configuration of agents will consist of agents being localised on the outer edge of R. The associated focal L function for this distribution of agents is

$$L(r) = \begin{cases} 0 & \text{if } r_B \le r < r_R \\ r_R - r_B & \text{if } r = r_R \end{cases}$$

The swarming metric $M(t)$ can then be produced for a given time point t by transforming the associated focal $L(r,t)$ function as follows:

$$M(t) = \frac{2}{(r_R - r_B)^2} \int_{r_B}^{r_R} \big(L(s,t) - (s - r_B)\big) ds. \tag{10}$$

As such, $M(t)$ has a range of $[-1, 1]$, with -1 being equivalent to agents being perfectly dispersed, and 1 being equivalent to agents being perfectly aggregated. Representative configurations of these are shown in Figure 1. The empirical metric $\hat{M}_i$ is produced using $\hat{L}_i(r)$ in Eq (10).

As a scalar summary, $M(t)$ does not contain the amount of information relating to agent aggregation as the functions $K(r,t)$ or $L(r,t)$. However, we suggest that as a scalar time series of $\hat{M}_i$ values may be easier to interpret than the functional time series $\hat{K}_i(r)$ or $\hat{L}_i(r)$. Furthermore, we suggest that the time series of $\hat{M}_i$ may be most useful in comparing realisations of different experimental conditions in which one expects a divergence between $M(t)$ for each condition.

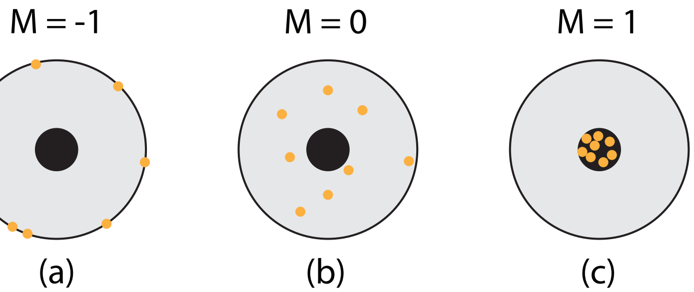


Figure 1. Representative images indicating agent arrangements and associated swarming metric, M, values. a) $M = -1$; b) $M = 0$; c) $M = 1$.

2.3 The relationship between f, D, and focal K functions

The functions f and D can be related to the focal K functions as follows. From Eq (5) we have

$$\frac{\partial K}{\partial t} = \frac{2\pi r}{\lambda_R}\big(- f(r,t)\lambda(r,t) + D(r,t)\lambda'(r,t)\big). \tag{11}$$

We introduce the function $\gamma(r,t) = -\frac{\lambda_R}{2\pi r}\frac{\partial K}{\partial t}$, such that

$$\gamma(r,t) = f(r,t)\lambda(r,t) - D(r,t)\lambda'(r,t). \tag{12}$$

Eq (12) shows that, under the above assumptions, there is a simple linear relationship between $\gamma(r,t)$ and the functions $\lambda(r,t)$ and $\lambda'(r,t)$. Given estimates $\hat{\gamma}_i(r)$, $\hat{\lambda}_i(r)$, and $\hat{\lambda}_i'(r)$ at each time point t_i, one can employ functional linear regression to produce estimates for f and D. This can be achieved using B-spline (or other) smoothing of empirical focal K functions as outlined in Supplementary 1. The major functions used to produce smoothed empirical focal K functions and to perform functional linear regression are found in the *fda* R package [47].

Below we apply two methods, again detailed in Supplementary 1, to estimate f and D: one in which it is assumed both are time-invariant, so that $f(r,t) = f(r)$ and $D(r,t) = D(r)$, and one in which these functions are allowed to be time-varying. Time-invariant estimates are denoted $\hat{f}(r)$ and $\hat{D}(r)$, whereas the time-varying estimates produced at each discrete viewing time are denoted $\hat{f}(r,t_i) = \hat{f}_i(r)$, and $\hat{D}(r,t_i) = \hat{D}_i(r)$.

It may be known *a priori* whether f and D are time-invariant or time-varying due to experimental conditions, e.g. due to the induction of a constant chemokine gradient. Also, if $K_i(r)$ strongly suggests that $f(r,t) = 0$, and D is assumed to be time-invariant, time-invariant estimates may be suitable. In general however, we suggest that most processes will likely demonstrate a degree of time dependence for f and D.

We suggest initially analysing data using time-varying f and D, and considering whether the functional estimates exhibit any obvious trends. If no obvious trends are present, it may be appropriate to proceed to using a time-invariant estimate for f and D. We suggest that even for cases that demonstrate trends in the time-varying estimates, the time-invariant estimates $\hat{f}(r)$ and $\hat{D}(r)$ may provide a useful measure of overall drift and diffusivity.

3 Simulations of simple 2D agent-based models of chemotaxis

We here apply the above approaches to simple agent-based models (ABMs) of chemotaxis. The models are based on other off-lattice ABMs of cell movement [43,48].

Each ABM consists of 1000 individual agents moving within an enclosed circular region, R, with radius $r_R = 100$, with a circular central region B, with radius $r_B = 20$, located at the centre of R.

Simulations commence with all agents randomly distributed within R under a uniform distribution. The models are discrete time processes with time steps of constant duration, τ. We set $\tau = 1$ for all simulations. Each simulation runs for 200 time steps.

Each ABM evolves via a similar underlying process. A random sequential update method is used to perform simulations [40,49]. That is, at each time step, 1000 sequential random selections of the 1000 agents are made, with selected agents performing a movement as described below. As such, an agent may move more than once, or not at all, for a given time step. If an agent is selected to move, they move in a direction, θ, drawn from a probability density function (PDF) that is dependent on an idealised chemokine gradient. Once a direction is selected the agent moves a constant distance, Δ, in that direction. We use a reflecting boundary at the edge of R, such that if an agent attempts to move out of R it is reflected back into R. These processes are not exclusion processes and agents can occupy the same position in space. We set $\Delta = 1$ for all simulations.

Zero chemotaxis

For the ABM of zero chemotaxis agents are not biased to move in any particular direction; θ is drawn from a uniform distribution over $[0, 2\pi]$.

Chemotaxis

To model chemotaxis we use a similar concept as in [48,50,51]. An idealised attractive chemokine gradient is established radiating from the origin. Agents are more likely to move up the gradient with the likelihood determined by the steepness of the local gradient at the position of the agent.

We take the idealised chemokine concentration, $v(r,t)$, to be the product

$$v(r,t) = V(r)T(t), \tag{13}$$

with $V(r)$ giving the distribution of chemokine with respect to r, and $T(t)$ giving a time dependent function that allows the chemokine concentration to change with time.

The steepness of $v(r,t)$ determines how biased agent movements are. Thus we use the absolute value $|\frac{\partial v}{\partial r}|$ to parameterise the distribution for θ.

For each agent the angle of movement is drawn from a von Mises distribution [20,52]. Specifically, for an agent at position (r, α) in polar coordinates, at time t, $\theta(r, \alpha, t)$ is drawn using the probability density function

$$p(\theta; r, \alpha, t) = \frac{\exp(|\frac{\partial v}{\partial r}| \cos(\theta - (\pi + \alpha)))}{2\pi I_0(|\frac{\partial v}{\partial r}|)}, \tag{14}$$

with $I_0(\cdot)$ giving the modified Bessel function of the first kind of order 0. As such, the expectation of agent movements is $\alpha + \pi$, i.e. towards the origin, with the variability around this value determined by the steepness of the chemokine gradient, $|\frac{\partial v}{\partial r}|$.

For our simulations we set $V(r)$ to be a Gaussian function;

$$V(r) = \beta e^{\frac{-r^2}{2\sigma^2}}, \tag{15}$$

with the parameter $\beta > 0$ increasing the steepness of the gradient for larger values and σ increasing the “diffusivity” of the chemokine, providing a measure of how far the chemokine signal is effective. For all chemotaxis simulations we set $\beta = 40$ and $\sigma = 30$.

3.1 Simulation results

For each case we visualise the empirical functional time series produced using rainbow plots [53, 54]. Rainbow plots are a useful approach in visualising functional time series data; the order in which functions occur in time is indicated by colour with early functions red, followed by orange, yellow, green, blue, indigo, with the latest functions violet.

For each simulation we then produce estimates $\hat{f}$ and $\hat{D}$ using functional linear modelling as described in Supplementary 1 using the *fda* R package [47]. For each simulation we use 40 splines and a smoothing parameter of 1 to produce the smoothed empirical focal K functions. We use order 6 splines, since this will ensure that the second derivatives of the smoothed focal K functions, which are used to produce the functional covariate $\hat{\lambda}'(r)$, will be smooth [55]. The number of splines used and the smoothing parameter employed for the estimated f and D must also be selected, and we use 20 splines and smoothing parameters of 10 for both f and D for all simulations.

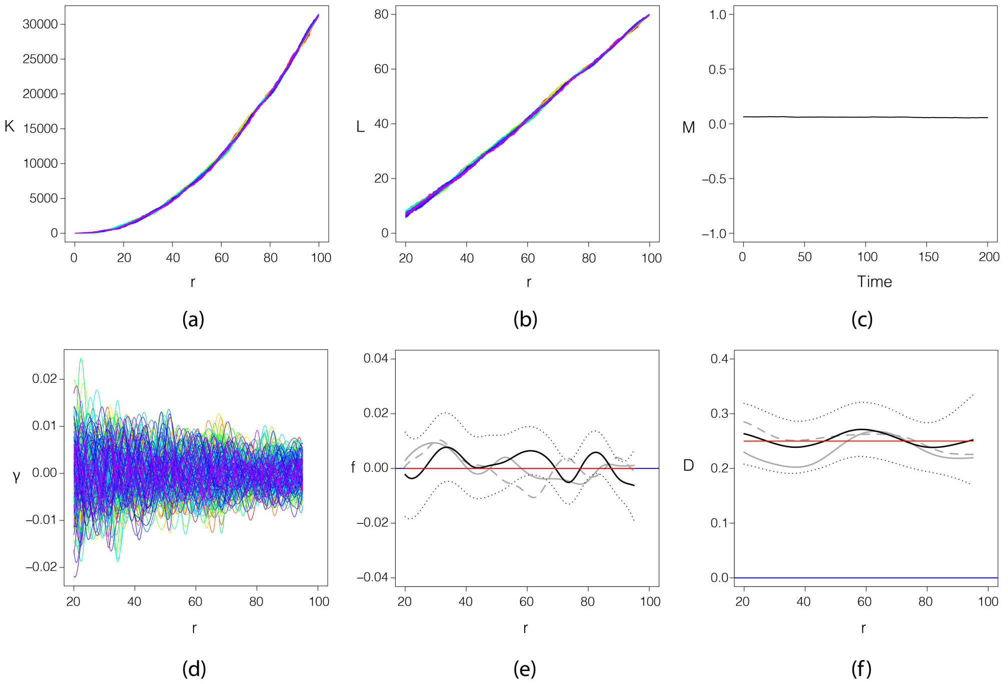
In addition, we note that we find estimates for f and D can sometimes be inaccurate around $r = 0$ and $r = r_R$, due the the inherent instability of higher order derivatives of splines at boundaries [55]. We find that estimating f and D over a truncated region, $[r_1, r_2]$, with $r_1 > 0$ and $r_2 < r_R$, but still close to these values, can produce better estimates. For all simulations we use $r_1 = 5$, and $r_2 = 95$.

For a discussion regarding implementation and robustness of estimates for f and D with respect to smoothness of the focal K functions, temporal resolution, agent density, and number of time steps, see Supplementary 2. We find that estimates are robust to even small agent densities and limited numbers of time steps. As demonstrated in Supplementary 2, we note that there is a dependence of the estimates for D on the smoothness of the focal K functions, and suggest that using relatively smooth focal K functions that still capture the variability of the underlying step functions.

We are able to compare $\hat{f}$ and $\hat{D}$ against approximations of the true drift and diffusivity functions associated with each ABM calculated using the first and second moment of agent movements [20, 40, 46, 52, 56]. See Supplementary 3 for details.

Zero chemotaxis

For the case of zero chemotaxis we run three replicates with different initial starting agent positions. For simplicity we present $\hat{K}_i(r)$, $\hat{L}_i(r)$, $\hat{M}_i$, and $\hat{\gamma}_i(r)$ for only the first replicate, see Figure 2. The estimates $\hat{f}(r)$ and $\hat{D}(r)$ obtained using functional linear modelling for each of the three simulations are plotted together on the relevant panels to allow for comparison. We see that for the given simulation the $\hat{K}_i(r)$ and $\hat{L}_i(r)$ functional time series do demonstrate and obvious trend. This is reflected in the stable trajectory of the swarming metric $\hat{M}_i$. We note that $\hat{M}_i > 0$ as agents are initially distributed uniformly throughout R, with some agents placed within B, and $\hat{M}_i$ compares the distribution of agents to the uniform distribution of agents within the region $R - B$, i.e. with 0 agents within B. The estimates $\hat{f}(r)$ vary around 0, strongly indicating that no bias of agent movement towards B is detected for this model. The estimates $\hat{D}(r)$ perform well in estimating the true $D(r)$ function. The estimates $\hat{f}(r)$ and $\hat{D}(r)$ are similar between the three simulations.



(a) (b) (c)

(d) (e) (f)

Figure 2. Zero chemotaxis: (a) $\hat{K}_i(r)$; (b) $\hat{L}_i(r)$; (c) $\hat{M}_i$; (d) $\hat{\gamma}_i(r)$; (e) $\hat{f}(r)$; and (f) $\hat{D}(r)$. The estimates $\hat{f}(r)$ and $\hat{D}(r)$ for the first replicate are given by solid black lines, with the dotted black lines giving the 95% confidence intervals for these estimates, the red solid lines giving the true $f(r)$ and $D(r)$ functions, and the blue lines giving 0. Estimates for $\hat{f}(r)$ and $\hat{D}(r)$ associated with the two additional replicates using different initial positions are added to panels (e) and (f) as solid grey and dashed grey lines.

Time-invariant chemotaxis

For the model of time-invariant chemotaxis we set $T(t) = 1$, such that attraction exerted on agents is time-invariant. We run three replicates with different initial starting agent positions. Similarly to above we present $\hat{K}_i(r)$, $\hat{L}_i(r)$, $\hat{M}_i$, and $\hat{\gamma}_i(r)$ for only the first replicate, see Figure 3. We can see that the $\hat{K}_i(r)$ and $\hat{L}_i(r)$ functions associated with the first experiment increase over time for all r values. This is reflected in the positive trend in the swarming metric $\hat{M}_i$. The estimates $\hat{f}(r)$ for each simulation perform very well in estimating the true $f(r)$ function. The estimates $\hat{D}(r)$ likewise perform well in estimating the true $D(r)$ function.

Linearly increasing chemotaxis

To model chemotaxis with linearly increasing strength we set

$$T(t) = 1 + 0.02t. \tag{16}$$

We perform and analyse only one simulation for this model, see Figure 4. The functional time series $\hat{f}_i(r)$ for $i = 15, ..., 186$ is produced using a moving window-type estimator, as discussed in Supplementary 1, using a window width of 29 time points. We observe that the functional time series $\hat{f}_i(r)$ performs well in estimating the functional time series of the true $f(r, t_i)$ functions for $i = 15, ..., 186$, with strength of attraction increasing with time for this simulation. We note similar results, with slightly reduced accuracy, are obtained for the estimates $\hat{D}_i(r)$.

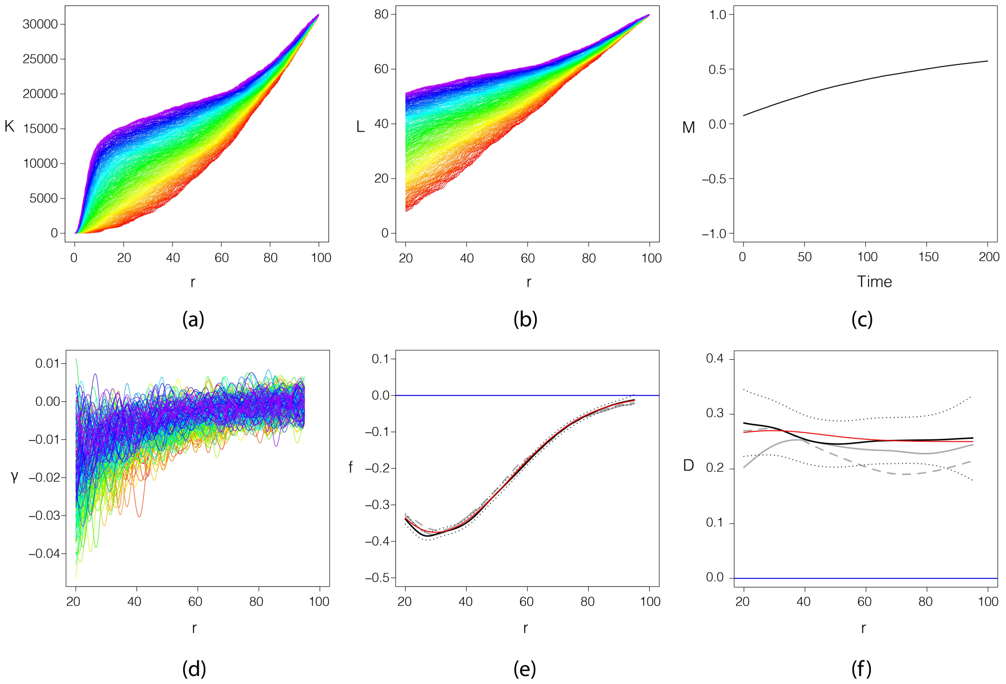


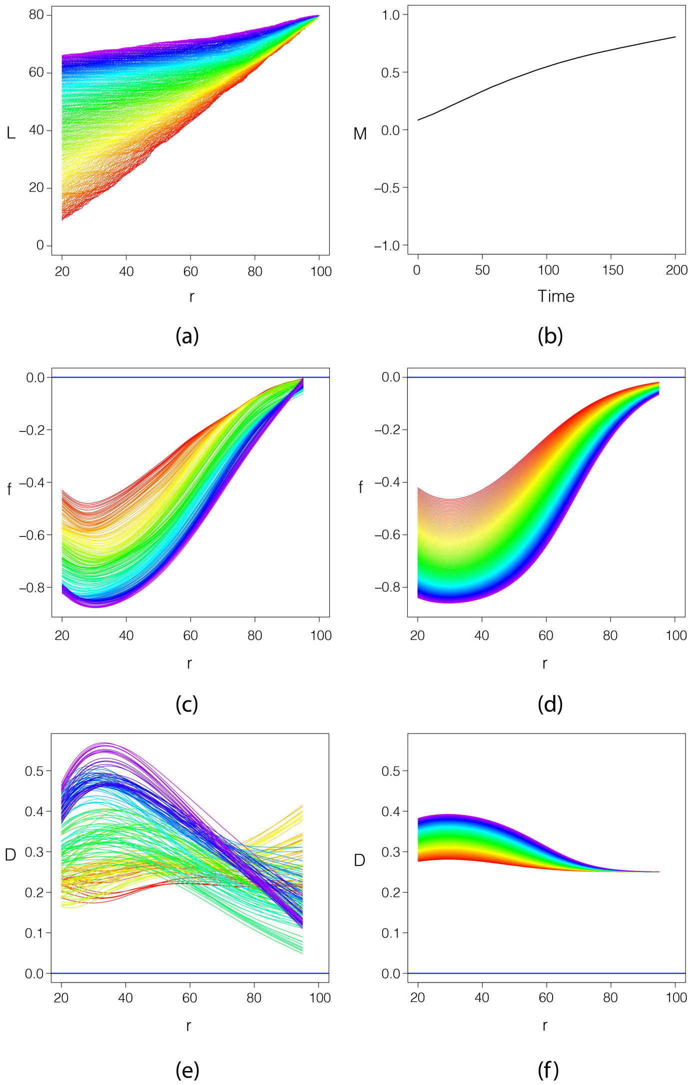
Figure 3. Time-invariant chemotaxis: (a) $\hat{K}_i(r)$; (b) $\hat{L}_i(r)$; (c) $\hat{M}_i$; (d) $\hat{\gamma}_i(r)$; (e) $\hat{f}(r)$; and (f) $\hat{D}(r)$. The estimates $\hat{f}(r)$ and $\hat{D}(r)$ for the first replicate are given by solid black lines, with the dotted black lines giving the 95% confidence intervals for these estimates, the red solid lines giving the true $f(r)$ and $D(r)$ functions, and the blue lines giving 0. Estimates for $\hat{f}(r)$ and $\hat{D}(r)$ associated with the two additional replicates using different initial positions are added to panels (e) and (f) as solid grey and dashed grey lines.

4 T Cell motility experiments

Cytotoxic T cells (CTLs) are specialised immune cells that seek out and destroy cancer cells, and they are the principal mediators of adoptive cell transfer immunotherapies [57]. These burgeoning therapies have proven revolutionary in the treatment of blood-borne cancers, yet have thus far not shown similar success when applied to solid tumours, which leukocytes infiltrate poorly [58, 59]. Therefore, understanding the spatial movements of T cell populations with respect to tumour masses is key to the development of effective immunotherapies targeting solid malignancies.

Recently, CTLs engaging tumour targets have been demonstrated to engage in chemokine signalling, attracting distant CTLs [7]. Here, we analysed *ex vivo* experiments of CTLs engaging tumour cells, presented in [7], (see Figure 1 in [7]). The experimental details, along with movies of the experiments, and additional related results, can be found in [7].

In brief the experiments occur within a single well of a 96-well optical plate, and consists of a central dense collection of tumour cells and extra-cellular matrix (referred to as tumouroid) surrounded by fluorescent effector CD8+ T cells embedded in a 3D collagen matrix. We consider two different experimental conditions. One we refer to as the pre-embedded-cognate experiment, in which the tumouroid consists of *cognate* tumour cells and pre-embedded, non-fluorescent, CTLs, i.e. the tumouroid already contains CTLs at the start of the experiment. The cognate tumour cells in this tumouroid present the activating antigen on their surface, thus leading to T cell arrest,



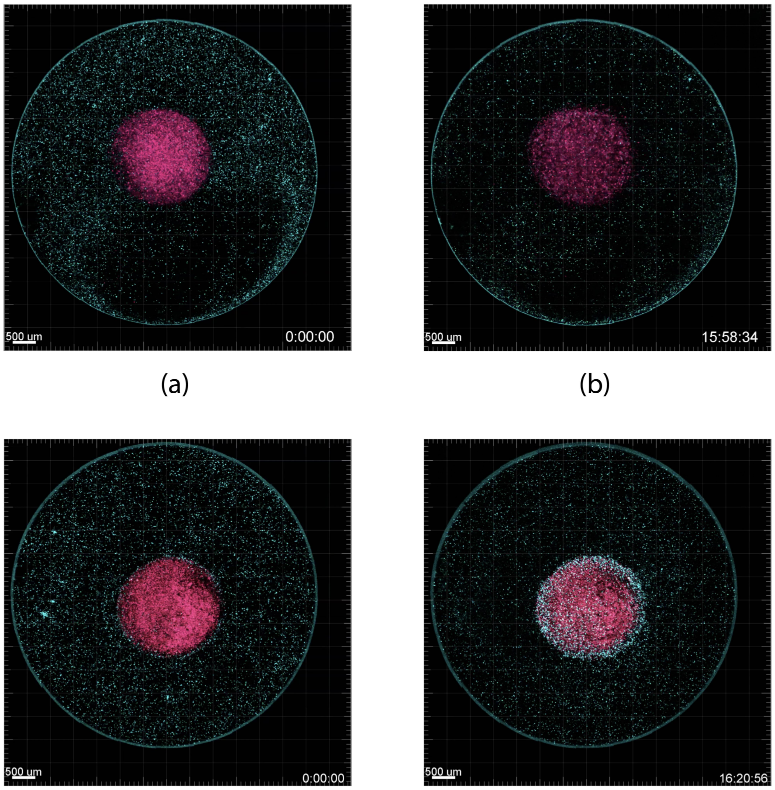
(a) (b) (c) (d) (e) (f)

Figure 4. Linearly increasing chemotaxis: (a) $\hat{L}_i(r)$; (b) $\hat{M}_i$; (c) $\hat{f}_i(r)$ for $i = 15, ..., 186$; (d) true $f(r, t_i)$ for $i = 15, ..., 186$; (e) $\hat{D}_i(r)$ for $i = 15, ..., 186$; (f) true $D(r, t_i)$ for $i = 15, ..., 186$.

activation and tumour cell clearance. The other experimental condition we refer to as the non-cognate experiment, in which the tumouroid contains *non-cognate* tumour cells only. Non-cognate tumour cells do not activate CTLs, which hence do not arrest on contact and thus continue their search.

We hypothesised that in the pre-embedded-cognate experiment the peripheral CTLs surrounding the tumouroid would be attracted to the tumouroid via chemotaxis. Specifically, we expected that CTLs initially pre-embedded within the tumouroid, and peripheral CTLs that made contact with the cognate cells making up the tumouroid during the experiment, would both become activated and secrete a chemokine to recruit more distant CTLs. In comparison, we hypothesised that in the non-cognate experiment CTLs would not become activated upon contact with the non-cognate cells making up the tumouroid, and hence there would be no detection of chemotaxis of CTLs towards the tumouroid.

Images were taken of the entire closed well every 5-6 minutes over approximately 16 hours and the positional data for the fluorescing CTLs, i.e. those initially dispersed around the tumouroid, were recorded during the experiment. We note that in contrast the non-fluorescent CTLs pre-embedded in the tumouroid were not imaged. Owing to the resolution required and the comparatively large area being imaged (>6mm in diameter), imaging was constrained to a 2D slice through the x-y plane of the entire well. Images of the first and last time point from both experiments are presented in Figure 5.



(a) (b)

(c) (d)

Figure 5. Images of non-cognate and pre-embedded-cognate experiments: (a) image from first time point of non-cognate experiment; (b) time point from final time point of non-cognate experiment; (c) image from first time point of pre-embedded-cognate experiment; (d) image from final time point of pre-embedded-cognate experiment. CTLs are blue, with tumouroids red. Scale bars: 500 µm. Time stamps indicate time of image acquisition. Images are from experiments reported in [7]

For each experiment we need to define the regions R and B. In 2D the tumoroids are approximately circular. We choose to set r_B equal to the average of longest and shortest distances from the centre of geometric mass to the edge of the tumouroid. Owing to the technicality of sample preparation, the tumouroid cannot be precisely positioned such that its centre of mass is at the exact centre of the well. We set r_R to the shortest distance from the tumouroid centre to the well edge. The radius of the base of the well is approximately $3150\mu m$. For the non-cognate experiment we set $r_B = 1076.75\mu m$ and $r_R = 2855\mu m$. For the pre-embedded-cognate experiment we set

$r_B = 1099$ and $r_R = 2985$.

While the imaged system is closed to the entry and egress of cells in the horizontal plane due to the walls of the well, it is open with respect to cell movements in the vertical axis due to the 3D nature of the collagen matrix. We expected that the net effect of this vertical movement would be negligible. We noted that for both experiments the trend in cell number is relatively stable for the first 10 hours of the experiments. Since we expect any evidence for the existence of chemotaxis to be evident during this period we continued by analysing only the first 10 hours of each experiment. In addition, a relatively sharp decrease in cell number was noted to occur over the first hour of the pre-embedded-cognate experiment. Since we still expected evidence of chemotaxis to still be evident after the first hour of the experiment we continued by removing the first hour of the pre-embedded-cognate experiment from our analysis. See Figure 6. for the associated cell numbers and $\hat{K}_i(r)$ functions for the remaining time points.

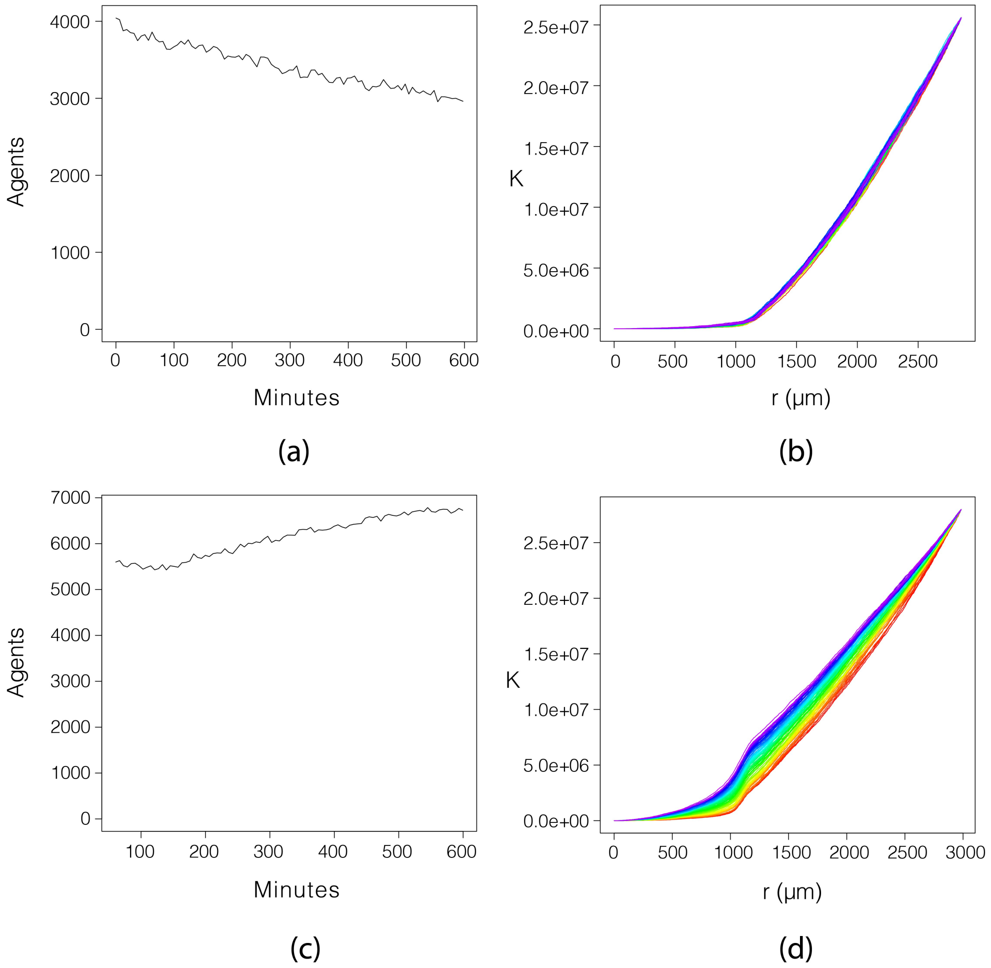


Figure 6. Cell number dynamics: (a) Total number of cells imaged within R for each time step analysed for the non-cognate experiment; (b) associated $\hat{K}_i(r)$ functions for the non-cognate experiment; (c) Total number of cells imaged within R for each time step analysed for the pre-embedded-cognate experiment; (d) associated $\hat{K}_i(r)$ functions for the pre-embedded-cognate experiment.

The approach presented in previous sections requires that the number of agents observed does not change with time. In Supplementary 4 we consider the case where agent numbers can change with time as a result of an agent birth-death process (in addition, we also consider the unrelated case where R has an open boundary). We demonstrate that if it can be assumed that the birth-death rate is spatially-invariant and linear with respect to $\lambda(r, t)$, then it has no impact on the interpretation of the change in K and L functions, the swarming metric M, or the approach to estimating f and D. For the purposes of this paper we continue with this assumption for the analysis of the T cell motility experiments.

4.1 T cell experiment results

For both experiments we set the number of splines to 200 and a smoothing parameter of 10^8. We use 20 spines for estimating both f and D, with smoothing parameters 10^{-1}. We estimate f and D over $[20, r_R - 20]$.

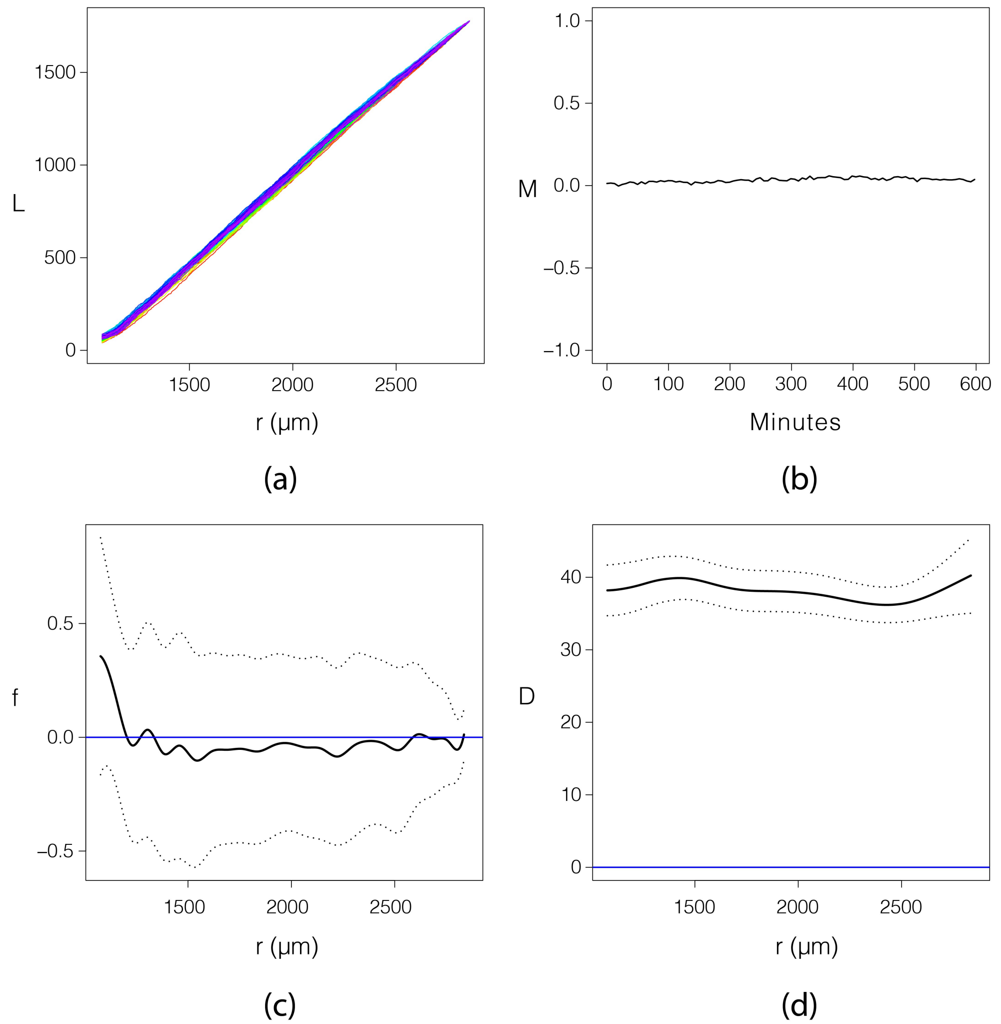
Relevant plots for the non-cognate experiment are displayed in Figure 7. The functional time series for $\hat{L}_i(r)$ has no apparent trend and appears nearly linear, suggesting that agents are approximately uniformly distributed outside the tumouroid. The associated time series for the swarming metric $\hat{M}_i$ varies around 0 for the duration of the experiment. Since $\hat{L}_i(r)$ and $\hat{M}_i$ do not demonstrate any trend we continue by producing time-invariant functional estimates for f and D. We note that the estimate $\hat{f}(r)$ is approximately 0 for all r values. This result was consistent with our expectation that CTLs would not be attracted towards the tumouroid for this experiment.

Relevant plots for the pre-embedded-cognate experiment are displayed in Figure 8. The functional time series for $\hat{L}_i(r)$ demonstrates a positive trend, with $\hat{L}_i(r)$ appearing to increase for all r values over time. The associated time series for the swarming metric $\hat{M}_i$ demonstrates an associated increase. Since these time series demonstrate obvious trends we first estimate time-varying f and D. The associated estimates $\hat{f}_i(r)$ and $\hat{D}_i(r)$ using a window width of 19 appear relatively stable over time. As such, we proceed to produce time-invariant estimates $\hat{f}(r)$ and $\hat{D}(r)$. The time-invariant $\hat{f}(r)$ is negative for all r values, and approaches 0 as r increases towards r_R. This result is consistent with our expectation that CTLs would be attracted to the tumouroid via chemotaxis for this experiment. We note that $\hat{f}(r)$ is strikingly similar in form to that in our simple ABM of chemotaxis, suggesting that the chemokine concentration in the experiment may be well approximated by a Gaussian distribution and that CTLs respond to the steepness of the chemokine gradient. The estimate for $\hat{D}(r)$ is similar to that for the non-cognate experiment, though we note that it is decreased for r values between r_B and approximately 1500.

5 Discussion

In this paper we have presented a novel approach for detecting and characterising directional bias in agent movements solely from positional data for agent-based processes. Our approach does not require the tracking of agents over time, meaning that our approach can be utilised when temporal tracking of individual agents is not available. For instance, in comparison to collecting positional data alone, individual cell tracking typically requires several undesirable features, including a high sampling rate with associated phototoxicity, a small field of view, and a relatively low cell density [10, 12–14].

Techniques from spatial statistics, such as the pair correlation function, are increasingly being applied to agent-based processes [51, 60–70]. Our approach employs a novel combination of both spatial statistics and functional data analysis. There is great scope to combine spatial statistical methods and functional data analysis to quantify



(a) (b)

(c) (d)

Figure 7. Non-cognate experiment: (a) rainbow plot of $\hat{L}_i(r)$; (b) $\hat{M}_i$; (c) time-invariant $\hat{f}(r)$ with 95% confidence intervals (dotted black lines); (d) time-invariant $\hat{D}(r)$ with 95% confidence intervals (dotted black lines).

the dynamics in agent-based processes. In previous work we combined such approaches in the analysis of spatially homogeneous agent-based processes [51]. As advances in microscopy, and also GPS tracking, yield detailed observations of dynamic and complex processes such as immune cell responses, animal behaviour, or search and rescue robot swarms, more advanced statistical metrics are increasingly needed.

The approach presented here introduces focal K functions that provide a measure of agent aggregation around an area of interest for a given time point. We use these functions to define a related scalar, M, which we refer to as the swarming metric. The functional and scalar time series associated with the focal K functions and the swarming metric, respectively, provide useful information regarding the evolution of an agent-based processes with respect to agent aggregation. These methods are general, and do not rely upon any assumptions regarding the underlying agent behaviour. We envisage that the swarming metric in particular may provide a useful means of comparing different experimental conditions of agent-based processes.

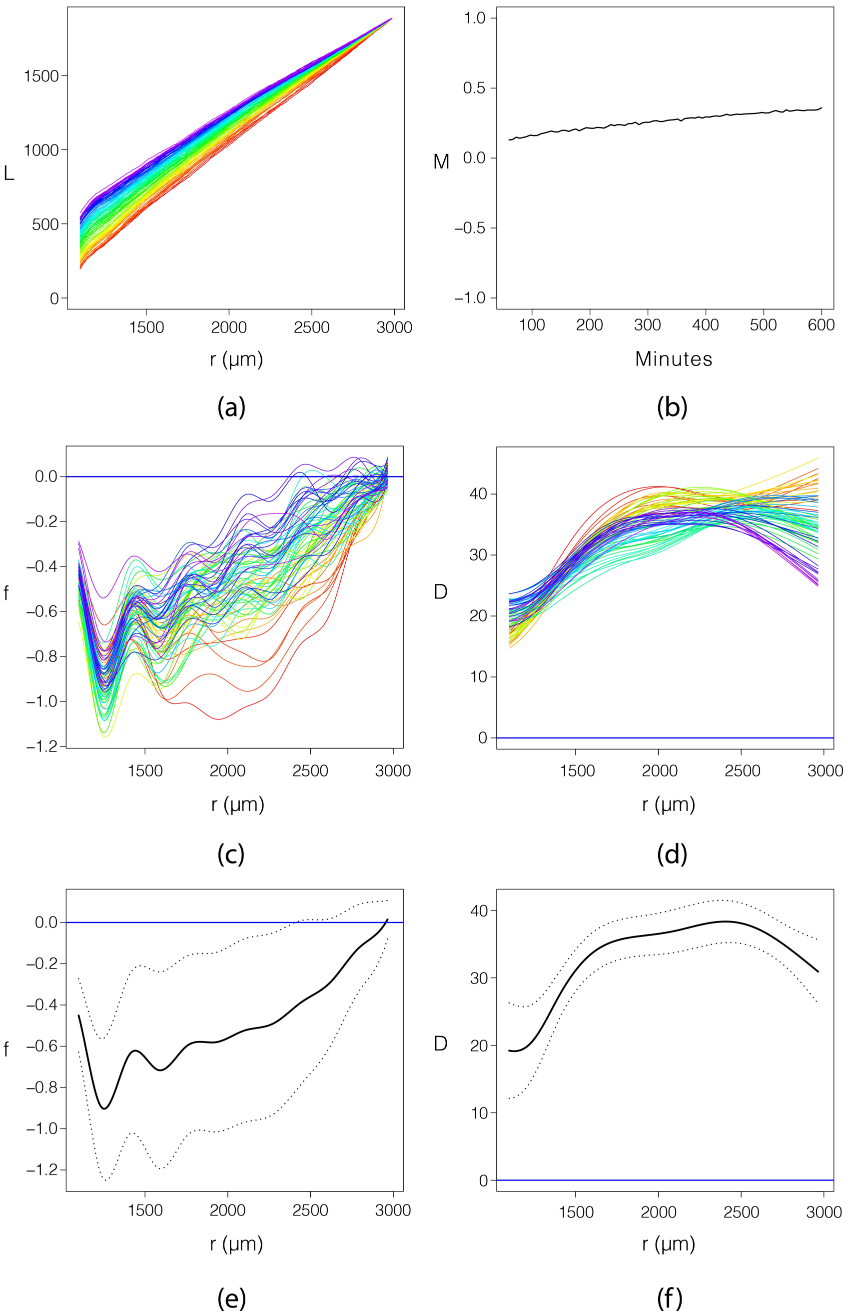


Figure 8. Pre-embedded-cognate experiment: (a) rainbow plot of $\hat{L}_i(r)$; (b) $\hat{M}_i$; (c) rainbow plot of $\hat{f}_i(r)$; (d) rainbow plot of $\hat{D}_i(r)$; (e) time-invariant $\hat{f}(r)$ with 95% confidence intervals (dotted black lines); (f) time-invariant $\hat{D}(r)$ with 95% confidence intervals (dotted black lines).

If the evolution of population density with time is well approximated by the Fokker-Planck equation, an advection-diffusion PDE, then these focal K functions can be analysed using a functional linear model to estimate the associated drift and diffusivity terms. Importantly, this allows for any bias in agent movements to be detected and characterised. Our approach allows for the analysis of agent-based processes with both time-invariant as well as time-varying attraction or repulsion. In addition, there are no requirements regarding whether the system is or is not in a steady state.

We demonstrate our approach using simulations, analysing a set of simple 2D random walk agent-based models. We find that our approach performs well for these sets of simulations. Moreover, we find that estimates appear to be robust with respect to the density of agents and the number of observations. We find that estimates for diffusivity are dependent on the degree of smoothing used in smoothing the focal K functions, and that this dependence is accentuated by reduced temporal resolution. Whilst we suggest that using a relatively large level of smoothing can ensure accurate estimates for diffusivity, determining an explicit approach for selecting the most appropriate degree of smoothing represents a useful extension.

We have here only analysed simulations of agents undergoing a random walk process without persistence or exclusion. Both these properties may be relevant in considering specific agent-based processes, such as bacterial chemotaxis, or cellular processes with high agent density. Resent results have demonstrated that drift-diffusion PDEs can be produced to approximate the evolution of agent-based processes exhibiting bias and persistence in the setting of modelling bacterial chemotaxis [71]. In addition, drift-diffusion type PDEs can be produced to approximate the evolution a lattice-based agent-based process that includes bias, persistence, and exclusion, with the drift and diffusion terms being dependent on agent density [69]. Determining whether our approach can be employed to analyse similar off-lattice processes represents a useful extension of this work.

We apply our approach to the analysis of *ex vivo* experiments consisting of live cell imaging of T cells interacting with a central tumouroid consisting of tumour cells. These represent a portion of novel experimental results used to demonstrate that cytotoxic T cells swarm to tumours [7]. Consistent with these findings, our analysis of these experiments identifies an attractive signal that biases cell movements towards the tumouroid. As outlined in [7], the attraction exhibited by T cells represents homotypic signalling; that is, the activated cytotoxic T cells in contact with cognate tumour cells produce the chemoattractant that induces chemotaxis in other cytotoxic T cells. As such, this process represents what we refer to as *agent-driven swarming*, such that the swarming behaviour exhibited is driven, at least in part, by the presence of other agents at the area of interest. This differentiates agent-driven swarming from other types of processes in which the attractive signal is derived from sources other than the agents. It can be of considerable interest to researchers whether observed swarming is agent-driven or otherwise. If it is known *a priori* that the agents within a given process represent the only possible source of attractant, or this can be proved experimentally, then detecting that agents movements are biased towards a given area is enough to determine agent-driven swarming. However, it may represent a useful extension of this work to produce a statistical technique for detecting agent-driven swarming in circumstances in which this is not known and cannot be demonstrated experimentally.

6 Acknowledgements

This work was supported by Australian Research Council Discovery Project grant DP180102458 to MR and MB. MR acknowledges support from the University of Sydney Centre for Advanced Food Enginomics. MB is additionally supported by funding via EMBL Australia. GR was partially supported by a Discovery grant and Accelerator grant from the Natural Science and Engineering Research Council of Canada. MB acknowledges Bitplane AG for an Imaris Developer Licence. The authors thank Jorge

Luis Galeano Niño for technical assistance with experimental work.

References

1. Schumacher LJ, Kulesa PM, McLennan R, Baker RE, Maini PK. Multidisciplinary approaches to understanding collective cell migration in developmental biology. Open biology. 2016;6(6):160056.

2. Lamalice L, Le Boeuf F, Huot J. Endothelial cell migration during angiogenesis. Circulation research. 2007;100(6):782–794.

3. Gillitzer R, Goebeler M. Chemokines in cutaneous wound healing. Journal of leukocyte biology. 2001;69(4):513–521.

4. Callaghan T, Khain E, Sander LM, Ziff RM. A stochastic model for wound healing. Journal of statistical physics. 2006;122(5):909–924.

5. Dianqing W. Signaling mechanisms for regulation of chemotaxis. Cell research. 2005;15(1):52.

6. Lämmermann T, Afonso PV, Angermann BR, Wang JM, Kastenmüller W, Parent CA, et al. Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo. Nature. 2013;498(7454):371–375.

7. Niño JLG, Pageon SV, Tay SS, Colakoglu F, Kempe D, Hywood J, et al. Cytotoxic T Cells swarm by homotypic chemokine signalling. eLife. 2020;9:e56554.

8. Chtanova T, Schaeffer M, Han SJ, van Dooren GG, Nollmann M, Herzmark P, et al. Dynamics of neutrophil migration in lymph nodes during infection. Immunity. 2008;29(3):487–496.

9. Ng LG, Qin JS, Roediger B, Wang Y, Jain R, Cavanagh LL, et al. Visualizing the neutrophil response to sterile tissue injury in mouse dermis reveals a three-phase cascade of events. Journal of Investigative Dermatology. 2011;131(10):2058–2068.

10. Meijering E, Dzyubachyk O, Smal I, et al. Methods for cell and particle tracking. Methods Enzymol. 2012;504(9):183–200.

11. Tong PL, Roediger B, Kolesnikoff N, Biro M, Tay SS, Jain R, et al. The skin immune atlas: three-dimensional analysis of cutaneous leukocyte subsets by multiphoton microscopy. Journal of Investigative Dermatology. 2015;135(1):84–93.

12. Beltman JB, Marée AF, De Boer RJ. Analysing immune cell migration. Nature Reviews Immunology. 2009;9(11):789–798.

13. Svensson CM, Medyukhina A, Belyaev I, Al-Zaben N, Figge MT. Untangling cell tracks: Quantifying cell migration by time lapse image data analysis. Cytometry Part A. 2018;93(3):357–370.

14. Gabriel EM, Fisher DT, Evans S, Takabe K, Skitzki JJ. Intravital microscopy in the study of the tumor microenvironment: from bench to human application. Oncotarget. 2018;9(28):20165.

15. Ripley BD. Modelling spatial patterns. Journal of the Royal Statistical Society Series B (Methodological). 1977;39(2):172–212.

16. Ripley BD. Statistical inference for spatial processes. Cambridge university press; 1991.
17. Diggle PJ. Statistical analysis of spatial point patterns. Hodder Arnold; 2003.
18. Diggle PJ. Statistical analysis of spatial and spatio-temporal point patterns. CRC Press; 2013.
19. Risken H. Fokker-planck equation. In: The Fokker-Planck Equation. Springer; 1996. p. 63–95.
20. Codling EA, Plank MJ, Benhamou S. Random walk models in biology. Journal of the Royal Society Interface. 2008;5(25):813–834.
21. Chavanis PH. Nonlinear mean field Fokker-Planck equations. Application to the chemotaxis of biological populations. The European Physical Journal B. 2008;62(2):179–208.
22. Kevrekidis IG, Gear CW, Hyman JM, Kevrekidid PG, Runborg O, Theodoropoulos C, et al. Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. Communications in Mathematical Sciences. 2003;1(4):715–762.
23. Erban R, Kevrekidis IG, Adalsteinsson D, Elston TC. Gene regulatory networks: A coarse-grained, equation-free approach to multiscale computation. The Journal of chemical physics. 2006;124(8):084106.
24. Yates CA, Erban R, Escudero C, Couzin ID, Buhl J, Kevrekidis IG, et al. Inherent noise can facilitate coherence in collective swarm motion. Proceedings of the National Academy of Sciences. 2009;106(14):5464–5469.
25. Weinan E, Engquist B, Li X, Ren W, Vanden-Eijnden E. The heterogeneous multiscale method: A review. In: Commun. Comput. Phys. Citeseer; 2007.
26. Lagergren JH, Nardini JT, Michael Lavigne G, Rutter EM, Flores KB. Learning partial differential equations for biological transport models from noisy spatio-temporal data. Proceedings of the Royal Society A. 2020;476(2234):20190800.
27. Nardini JT, Baker RE, Simpson MJ, Flores KB. Learning differential equation models from stochastic agent-based model simulations. arXiv preprint arXiv:201108255. 2020;.
28. Messenger DA, Bortz DM. Weak SINDy For Partial Differential Equations. arXiv preprint arXiv:200702848. 2020;.
29. Lagergren JH, Nardini JT, Baker RE, Simpson MJ, Flores KB. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. arXiv preprint arXiv:200513073. 2020;.
30. Rudy S, Alla A, Brunton SL, Kutz JN. Data-driven identification of parametric partial differential equations. SIAM Journal on Applied Dynamical Systems. 2019;18(2):643–660.
31. Keller EF, Segel LA. Initiation of slime mold aggregation viewed as an instability. Journal of Theoretical Biology. 1970;26(3):399–415.
32. Keller EF, Segel LA. Model for chemotaxis. Journal of theoretical biology. 1971;30(2):225–234.

33. Hillen T, Painter KJ. A user's guide to PDE models for chemotaxis. Journal of mathematical biology. 2009;58(1-2):183.

34. Ferguson EA, Matthiopoulos J, Insall RH, Husmeier D. Statistical inference of the mechanisms driving collective cell movement. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2017;66(4):869–890.

35. Alber M, Chen N, Lushnikov PM, Newman SA. Continuous macroscopic limit of a discrete stochastic model for interaction of living cells. Physical review letters. 2007;99(16):168102.

36. Lushnikov PM, Chen N, Alber M. Macroscopic dynamics of biological cells interacting via chemotaxis and direct contact. Physical Review E. 2008;78(6):061904.

37. Binder BJ, Landman KA. Exclusion processes on a growing domain. Journal of theoretical biology. 2009;259(3):541–551.

38. Bruna M, Chapman SJ. Excluded-volume effects in the diffusion of hard spheres. Physical Review E. 2012;85(1):011103.

39. Yates CA, Baker RE, Erban R, Maini PK. Going from microscopic to macroscopic on nonuniform growing domains. Physical Review E. 2012;86(2):021921.

40. Hywood JD, Landman KA. Biased random walks, partial differential equations and update schemes. The ANZIAM Journal. 2013;55(2):93–108.

41. Middleton AM, Fleck C, Grima R. A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. Journal of theoretical biology. 2014;359:220–232.

42. Dyson L, Baker RE. The importance of volume exclusion in modelling cellular migration. Journal of mathematical biology. 2015;71(3):691–711.

43. Irons C, Plank MJ, Simpson MJ. Lattice-free models of directed cell motility. Physica A: Statistical Mechanics and its Applications. 2016;442:110–121.

44. Matsiaka OM, Penington CJ, Baker RE, Simpson MJ. Continuum approximations for lattice-free multi-species models of collective cell migration. Journal of theoretical biology. 2017;422:1–11.

45. R Core Team. R: A Language and Environment for Statistical Computing; 2018. Available from: `https://www.R-project.org/`.

46. Hywood JD, Hackett-Jones EJ, Landman KA. Modeling biological tissue growth: Discrete to continuum representations. Physical Review E. 2013;88(3):032704.

47. Ramsay JO, Wickham H, Graves S, Hooker G. fda: Functional Data Analysis; 2017. Available from: `https://CRAN.R-project.org/package=fda`.

48. Binny RN, Haridas P, James A, Law R, Simpson MJ, Plank MJ. Spatial structure arising from neighbour-dependent bias in collective cell movement. PeerJ. 2016;4:e1689.

49. Chowdhury D, Schadschneider A, Nishinari K. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. Physics of Life reviews. 2005;2(4):318–352.

50. Binny RN, Plank MJ, James A. Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. Journal of The Royal Society Interface. 2015;12(106):20150228.

51. Hywood JD, Read MN, Rice G. Statistical analysis of spatially homogeneous dynamic agent-based processes using functional time series analysis. Spatial Statistics. 2016;17:199–219.

52. Codling EA, Bearon RN, Thorn GJ. Diffusion about the mean drift location in a biased random walk. Ecology. 2010;91(10):3106–3113.

53. Hyndman RJ, Shang HL. Rainbow plots, bagplots, and boxplots for functional data. Journal of Computational and Graphical Statistics. 2010;19(1).

54. Shang HL, Hyndman RJ. rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data; 2015. Available from: `\http://CRAN.R-project.org/package=rainbow`.

55. Ramsay JO, Hooker G, Graves S. Functional Data Analysis with R and MATLAB. 1st ed. Springer Publishing Company, Incorporated; 2009.

56. Plank MJ, Simpson MJ. Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. Journal of the Royal Society Interface. 2012;9(76):2983–2996.

57. Guedan S, Ruella M, June CH. Emerging cellular therapies for cancer. Annual review of immunology. 2019;37:145–171.

58. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313(5795):1960–1964.

59. van der Woude LL, Gorris MA, Halilovic A, Figdor CG, de Vries IJM. Migrating into the tumor: a roadmap for T cells. Trends in cancer. 2017;3(11):797–808.

60. Cavagna A, Cimarelli A, Giardina I, Orlandi A, Parisi G, Procaccini A, et al. New statistical tools for analyzing the structure of animal groups. Mathematical biosciences. 2008;214(1):32–37.

61. Buhl J, Sword GA, Simpson SJ. Using field data to test locust migratory band collective movement models. Interface Focus. 2012;2(6):757–763.

62. Binder BJ, Simpson MJ. Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. Physical Review E. 2013;88(2):022705.

63. Johnston ST, Simpson MJ, McElwain DS, Binder BJ, Ross JV. Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. Open biology. 2014;4(9):140097.

64. Agnew D, Green J, Brown T, Simpson M, Binder B. Distinguishing between mechanisms of cell aggregation using pair-correlation functions. Journal of theoretical biology. 2014;352:16–23.

65. Binder BJ, Simpson MJ. Spectral analysis of pair-correlation bandwidth: application to cell biology images. Royal Society Open Science. 2015;2(2):140494.

66. Dini S, Binder B, Green J. Understanding interactions between populations: Individual based modelling and quantification using pair correlation functions. Journal of theoretical biology. 2018;439:50–64.

67. De Oliveira AL, Binder BJ. Discrete Manhattan and Chebyshev pair correlation functions in k dimensions. Physical Review E. 2020;102(1):012130.

68. Gavagnin E, Owen JP, Yates CA. Pair correlation functions for identifying spatial correlation in discrete domains. Physical Review E. 2018;97(6):062104.

69. Zhang S, Chong A, Hughes BD. Persistent exclusion processes: Inertia, drift, mixing, and correlation. Physical Review E. 2019;100(4):042415.

70. Johnston ST, Crampin EJ. Corrected pair correlation functions for environments with obstacles. Physical Review E. 2019;99(3):032124.

71. Bearon R, Durham W. A model of strongly biased chemotaxis reveals the trade-offs of different bacterial migration strategies. Mathematical Medicine and Biology: A Journal of the IMA. 2020;37(1):83–116.