

## Design and interpretation of cell trajectory assays

Lucie G. Bowden, Matthew J. Simpson and Ruth E. Baker

*J. R. Soc. Interface* 2013 **10**, 20130630, published 28 August 2013

---

### Supplementary data

["Data Supplement"](#)

[http://rsif.royalsocietypublishing.org/content/suppl/2013/08/22/rsif.2013.0630.DC1.htm](http://rsif.royalsocietypublishing.org/content/suppl/2013/08/22/rsif.2013.0630.DC1.html)  
l

### References

[This article cites 28 articles, 7 of which can be accessed free](#)

<http://rsif.royalsocietypublishing.org/content/10/88/20130630.full.html#ref-list-1>

### Subject collections

Articles on similar topics can be found in the following collections

[biomathematics](#) (270 articles)

[computational biology](#) (260 articles)

[systems biology](#) (146 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

## Research



**Cite this article:** Bowden LG, Simpson MJ, Baker RE. 2013 Design and interpretation of cell trajectory assays. *J R Soc Interface* 10: 20130630.  
<http://dx.doi.org/10.1098/rsif.2013.0630>

Received: 15 July 2013

Accepted: 2 August 2013

### Subject Areas:

systems biology, biomathematics,  
computational biology

### Keywords:

cell migration assay, cell trajectory, confidence,  
random walk model

### Author for correspondence:

Ruth E. Baker

e-mail: [ruth.baker@maths.ox.ac.uk](mailto:ruth.baker@maths.ox.ac.uk)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2013.0630> or via <http://rsif.royalsocietypublishing.org>.

# Design and interpretation of cell trajectory assays

Lucie G. Bowden<sup>1</sup>, Matthew J. Simpson<sup>2,3</sup> and Ruth E. Baker<sup>1</sup>

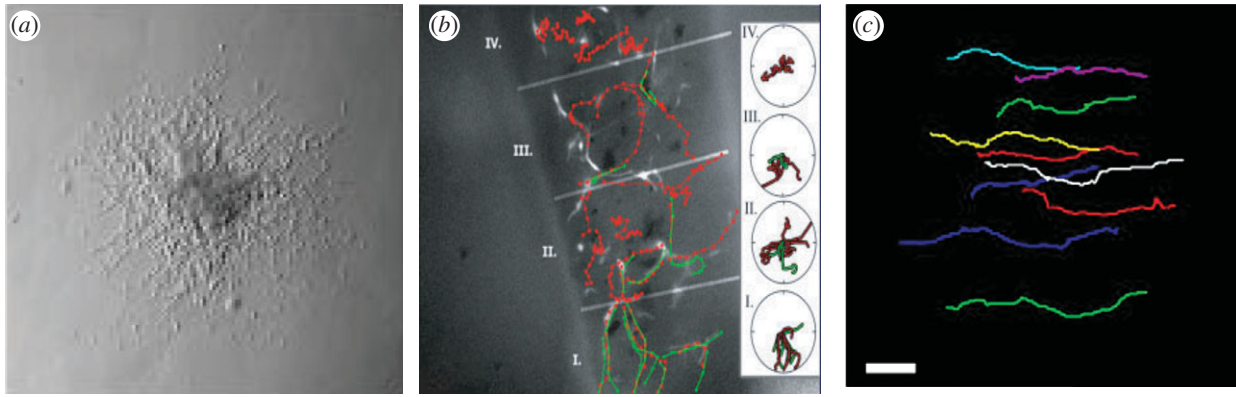
<sup>1</sup>Centre for Mathematical Biology, Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

<sup>2</sup>Mathematical Sciences, and <sup>3</sup>Tissue Repair and Regeneration Program, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4059, Australia

Cell trajectory data are often reported in the experimental cell biology literature to distinguish between different types of cell migration. Unfortunately, there is no accepted protocol for designing or interpreting such experiments and this makes it difficult to quantitatively compare different published data-sets and to understand how changes in experimental design influence our ability to interpret different experiments. Here, we use an individual-based mathematical model to simulate the key features of a cell trajectory experiment. This shows that our ability to correctly interpret trajectory data is extremely sensitive to the geometry and timing of the experiment, the degree of motility bias and the number of experimental replicates. We show that cell trajectory experiments produce data that are most reliable when the experiment is performed in a quasi-one-dimensional geometry with a large number of identically prepared experiments conducted over a relatively short time-interval rather than a few trajectories recorded over particularly long time-intervals.

## 1. Introduction

Cell migration is an essential feature of tissue repair [1,2], development [3,4] and disease [5]. *In vitro* assays are often designed to identify different types of cell migration, such as distinguishing between undirected random motility and directed motility (e.g. chemotaxis). Individual cell-tracking assays, where the trajectories of individual cells are recorded over time [6–13], are often reported as a way to distinguish between different cell migration mechanisms [14,15] or to explore how cell behaviour is affected by different local conditions, such as differences in crowding effects [16]. For example, a snapshot from an *in vitro* migration assay in figure 1*a* shows an explant of neurons studied by Ward *et al.* [17]. The initially circular explant was placed on a substrate for 24 h, after which a source of the protein slit was placed next to the explant. During the following 24 h, the trajectories of a relatively small number of individual cells within the population were recorded, so that the effect of slit could be measured. An image from a similar *in vivo* assay in figure 1*b* shows individual cell trajectories within an invasion wave of cells associated with development of the enteric nervous system. This developmental process involves a population of precursor cells, called neural crest cells, invading the entire length of the intestine as a wave of cells that moves along the developing gut tissues. Individual cells within the population move and proliferate, with the net result being the formation of a wave of cells that advances along the gut tissue. The precise details of the cell migration mechanism, such as whether cells are undirected or directed by chemical signals, are poorly understood [4]. To provide insights into these details, Druckenbrod & Epstein [3] labelled a small number of cells within the population and recorded their trajectories. Their results indicated that cells at the leading edge seemed to follow relatively directed trajectories, whereas cells behind the leading edge followed less directed, more random trajectories [3]. The image in figure 1*c* shows the trajectories of endoderm cells during gastrulation in a zebra-fish embryo. These images indicate that the motion of individual cells is highly directed, and these kinds of trajectories were used by Mizoguchi *et al.* [18] to



**Figure 1.** A suite of cell trajectory assays: (a) an explant of neurons used to examine how slit influences cell migration [17] (reproduced with permission from the Society of Neuroscience); (b) *in vivo* trajectories of neural crest cells moving along the developing intestine [3] (reproduced with permission from John Wiley and Sons); (c) trajectories of endoderm cells during zebrafish gastrulation [18] (reproduced with permission from the Company of Biologists). (Online version in colour.)

study the effect of secreted chemokines. Mizoguchi *et al.*'s experiments showed that normal development was associated with highly directed endoderm cell motion, whereas abnormal development was associated with random motion of endoderm cells [18].

The three experiments in figure 1*a–c* are representative examples of the kinds of broad cell biology systems and research questions that are often explored using cell trajectory data. Although collecting and analysing cell trajectory data can provide insightful information, currently there are no accepted protocols for designing and interpreting such cell trajectory assays. Such a protocol is, however, important because there are many variables to be considered when designing these experiments, and these variables can have a significant impact on the measurements and subsequent interpretation. For example, if we consider the data reported by Ward *et al.* [17] (figure 1*a*), several judgements about how to design and collect these data had to be made. These choices include the following:

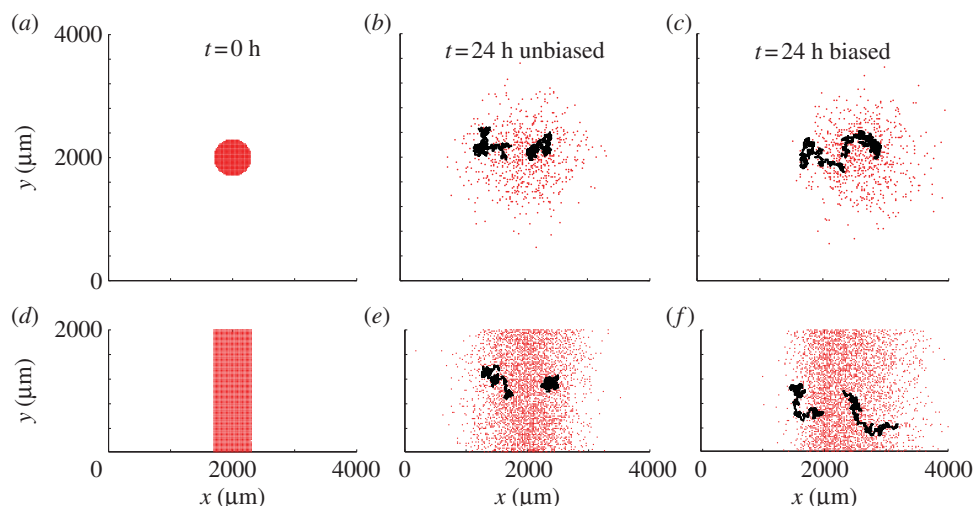
- What duration of time should an individual cell trajectory be measured over?
- Should we measure trajectories originating from the leading edge of the population, or should we focus on cell trajectories originating from the centre of the population?
- If we are measuring the trajectory of a cell in such an explant experiment placed next to a chemical source that is thought to regulate motility, should we focus on cells located close to the chemical signal, or should we focus on cells that originate from a location further away from the chemical source?
- If we are to draw useful quantitative conclusions from cell trajectory data, then how many trajectories must be recorded before reliable conclusions can be drawn about the relevant mechanisms?

There is no current consensus on the appropriate design of cell trajectory experiments. To provide insights into various choices associated with the design of a trajectory experiment, we analyse trajectory data from an individual-based model [19,20] designed to mimic the key features of a two-dimensional cell migration assay. We use a lattice-based exclusion process [21], which incorporates cell-to-cell crowding effects by allowing, at most, one agent to occupy each

site. The experiments are simulated using a two-dimensional square lattice with spacing  $\Delta = 20 \mu\text{m}$  to represent a typical cell diameter [22]. A different lattice structure, such as a hexagonal or unstructured lattice, could be used [19]; however, we take the most straightforward modelling approach and use a square lattice. Each site is indexed  $(i, j)$  and has position  $(x, y) = (i\Delta, j\Delta)$ . Simulations are performed in the following way: for a system with  $N$  cells on the lattice at time  $t$ , during the next time step of duration  $\tau$ ,  $N$  agents are selected independently at random, one at a time. When chosen, an agent attempts to step to a nearest-neighbour site with probability  $P_m \in [0,1]$ . We allow the motion of agents to be biased, so that our model can be used to represent migration assays where the motion of cells can be directed by some external chemical signal [17]. Accordingly, a motile agent at  $(x, y)$  will attempt to step to  $(x, y \pm \Delta)$  with probability  $(1 \pm \rho_y)/4$ , or to  $(x \pm \Delta, y)$  with probability  $(1 \pm \rho_x)/4$  [23].

Using this modelling framework, we can recreate a two-dimensional cell migration assay and examine the associated cell trajectory data. For example, the snapshot in figure 2*a* is a model of a two-dimensional explant with dimensions chosen to replicate Ward *et al.*'s experiments [17] (figure 1*a*). Here, we consider a circular explant of diameter  $400 \mu\text{m}$ , composed of cells that are diameter of  $20 \mu\text{m}$ . Given that a typical cell diffusivity is approximately  $D = 1 \times 10^{-6} \text{mm}^2 \text{s}^{-1}$  [22], we simulate the experiments with  $\Delta = 20 \mu\text{m}$ ,  $P_m = 1$  and  $\tau = 100 \text{s}$ . Snapshots of a single realization of the model after 24 h are given for unbiased ( $\rho_x = \rho_y = 0$ , figure 2*b*) and biased motion ( $\rho_x = 0.5$ ,  $\rho_y = 0$ , figure 2*c*). We consider trajectory experiments over a period of no longer than 24 h to be consistent with Ward *et al.*'s experiments [17], and we note that this time-interval means that we can neglect the influence of cell proliferation in these experiments, because a typical cell doubling time is around 24 h [24,25].

The influence of the motility bias on the spreading populations is obvious in figure 2, because the unbiased spreading population remains radially symmetric (figure 2*b*), whereas the biased spreading population becomes asymmetric (figure 2*c*). We also superimpose two sample trajectories on each snapshot, one trajectory starting at the left-most edge of the explant where  $x = 1800 \mu\text{m}$ , and the other starting on the right-most edge of the explant where  $x = 2200 \mu\text{m}$ . The precise details of the mechanisms driving these single trajectories are difficult to interpret, because all individual trajectories follow very unpredictable paths. Some general trends,



**Figure 2.** A suite of simulated cell trajectory assays: (a) a round explant of diameter 400  $\mu\text{m}$  similar to the explant shown in figure 1a [17]; (b) a snapshot after 24 h for unbiased motility including two trajectories recorded over 24 h; (c) a snapshot after 24 h for biased motility including two trajectories recorded over 24 h. (d–f) A suite of simulated cell trajectory assays in a quasi-one-dimensional geometry where the initial condition is independent of the vertical coordinate. The quasi-one-dimensional simulations show (d) an initial strip explant, of width 400  $\mu\text{m}$ ; (e) a snapshot after 24 h for unbiased motility including two trajectories recorded over 24 h; (f) a snapshot after 24 h for biased motility including two trajectories recorded over 24 h. Results correspond to  $P_m = 1$ ,  $\Delta = 20 \mu\text{m}$ ,  $\tau = 100 \text{ s}$ , cells are indicated by red (grey) discs and trajectories are shown in black. (b,e) correspond to  $\rho_x = \rho_y = 0$ , (c,f) correspond to  $\rho_x = 0.5$  and  $\rho_y = 0$ . (Online version in colour.)

however, are discernable. The trajectories in figure 2b appear to be directed away from the location of the explant, because the left-most trajectory appears to move in the negative  $x$ -direction, whereas the right-most trajectory appears to move in the positive  $x$ -direction. This observation is consistent with previous analysis of the random walk model, and this drift is caused by cell-to-cell crowding effects [23]. This observation illustrates that the initial position of the trajectory, relative to the bulk population, is an important consideration in the design and interpretation of cell trajectory assays.

Our aim in this work is to analyse the design and interpretation of cell trajectory experiments. To achieve this, we study trajectory data from a discrete stochastic exclusion process model, allowing us to examine both the motion of the bulk population and the motion of individual trajectories within the bulk population. We choose to study an exclusion process instead of a standard non-interacting random walk [26], because the exclusion process explicitly accounts for crowding effects induced by the finite size of cells [27,28]. These models have been used previously to study cell migration in several contexts such as the collective migration of glioma cells [27,29], breast cancer cells [25] and fibroblast cells [2,30]. While some cell trajectory assays are performed on a two-dimensional substrate (e.g. figure 2a–c), we also consider experiments with a special initial condition where the initial density of cells is independent of the vertical location in the domain (e.g. figure 2d–f), and either periodic or reflecting boundary conditions are applied on both horizontal boundaries of the domain. Under these quasi-one-dimensional conditions, the cell density remains independent of vertical location for all  $t > 0$  [23]. Such quasi-one-dimensional experiments are relevant when considering cell migration assays performed in a narrow channel geometry, for example scratch assays [1,29]. The modelling data in figure 2d–f are identical to those in figure 2a–c except that we now consider the initial explant to be a column of cells rather than be a circular explant. This kind of simplification allows us only to examine properties of cell trajectory assays as a function of the horizontal

coordinate. Once we have completely characterized this kind of experiment in the simplified geometry, we will examine the more detailed, genuinely two-dimensional counterpart.

## 2. Modelling methods

### 2.1. Discrete cell migration model

The discrete model, described briefly in §1, is a biased exclusion process that has been described previously [23]. The model treats a population of cells as a system of discrete uniformly sized agents, and allows each cell the opportunity to undergo motility events to simulate cell migration with probability  $P_m$  per time step. The motion of agents can be unbiased ( $\rho_x = \rho_y = 0$ ) or biased ( $\rho_x \neq 0$  and/or  $\rho_y \neq 0$ ). Unlike standard random walk models that treat agents as point particles without any volume [26], our model explicitly incorporates crowding effects, because potential motility events are permitted only if the target site is vacant [23,24,27].

Previous work on this discrete model has focused on the relationship between averaged data from a large number of identically prepared realizations of the discrete model and the solution of relevant continuum descriptions [23,24,27]. Here, we take a different approach, because our focus is to understand and quantify how our ability to correctly interpret the cell migration mechanism using a limited, more realistic, number of observations from the discrete model.

### 2.2. Continuum description

Averaged agent (or cell) density data from the discrete simulations are related to a partial differential equation (PDE) description [23,24,27]. If the average occupancy of site  $(i, j)$  in the  $k$ th realization is  $\langle C_{i,j} \rangle = (1/M) \sum_{k=1}^M C_{i,j}^k$ , then, as  $M \rightarrow \infty$ , the average agent density is governed by the equation

$$\frac{\partial C}{\partial t} = D \nabla^2 C - \nabla \cdot [VC(1 - C)], \quad (2.1)$$



where  $C(x, y, t)$  is the population density,  $D$  is the cell diffusivity and  $V = (v_x, v_y)$  is a drift velocity reflecting the presence of motility bias with

$$v_x = \lim_{\Delta, \tau \rightarrow 0} \left( \frac{P_m \rho_x \Delta}{2\tau} \right), \quad v_y = \lim_{\Delta, \tau \rightarrow 0} \left( \frac{P_m \rho_y \Delta}{2\tau} \right) \quad (2.2)$$

and  $D = \lim_{\Delta, \tau \rightarrow 0} \left( \frac{P_m \Delta^2}{4\tau} \right).$

The relationship between the averaged discrete data and the solution of equation (2.1) was analysed and discussed, in detail, previously [24,25].

We also consider the averaged behaviour of cell trajectories within the population [23]. If the coordinates of a tagged agent in the  $k$ th realization are  $x(t)^k$  and  $y(t)^k$ , then the averaged position coordinates are  $\langle x(t) \rangle = (1/M) \sum_{k=1}^M x(t)^k$  and  $\langle y(t) \rangle = (1/M) \sum_{k=1}^M y(t)^k$ . Then, considering  $M \rightarrow \infty$ , the average coordinates of this trajectory are governed by [23]

$$\frac{dp_x}{dt} = -2D \frac{\partial C}{\partial x} + v_x(1 - C) \quad (2.3)$$

and  $\frac{dp_y}{dt} = -2D \frac{\partial C}{\partial y} + v_y(1 - C).$

If a tagged agent is initially at  $(p_x(0), p_y(0))$ , then, on average, it will follow the trajectory given by  $(p_x(t), p_y(t))$  for  $t > 0$ . To explore the average behaviour of the system, we solve equation (2.1) using a finite difference approximation with discretization  $\delta x = \delta y = \delta$ . Temporal integration is performed using Crank–Nicolson time integration with time step  $\delta t$  [31]. The nonlinear algebraic equations are solved using Picard iteration. The solution of equation (2.3) is approximated numerically using techniques outlined previously [23].

These continuum descriptions, equations (2.1) and (2.3), describe averaged simulation data,  $\langle C_{i,j} \rangle$ ,  $\langle x(t) \rangle$  and  $\langle y(t) \rangle$ , under highly idealized conditions where  $M \rightarrow \infty$  [23]. Our aim is to understand and quantify how well we can make inferences from more realistic experimental conditions where we have access to a modest number of observations. We note that it is convenient to have a mathematical description of the idealized case,  $M \rightarrow \infty$ , so that we can quantify how our confidence in our ability to correctly interpret finite amounts of trajectory data varies as we change the design of the experiment.

## 2.3. Confidence intervals

Under typical experimental conditions, cell trajectories are collected using a relatively small number of experimental realizations [6–13]. Therefore, to quantify the implications of having access to limited data, we calculate confidence intervals for samples of trajectory data from our discrete model using a realistic number of realizations. We make the standard assumption that the trajectory data samples are normally distributed about the mean [32] (see electronic supplementary material). A confidence interval for the location of the mean trajectory  $\langle x(t) \rangle$ , with  $\bar{x} = \lim_{M \rightarrow \infty} \langle x(t) \rangle$ , is given by [32]

$$\left[ \langle x(t) \rangle - z_{\alpha/2} \frac{\sigma}{\sqrt{M}} \leq \bar{x} \leq \langle x(t) \rangle + z_{\alpha/2} \frac{\sigma}{\sqrt{M}} \right],$$

where  $z_{\alpha/2}$  is the inverse of the normal distribution,  $\sigma$  is the standard deviation of the distribution and  $M$  is the sample size. Rearranging gives  $z_{\alpha/2} \sigma / \sqrt{M} = \beta$ , where  $2\beta$  is the interval about which we wish to estimate the confidence in

our observations. This gives  $\alpha = 2 \exp(-[z_{\alpha/2}]^2/2)$  and a confidence interval of [32]

$$\text{Confidence} = 100(1 - \alpha)\%.$$

This definition allows us to quantitatively examine how our confidence in our partial observations reflects the true cell migration mechanism as we systematically vary the design of the trajectory experiment.

## 2.4. Parameter fitting

To complement our confidence interval results, we also quantify our ability to interpret cell trajectory data with a relatively small number of experimental realizations using parameter fitting. To do this, we simulate trajectory data with known values of  $P_m$ ,  $\tau$  and  $\rho_x$  (or, equivalently, known values of  $D$  and  $v_x$ ) for a range of experimental conditions and then use a standard parameter estimation algorithm to calibrate the solution of the continuum model, equations (2.1) and (2.3), to that data. This allows us to measure how estimates of  $P_m$  and  $\rho_x$  depend on the details of the experimental design. We use Matlab's LSQNONLIN routine [33] to find least-squares estimates of  $P_m$  and  $\rho_x$  to match the simulated trajectories. If  $\langle x(t) \rangle$  denotes the mean pathline of sample size  $M$ , then  $p_x(t)$  is the solution of equation (2.3) and  $t_k = 0, 1, 2, \dots, T$  is the set of time points over which we perform the fitting, we define the mean squared error (m.s.e.) to be

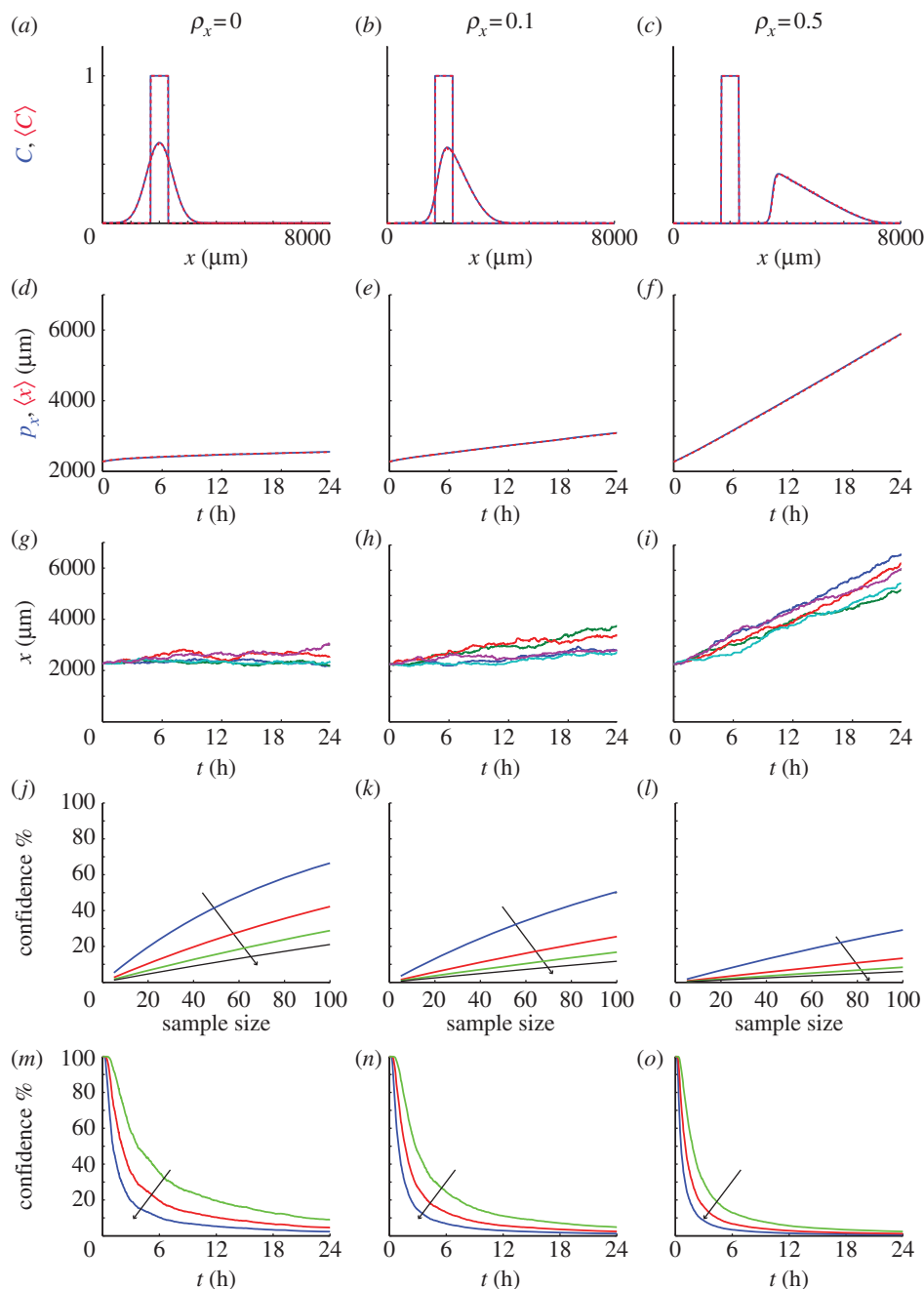
$$\text{m.s.e.} = \frac{1}{T} \sum_{k=0}^T \left[ \frac{\langle x(t_k) \rangle - p_x(t_k)}{\langle x(t_0) \rangle - \langle x(t_T) \rangle} \right]^2, \quad (2.4)$$

where we have normalized by the length of  $\langle x(t) \rangle$ . Using this definition we use Matlab's LSQNONLIN to find the estimates of  $P_m$  and  $\rho_x$  such that the solution of equation (2.3) minimizes the m.s.e.

## 3. Results and discussion

### 3.1. Confidence in one-dimension over one cell diameter

To address the four questions regarding the design and interpretation of cell trajectory experiments that were outlined previously in the Introduction, we present a suite of data in figure 3 for the quasi-one-dimensional initial condition described previously in figure 2d–f. This corresponds to an explant of width 400  $\mu\text{m}$ . We consider different amounts of bias,  $\rho_x = 0, 0.1$  and  $0.5$ , and we present averaged density profiles superimposed on the relevant solution of equation (2.1) in figure 3a–c, which shows that the density spreads symmetrically for  $\rho_x = 0$ , whereas the density drifts in the positive  $x$ -direction for  $\rho_x > 0$ . Comparing the averaged density data and the solution of equation (2.1) confirms that the PDE model predicts the averaged behaviour. Results in figure 3d–f compare averaged trajectory data from a very large number ( $M = 5000$ ) of identically prepared discrete simulations for a tagged cell initially located at the right-most edge of the explant,  $x(0) = 2200 \mu\text{m}$ . We note that the average trajectory in figure 3d drifts in the positive  $x$ -direction owing to crowding effects [23]. For all values of  $\rho_x$ , the averaged trajectory is superimposed on the solution of equation (2.3), confirming that the averaged data and the solution of the continuum description compare very well, provided



**Figure 3.** (a–c) Comparison of averaged agent density information from the discrete simulations (blue (dark grey), solid) for the same initial condition shown in figure 2d superimposed on the corresponding solution of equation (2.1), (red (grey), dashed) at  $t = 24$  h for different amounts of bias with  $P_m = 1$ ,  $\Delta = 20$   $\mu\text{m}$  and  $\tau = 100$  s. Average density profiles are constructed using  $M = 5000$  realizations. (d–f) Comparison of averaged trajectories from the simulations in (a–c) (blue (dark grey), solid) where the initial location of the tagged cells is at the right-most edge of the population,  $x(0) = 2200$   $\mu\text{m}$ , is superimposed on the corresponding solution of equation (2.3) (red (grey), dashed) for different amounts of bias. Five sample trajectories from the results used to construct the average trajectories in (d–f) are shown in (g–i). Results in (j–l) show how the confidence varies when we identify the true average trajectory to within an interval of  $\pm 1$  cell diameter as a function of the number of identically prepared trajectories considered. Confidence results are given at  $t = 6, 12, 18$  and  $24$  h with the arrows showing the direction of increasing  $t$ . Results in (m–o) show how the confidence decreases with  $t$  when we consider identifying the average trajectory to within  $\pm 1$  cell diameter with  $M = 10, 20$  and  $40$  experimental realizations with the arrows showing the direction of decreasing  $M$ . (Online version in colour.)

that we have access to a very large number of identically prepared realizations of the experiments.

The trajectory results in figure 3d–f were obtained using an impractically large number of identically prepared discrete simulations ( $M = 5000$ ). Instead, we wish to analyse how our confidence in the properties of the observed trajectories is influenced by having limited amounts of experimental data. To illustrate the stochasticity in the trajectory data, we show in figure 3g–i trajectories from five such realizations of the results in figure 3d–f. Comparing the individual realizations with the averaged data confirms that the individual trajectories are

relatively noisy, and it is difficult to draw definitive conclusions from these limited observations. For example, we expect that the trajectories in figure 3g, for unbiased motility ( $\rho_x = 0$ ), to drift in the positive  $x$ -direction owing to crowding effects in the stochastic model [23]. However, close inspection of the five realizations in figure 3g indicates that some trajectories drift in the positive  $x$ -direction, whereas others drift in the negative  $x$ -direction for some periods of time. This means that limited amounts of trajectory data could lead us to make incorrect conclusions about the cell migration mechanism acting in this system. All results in figure 3g–i indicate

that the observed spread about the mean trajectory increases with time. This suggests that the design of cell trajectory assays needs to consider a trade-off between allowing a sufficient amount of time to observe meaningful trajectories while ensuring that the timescale of the experiment is not too long that the trajectory data become too noisy to draw useful conclusions.

To explore the trade-off between the timescale of the experiment, the number of cell trajectories and our confidence in the results, we present confidence interval data at  $t = 6, 12, 18$  and  $24$  h in figure 3*j–l*, estimated for the same problem considered in figure 3*d–f*. These confidence intervals are calculated using a strict observation window where we require that the observations be correct to within  $\pm 1$  cell diameter about the mean, corresponding to  $\beta = 1$  in the confidence interval calculation. Our results indicate that we have a remarkable reduction in confidence as we decrease the number of trajectories as well as a remarkable reduction in confidence as we increase the timescale of the experiment. In particular, we observe a very low confidence, less than 10%, for experiments with  $M = 10$  trajectories, regardless of the timescale of the experiment. This is important, because many experimental observations are reported for relatively low numbers of trajectories [3,18]. Instead, our modelling indicates that we need to use around 100 trajectories to have significant confidence in our results. Comparing results for different amounts of bias indicates that we see a remarkable reduction in the confidence for biased data. In summary, optimal results correspond to unbiased trajectories collected over relatively short timescales. However, even under these highly idealized conditions, we need to collect very large numbers of trajectories to be able to draw reliable conclusions.

To explore the relationship between confidence, the timescale of the experiment and the number of experimental replicates, we present a final set of results in figure 3*m–o* showing how the confidence decreases as a function of time for different experiments where different numbers of trajectories are considered. These results confirm that on very short timescales we are very confident in our experimental observations and that our confidence decays very quickly with time. Again, we observe that the decay in confidence with time occurs fastest for the more biased motion.

All results presented in figure 3 correspond to cell trajectories originating from the right-most edge of the explant. Intuitively, we might expect that our results will depend on the initial location of the tagged cell, and, for completeness, we present equivalent results illustrating how predictions vary as a function of the initial location by considering tracking a cell at the centre (figure 4) and the left-most edge (figure 5) of the explant. These additional results confirm that the details of the trajectory data depend on the initial location of the tagged cell. However, regardless of these details, we still observe that our confidence decays rapidly with time, as the number of observations decreases and with increasing motility bias.

### 3.2. Parameter fitting

All results in figures 3–5 quantify our ability to infer the cell migration mechanism in a cell trajectory experiment using a measure of confidence. We can also quantify our ability to infer the cell migration mechanism using parameter fitting. To illustrate this approach, we consider 12 different

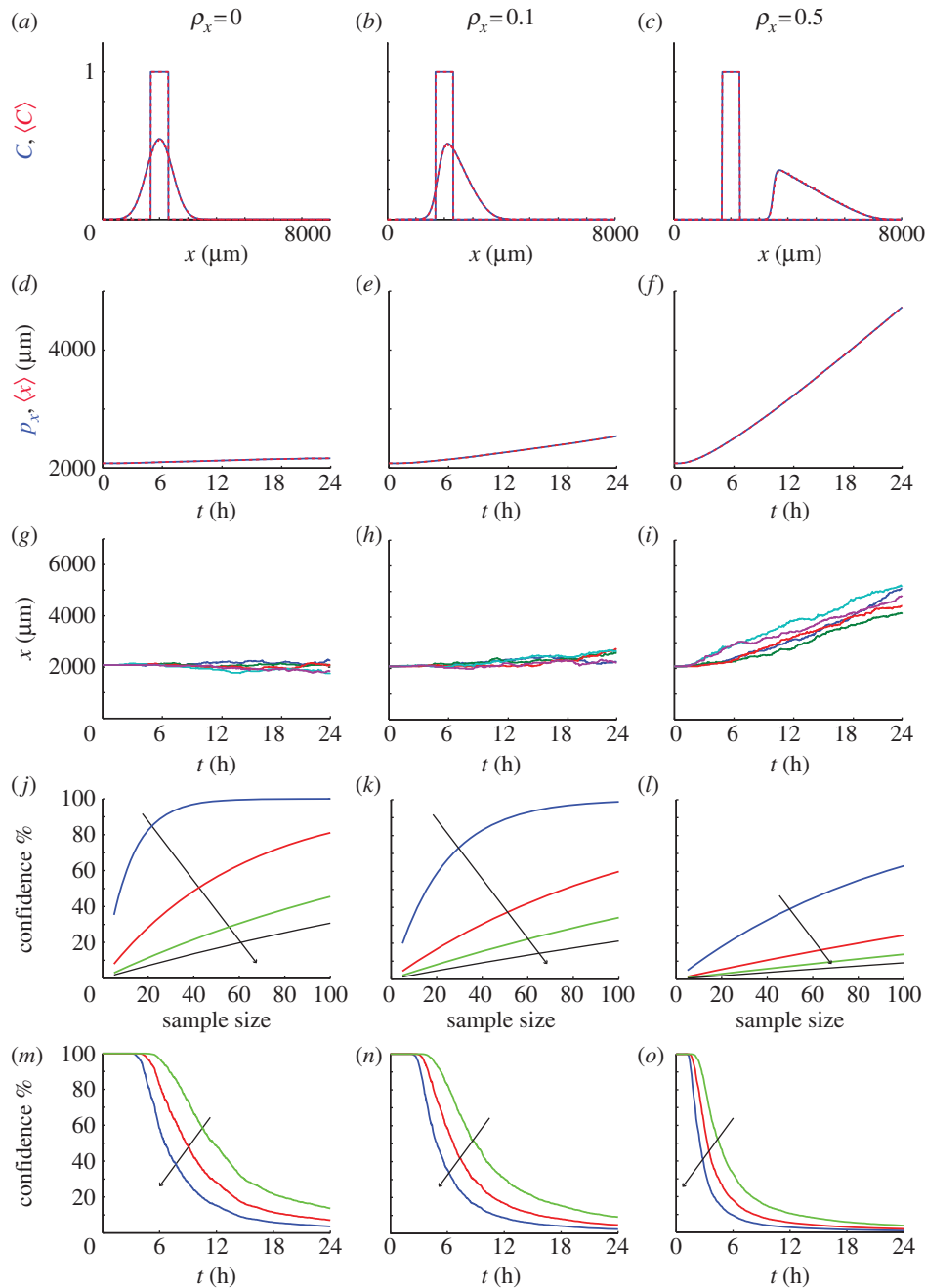
experiments corresponding to three different amounts of bias ( $\rho_x = 0, 0.1, 0.5$ ) and four different time points ( $t = 6, 12, 18, 24$  h), each with  $P_m = 1$  and  $\tau = 100$  s. For each experimental condition, we considered 121 equally spaced values of  $P_m$  in the range  $P_m \in [0, 1.2]$ , and 81 equally spaced values of  $\rho_x$  in the range  $\rho_x \in [-0.2, 0.6]$ . For each of the  $121 \times 81$  parameter combinations, we performed an impractically large number of realizations ( $M = 5000$ ) to quantify the m.s.e. according to equation (2.4). These data are reported in figure 6 as a series of error surfaces showing the m.s.e. as a function of  $P_m$  and  $\rho_x$ . The error surfaces display some revealing trends. First, none of these error surfaces contains a single, well-defined minimum, which suggests that there are many parameter combinations that give a good match to our observations, regardless of the amount of experimental data. Second, as  $\rho_x$  increases, the ‘shallow’ region of the error surface surrounding the true minimum becomes larger. This indicates that our ability to recover the true parameters becomes increasingly difficult as  $\rho_x$  increases.

To explore our ability to fit parameters using realistic numbers of experimental replicates, we used Matlab’s `LSQNONLIN` routine [33] to obtain least-squares estimates of  $P_m$  and  $\rho_x$  for trajectory data averaged over just  $M = 10, 20$  and  $40$  samples. The results of this fitting exercise are superimposed on the error surfaces in figure 6, where we see that such parameter estimates can give misleading results. For example, results in figure 6*a* for  $t = 6$  h with true values of  $(P_m, \rho_x) = (1.0, 0)$  indicate that we estimate  $(P_m, \rho_x) \approx (1.2, -0.1), (0.81, -0.02)$  and  $(0.97, 0.01)$  with  $M = 10, 20$  and  $40$ , respectively. These parameter fitting results are consistent with our confidence interval results in figure 3 where we observed that our confidence decreased rapidly as  $M$  decreased. The same trends are observed in other subfigures in figure 6 for different amounts of bias and different experimental durations.

Data in figure 6 also indicate that the amount of bias present in the experiment has an influence on our ability to recover parameters from limited observations. For example, the least-squares parameter estimates in the left column of figure 6 with  $M = 40$  are consistently closer to the expected minimum compared with our least-squares estimates with  $M = 40$  in the middle and right columns, for larger values of  $\rho_x$ . These trends are also consistent with our confidence interval results in figure 3*j–o* where we saw that, in general, our estimates of confidence decreased as  $\rho_x$  increased. In summary, we find that the trends in figures 3 and 6 confirm that we can best infer the mechanisms from unbiased trajectory data collected over relatively short time periods with a large number of replicates. Conversely, the trends in figures 3 and 6 indicate that we are least able to infer the mechanisms from biased trajectory data collected over long time periods with a small number of replicates. Because our parameter estimation approach yields results that are consistent with our confidence data, we shall focus only on the confidence results for the remainder of this work.

### 3.3. Confidence in one dimension over variable cell diameters

Results in figures 3–5 indicate that our confidence in observed trajectory data is very low for typical conditions where modest numbers of trajectories are recorded over typical experimental timescales [6–13]. In particular, for the best-case scenario where the system is characterized by



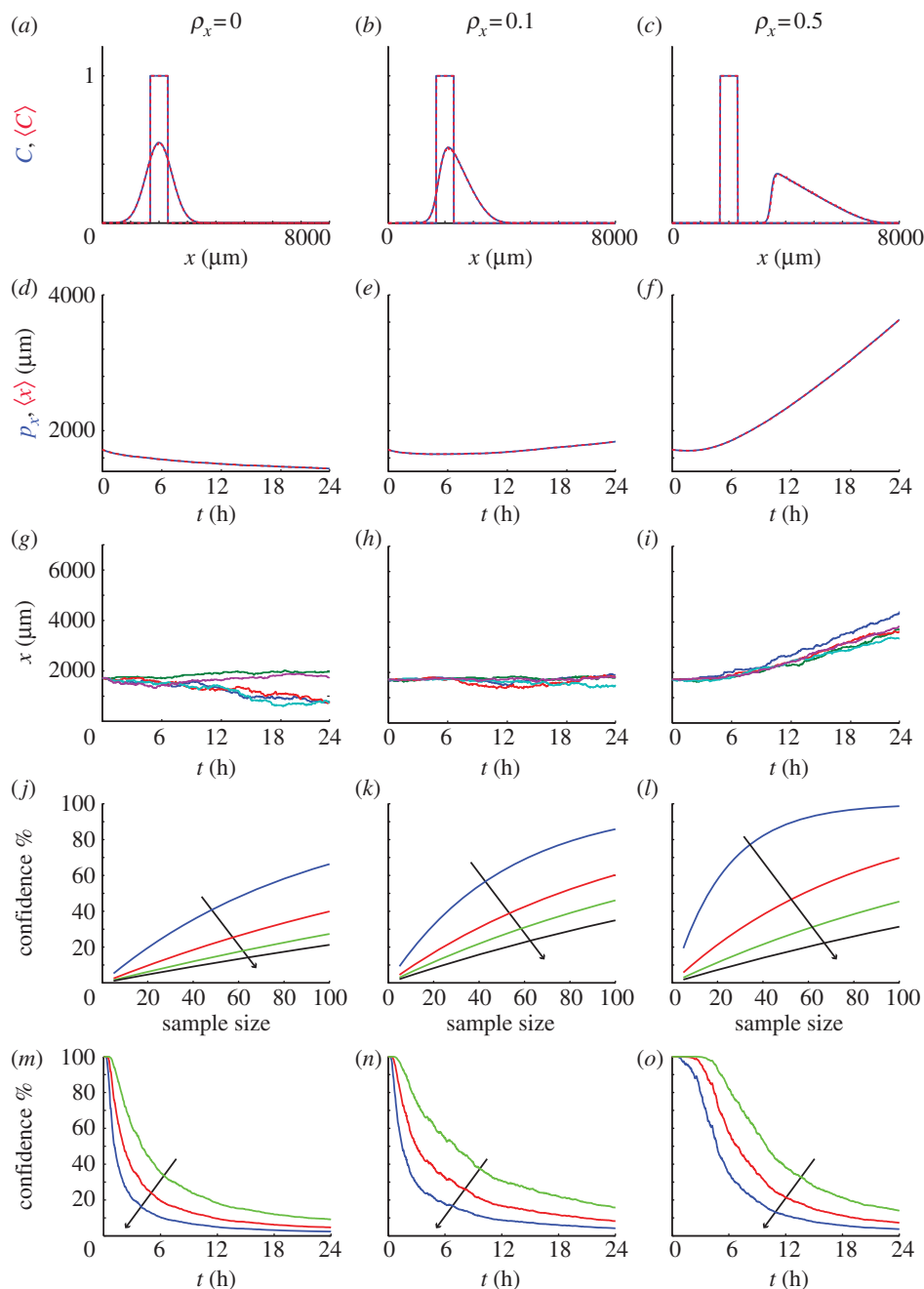
**Figure 4.** (a–c) Comparison of averaged agent density information from the discrete simulations (blue (dark grey), solid) for the same initial condition shown in figure 2d superimposed on the corresponding solution of equation (2.1) (red (grey), dashed) at  $t = 24$  h for different amounts of bias with  $P_m = 1$ ,  $\Delta = 20 \mu\text{m}$  and  $\tau = 100$  s. Average density profiles are constructed using  $M = 5000$  realizations. (d–f) Comparison of averaged trajectories from the simulations in (a–c) (blue (dark grey), solid) where the initial location of the tagged cells is at the centre of the population,  $x(0) = 2000 \mu\text{m}$ , is superimposed on the corresponding solution of equation (2.3) (red (grey), dashed) for different amounts of bias. Five sample trajectories from the results used to construct the average trajectories in (d–f) are shown in (g–i). Results in (j–l) show how the confidence varies when we identify the true average trajectory to within an interval of  $\pm 1$  cell diameter as a function of the number of identically prepared trajectories considered. Confidence results are given at  $t = 6, 12, 18$  and  $24$  h with the arrows showing the direction of increasing  $t$ . Results in (m–o) show how the confidence decreases with  $t$  when we consider identifying the average trajectory to within  $\pm 1$  cell diameter with  $M = 10, 20$  and  $40$  experimental realizations with the arrows showing the direction of decreasing  $M$ . (Online version in colour.)

unbiased motility and we have access to 100 identically prepared trajectories, we never observe any confidence intervals greater than 65%. We now seek to relax the conditions imposed in figures 3–5 by exploring the sensitivity of our confidence results as a function of the desired interval about the mean. If, say, we consider interpreting trajectory data from figure 3 within a window of  $\pm 5$  cell diameters instead of  $\pm 1$  cell diameters, the results in figure 7a,b indicate that our confidence in the trajectory data increases significantly. Similarly, the results in figure 7c indicate that if we

are satisfied with a sample interval of  $\pm 10$  cell diameters, then our confidence increases towards 100% quite rapidly.

The results in figure 7d–f indicate that if we consider the trajectory data from figure 3 and we relax the sample interval from  $\beta = 1$  to  $\beta = 5$  or  $\beta = 10$ , we observe a much slower decay in the confidence over time. These results indicate that it is infeasible for us to expect to make very accurate results over typical timescales and using typical numbers of experimental trajectories if we expect to be confident within  $\pm 1$  cell diameter, whereas it is more realistic to expect to





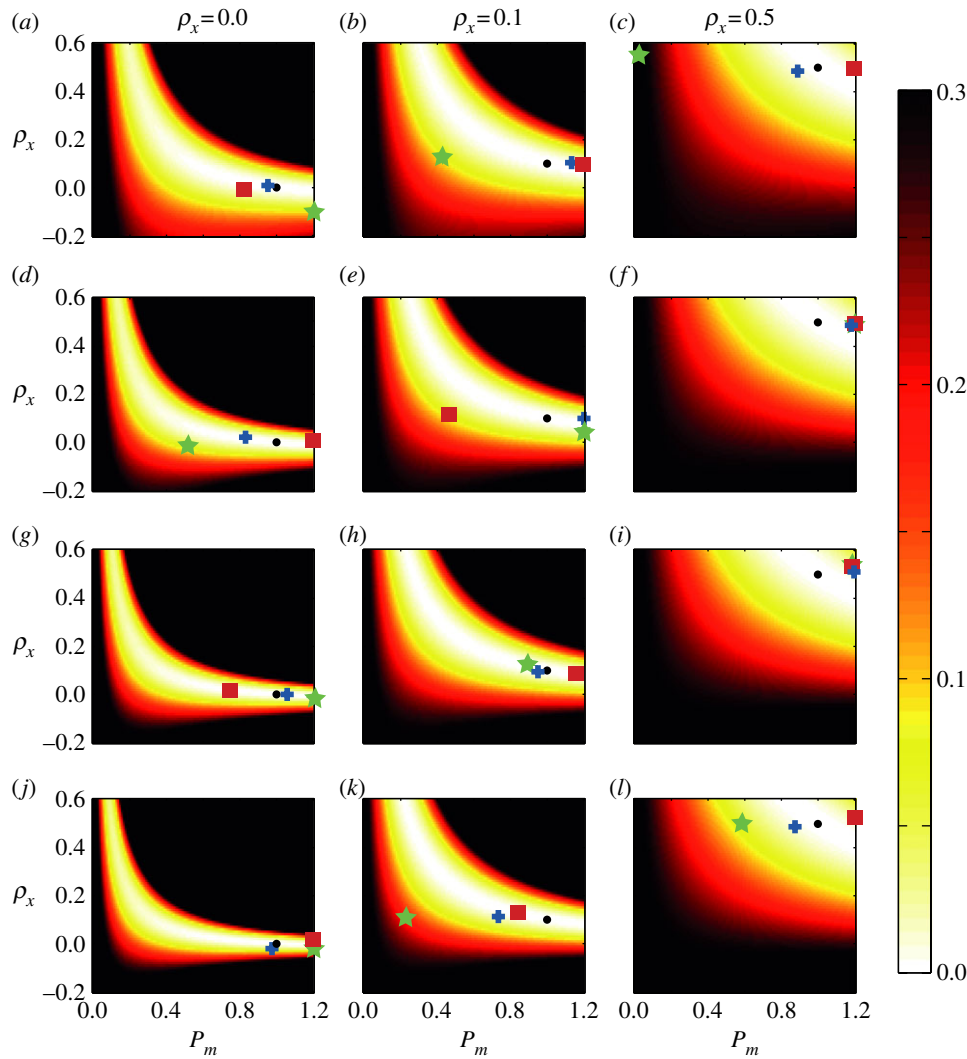
**Figure 5.** (a–c) Comparison of averaged agent density information from the discrete simulations (blue (dark grey), solid) for the same initial condition shown in figure 2d superimposed on the corresponding solution of equation (2.1) (red (grey), dashed) at  $t = 24$  h for different amounts of bias with  $P_m = 1$ ,  $\Delta = 20 \mu\text{m}$  and  $\tau = 100$  s. Average density profiles are constructed using  $M = 5000$  realizations. (d–f) Comparison of averaged trajectories from the simulations in (a–c) (blue (dark grey), solid) where the initial location of the tagged cells is at the left-most edge of the population,  $x(0) = 1800 \mu\text{m}$ , is superimposed on the corresponding solution of equation (2.3) (red (grey), dashed) for different amounts of bias. Five sample trajectories from the results used to construct the average trajectories in (d–f) are shown in (g–i). Results in (j–l) show how the confidence varies when we identify the true average trajectory to within an interval of  $\pm 1$  cell diameter as a function of the number of identically prepared trajectories considered. Confidence results are given at  $t = 6, 12, 18$  and  $24$  h with the arrows showing the direction of increasing  $t$ . Results in (m–o) show how the confidence decreases with  $t$  when we consider identifying the average trajectory to within  $\pm 1$  cell diameter with  $M = 10, 20$  and  $40$  experimental realizations with the arrows showing the direction of decreasing  $M$ . (Online version in colour.)

be able to make relatively confident predictions using limited data over typical timescales if we only expect to be confident within a larger window.

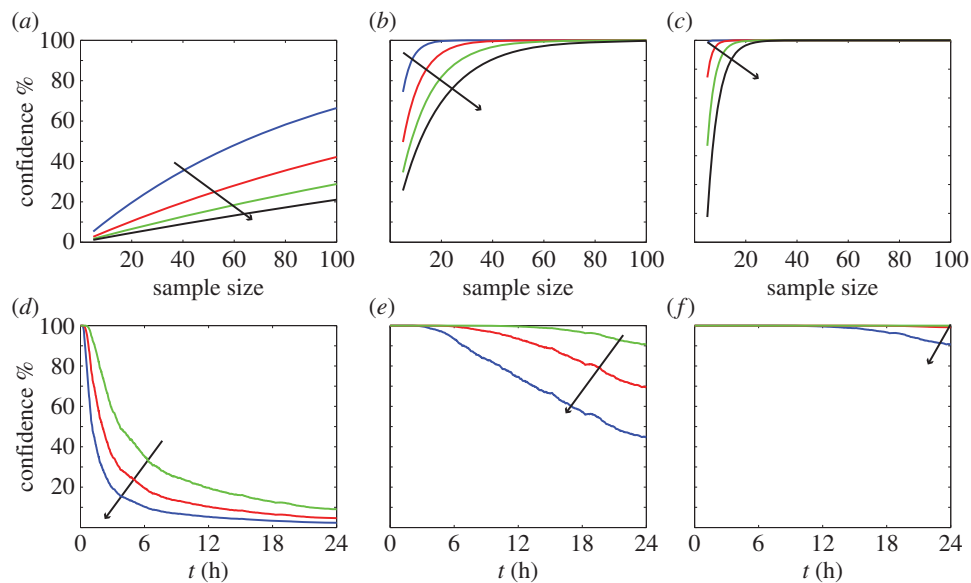
### 3.4. Confidence in two-dimensions

We now analyse trajectory data from a two-dimensional explant equivalent to the snapshots shown in figure 2a–c. As in figure 2, we consider a two-dimensional explant of diameter  $400 \mu\text{m}$ . Results in figure 8a–c show the averaged cell density profile and the corresponding solution of equation (2.1) at  $t = 12$  h

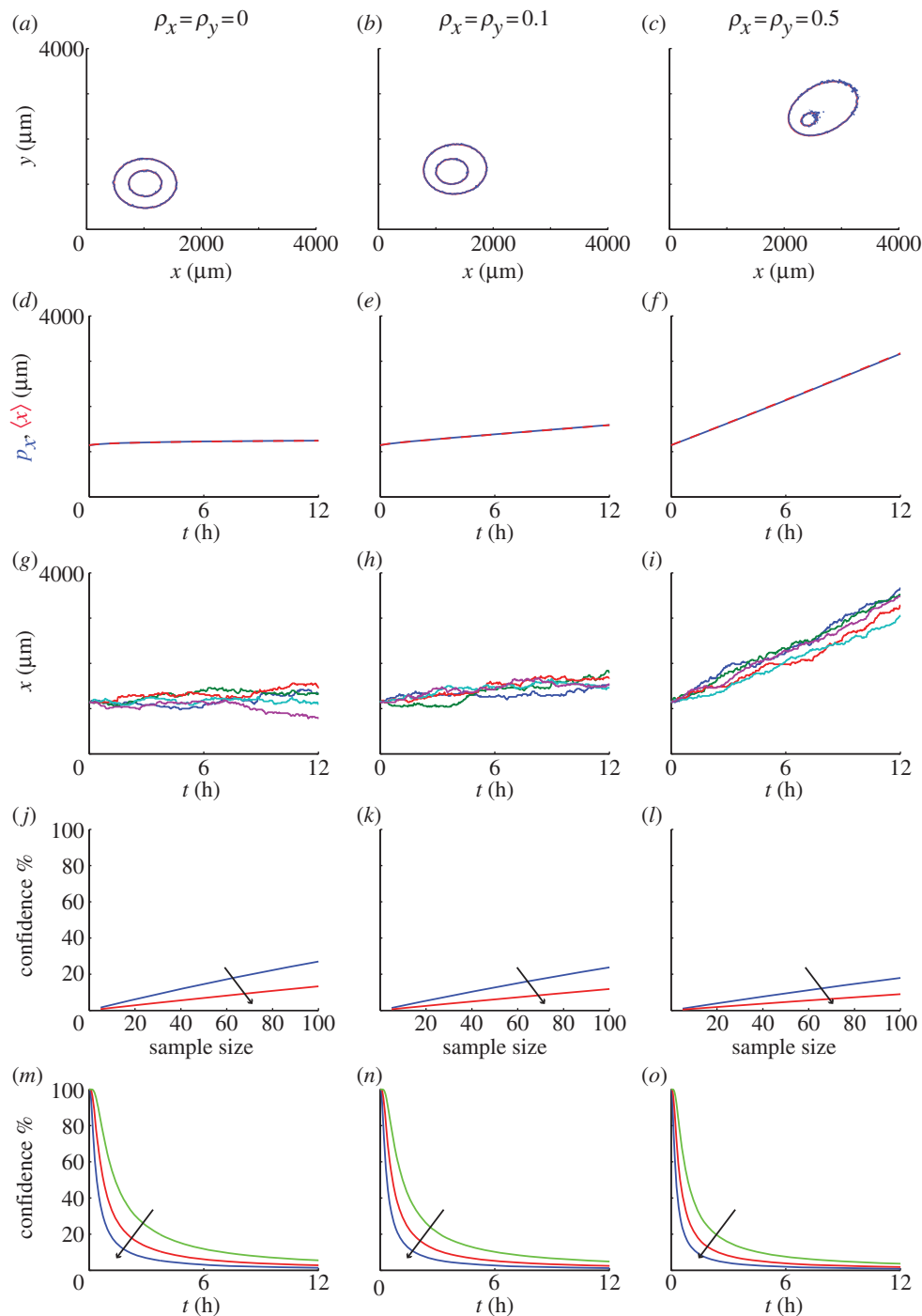
for different amounts of bias:  $\rho_x = \rho_y = 0$ ,  $\rho_x = \rho_y = 0.1$  and  $\rho_x = \rho_y = 0.5$ , respectively. This comparison confirms that equation (2.1) accurately describes the temporal evolution of the cell density profile in such a two-dimensional explant. Within these simulations, we tag a cell at the leading edge of the explant, with initial location  $(x(0), y(0)) = (1140, 1140 \mu\text{m})$ , and we show the averaged  $x$ -coordinate of the cell trajectory together with the solution of equation (2.3) in figure 8d–f. We have omitted to show the details of the  $y$ -coordinate of the trajectory because, owing to symmetry, the cell trajectory information in the  $y$ -direction is equivalent to the  $x$ -direction.



**Figure 6.** Parameter fitting results for the cell-tracking experiment shown in figure 3. Each subfigure shows a surface of the m.s.e., given by equation (2.4), for cell-tracking experiments corresponding to  $P_m = 1$ , various  $\rho_x$  as indicated, and for different experimental durations: (a–c)  $t = 6$  h, (d–f)  $t = 12$  h, (g–i)  $t = 18$  h and (j–l)  $t = 24$  h. The m.s.e. surface was constructed using  $M = 5000$  identically prepared realizations, and the black dot on each subfigure shows the expected minimum of the m.s.e. Least-squares parameter estimates were obtained and superimposed on each subfigure using different sample sizes,  $M = 10, 20$  and  $40$ , and these estimates are indicated using a star (green), square (red) and cross (blue), respectively. The colour bar used to define the m.s.e. surface is given on the right. (Online version in colour.)



**Figure 7.** Results in (a–c) are equivalent to the confidence results in figure 3j except that we estimate the confidence within an interval about the mean  $\pm \beta$  with (a)  $\beta = 1$ , (b)  $\beta = 5$  and (c)  $\beta = 10$ . The arrows show the direction of increasing time ( $t = 6, 12, 18$  and  $24$  h). Results in (d–f) are equivalent to the confidence results in figure 3m except that we estimate the confidence within an interval about the mean  $\pm \beta$ , with (d)  $\beta = 1$ , (e)  $\beta = 5$  and (f)  $\beta = 10$ . The arrows show the direction of decreasing  $M$  ( $M = 40, 20, 10$ ). (Online version in colour.)



**Figure 8.** (a–c) Comparison of averaged discrete agent density profiles (blue (dark grey), solid) at  $t = 12$  h using  $M = 5000$  and the corresponding solution of equation (2.1) (red (grey), dashed) for different amounts of bias with  $P_m = 1$ ,  $\Delta = 20 \mu\text{m}$  and  $\tau = 100$  s. The inner and outer contours correspond to  $C(x, y, t) = 0.15$  and  $0.05$ , respectively. (d–f) Comparison of averaged trajectories from the simulations in (a–c) (blue (dark grey), solid) where the initial location of the tagged cells is at the edge of the population at  $(x, y) = (1140, 1140 \mu\text{m})$  and the solution of equation (2.3) (red (grey), dashed). Five sample trajectories from the results used to construct the average trajectories in (d–f) are shown in (g–i). Results in (j–l) show how the confidence varies when we identify the true average trajectory to within an interval of  $\pm 1$  cell diameter as a function of  $M$ . Confidence results are given at  $t = 6$  and  $12$  h with the arrows showing the direction of increasing  $t$ . Results in (m–o) show how the confidence decreases with  $t$  when we consider identifying the average trajectory to within  $\pm 1$  cell diameter with  $M = 10, 20$  and  $40$  experimental replicates, with the arrows show the direction of decreasing  $M$ . (Online version in colour.)

Comparing the averaged simulation results and the solution of equation (2.3) in figure 8d–f confirms that the continuous description accurately describes the average behaviour of the discrete model for sufficiently large  $M$ .

Results in figure 8g–i show five sample trajectories corresponding to the trajectories used to construct the average cell trajectories in figure 8d–f. Similar to the results in figures 3–5, for the two-dimensional problem, we see that the individual trajectories are noisy, with some individual trajectories

displaying behaviour that is quite different to the average results. To quantify our ability to predict the expected mean behaviour of the system using a limited number of experimental trajectories, we present confidence intervals in figure 8j–l at  $t = 6$  and  $12$  h, as a function of the number of experimental realizations,  $M$ . We observe the same qualitative trends in the two-dimensional confidence interval data that was observed previously for the quasi-one-dimensional results, namely that our confidence decreases with decreasing

$M$ , with increasing time and with increasing motility bias. Importantly, when we compare our confidence intervals in figure 8j–l with results in figure 3, we see that the confidence in the two-dimensional experiment is lower than in the corresponding one-dimensional experiment. Similar trends are clear when we present our confidence as a function of time in figure 8m–o where we see that the two-dimensional confidence trends are qualitatively similar to the trends illustrated previously for the quasi-one-dimensional problem except that the two-dimensional confidence results are much lower. The implications of comparing the quasi-one- and two-dimensional confidence intervals are that there is an advantage in performing cell trajectory experiments in a one-dimensional geometry.

## 4. Conclusions

Analysing cell trajectories is a common approach used to infer different cell migration mechanisms [6–13]. A major challenge in the interpretation of cell trajectory experiments is that there is no standard protocol for designing or interpreting such assays and this means that it is often tricky to interpret and compare different sets of published cell trajectory data. Further, our modelling demonstrates that it is very difficult to extract meaningful

information from limited amounts of trajectory data, because these data are inherently noisy and become increasingly noisy with time and with increasing motility bias. This implies that the collection of such data ought to be designed with care by ensuring that trajectory data are collected over sufficiently long periods of time that we can observe meaningful trajectories, but not too long that the trajectory data become too noisy. We also show that cell trajectories associated with biased migration are more sensitive to noise than cell trajectories associated with undirected cell migration and that trajectory data obtained from two-dimensional migration assays are more sensitive to noise than trajectories from quasi-one-dimensional assays conducted in a channel geometry where the initial condition is independent of vertical location. This is a standard way to perform cell migration assays, such as a scratch assay [1,29], and essentially allows us to reduce the problem to a one-dimensional problem. Our results show that focusing on this type of channel assay problem gives more reliable data than genuinely two-dimensional assays.

**Acknowledgements.** We acknowledge support from the Australian Research Council (grant no. DP120100551), and the 2011 Royal Society International Exchange Scheme.

## References

- Maini PK, McElwain DLS, Leavesley DI. 2004 Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.* **10**, 475–482. (doi:10.1089/107632704323061834)
- Treloar KK, Simpson MJ. 2013 Sensitivity of edge detection methods for quantifying cell migration assays. *PLoS ONE* **8**, e67389. (doi:10.1371/journal.pone.0067389)
- Druckner NR, Epstein ML. 2007 Behavior of enteric neural crest-derived cells varies with respect to the migratory wavefront. *Dev. Dyn.* **236**, 84–92. (doi:10.1002/dvdy.20974)
- Young HM, Bergner AJ, Anderson RB, Enomoto H, Milbrandt J, Newgreen DF, Whittington PM. 2004 Dynamics of neural crest-derived cell migration in the embryonic mouse gut. *Dev. Biol.* **270**, 455–473. (doi:10.1016/j.ydbio.2004.03.015)
- Hannan D, Weinberg RA. 2000 The hallmarks of cancer. *Cell* **100**, 57–70. (doi:10.1016/S0092-8674(00)81683-9)
- Britto JM, Tait KJ, Johnston LA, Hammond VE, Kalloniatis M, Tan S-S. 2012 Altered speeds and trajectories of neurons migrating in the ventricular and subventricular zones of the reeler neocortex. *Cereb. Cortex* **21**, 1018–1027. (doi:10.1093/cercor/bhq168)
- Decaestecker C, Debeir O, Van Ham P, Kiss R. 2006 Can anti-migratory drugs be screened *in vitro*? A review of 2D and 3D assays for the quantitative analysis of cell migration. *Med. Res. Rev.* **27**, 149–176. (doi:10.1002/med.20078)
- Gilles C, Polette M, Zahm J-M, Tournier J-M, Volders L, Foidart J-M, Birembaut P. 1999 Vimentin contributes to human mammary epithelial cell migration. *J. Cell Sci.* **112**, 4615–4625.
- Hayot C, Debeir O, Van Ham P, Van Damme M, Kiss R, Decaestecker C. 2006 Characterization of the activities of actin-affecting drugs on tumor cell migration. *Toxicol. Appl. Pharm.* **211**, 30–40. (doi:10.1016/j.taap.2005.06.006)
- Kouvroukoglou S, Dee KC, Bizios R, McIntire LV, Zygorakis K. 2000 Endothelial cell migration on surfaces modified with immobilized adhesive peptides (2000). *Biomaterials* **21**, 1725–1733. (doi:10.1016/S0142-9612(99)00205-7)
- Stokes CL, Lauffenburger DA, Williams SK. 1991 Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *J. Cell Sci.* **99**, 419–430.
- Wang S-J, Saadi W, Lin F, Nguyen CM-C, Jeon NL. 2004 Differential effects of EGF gradient profiles on MDA-MB-231 breast cancer cell chemotaxis. *Exp. Cell Res.* **300**, 180–189. (doi:10.1016/j.yexcr.2004.06.030)
- Zahm J-M, Kaplan H, Hérard A-L, Doriot F, Pierrot D, Somelette P, Puchelle E. 1997 Cell migration and proliferation during the *in vitro* wound repair of the respiratory epithelium. *Cell Motil. Cytoskeleton* **37**, 33–43. (doi:10.1002/(SICI)1097-0169(1997)37:1<33::AID-CM4>3.0.CO;2-I)
- Kulesa PM, Fraser SE. 1998 Neural crest cell dynamics revealed by time-lapse video microscopy of whole embryo chick explant cultures. *Dev. Biol.* **204**, 327–344. (doi:10.1006/dbio.1998.9082)
- Kulesa P, Bronner-Fraser M, Fraser S. 2000 *In ovo* time-lapse analysis after dorsal neural tube ablation shows rerouting of chick hindbrain neural crest. *Development* **127**, 2843–2852.
- Tremel A, Cai A, Tirtaatmadja N, Hughes BD, Stevens GW, Landman KA, O'Connor AJ. 2009 Cell migration and proliferation during monolayer formation and wound healing. *Chem. Eng. Sci.* **64**, 247–253. (doi:10.1016/j.ces.2008.10.008)
- Ward M, McCann C, DeWulf M, Wu JY, Rao Y. 2003 Distinguishing between directional guidance and motility regulation in neuronal migration. *J. Neurosci.* **23**, 5170–5177.
- Mizoguchi T, Verkade H, Heath JK, Kuroiwa A, Kikuchi Y. 2008 Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521–2529. (doi:10.1242/dev.020107)
- Aubert M, Badoual M, Grammaticos B. 2009 A model for short- and long-range interactions of migrating tumour cell. *Acta Biotheor.* **56**, 297–314. (doi:10.1007/s10441-008-9061-x)
- Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834. (doi:10.1098/rsif.2008.0014)
- Liggett TM. 1999 *Stochastic interacting systems: contact, voter, and exclusion processes*. Berlin, Germany: Springer.
- Murray JD. 2002 *Mathematical biology I: an introduction*. Heidelberg, Germany: Springer.
- Simpson MJ, Landman KA, Hughes BD. 2009 Pathlines in exclusion processes. *Phys. Rev. E* **79**, 031920. (doi:10.1103/PhysRevE.79.031920)
- Simpson MJ, Landman KA, Hughes BD. 2010 Cell invasion with proliferation mechanisms motivated by time-lapse data. *Physica A* **389**, 3779–3790. (doi:10.1016/j.physa.2010.05.020)
- Simpson MJ, Towne C, McElwain DLS, Upton Z. 2010 Migration of breast cancer cells: understanding



- the roles of volume exclusion and cell-to-cell adhesion. *Phys. Rev. E* **82**, 041901. (doi:10.1103/PhysRevE.82.041901)
26. Hughes BD. 1995 *Random walks and random environments*, vol. 1. Oxford, UK: Oxford University Press.
  27. Deroulers C, Aubert M, Badoual M, Grammaticos B. 2009 Modeling tumor cell migration: from microscopic to macroscopic models. *Phys. Rev. E* **79**, 031917. (doi:10.1103/PhysRevE.79.031917)
  28. Painter KJ, Hillen T. 2002 Volume-filling and quorum-sensing in models for chemosensitive movement. *Can. Appl. Math. Q.* **10**, 501–543.
  29. Khain E, Katakowski M, Hopkins S, Szalad A, Zheng X, Jiang F, Chopp M. 2011 Collective behavior of brain tumor cells: the role of hypoxia. *Phys. Rev. E* **83**, 031920. (doi:10.1103/PhysRevE.83.031920)
  30. Simpson MJ, Treloar KK, Binder BJ, Haridas P, Manton KJ, Leavesley DI, McElwain DLS, Baker RE. 2013 Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *J. R. Soc. Interface* **10**, 20130007. (doi:10.1098/rsif.2013.0007)
  31. Chapra SC, Canale RP. 1998 *Numerical methods for engineers*. Singapore: McGraw-Hill.
  32. Casella G, Berger R. 2001 *Statistical inference*. Pacific Grove, CA: Duxbury Press.
  33. Coleman TF, Li Y. 1996 An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM J. Opt.* **6**, 418–445. (doi:10.1137/0806023)