
Proceedings Paper:
Kadochnikova, A., Isles, H.M., Renshaw, S.A. orcid.org/0000-0003-1790-1641 et al. (1 more author) (2018) Estimation of hidden chemoattractant field from observed cell migration patterns. In: Rojas, C., (ed.) IFAC-PapersOnLine. 18th IFAC Symposium on System Identification SYSID 2018, 09-11 Jul 2018, Stockholm, Sweden. IFAC , pp. 766-771.

https://doi.org/10.1016/j.ifacol.2018.09.161

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-15 (2018) 766–771

Estimation of Hidden Chemoattractant Field from Observed Cell Migration Patterns ⋆

Anastasia Kadochnikova * **Hannah M. Isles** **
Stephen A. Renshaw ** **Visakan Kadirkamanathan** *
* *Department of Automatic Control and Systems Engineering, the University of Sheffield, UK (e-mail: {akadochnikova1, visakan}@sheffield.ac.uk).*
** *Department of Infection, Immunity and Cardiovascular Disease, The University of Sheffield, UK (e-mail: {hmisles1, s.a.renshaw}@sheffield.ac.uk)*

Abstract: Neutrophilic chemotaxis is essential to immune system response to external threats. During this process cells alternate between directed motion towards the higher concentration of external stimuli and correlated random walk. An individual neutrophil migration can thus be characterised as a stochastic dynamical process driven by an external chemotactic environment that is typically not measured. This introduces the problem of estimating spatially-varying chemoattractant concentration field from the observed migration patterns of cell populations. We propose a solution to this estimation problem in a statistical inference framework. The framework has measured cell positions in the field as inputs and employs the expectation-maximisation algorithm for joint estimation of full cell states and parameters of the chemoattractant field decomposed with cubic B-splines. The performance of the developed algorithm is accessed via process *in vivo* measurements of cell positions in the injured tail fin of zebrafish. Estimation results for different injury types evidence that the proposed estimation algorithm provides a rigorous connection between mathematical modelling and experimental data.

Keywords: Parameter estimation, maximum likelihood methods, basis functions, filtering and smoothing, biological systems.

1. INTRODUCTION

The use of *in silico* analytical tools is widely accepted in biomedical and clinical research. They support decision-making process by employing mathematical modelling that allows the interpretation of biological processes via simplified prototypes (Bekey and Beneken, 1978). The rapid development of the *in silico* biology is driven by three key processes: novel measurement technologies, continuous improvement and training of models representing complex biological processes, and fitting of the existing models to available experimental data. Both the task of model fitting and the model training can be viewed as a parameter estimation problem and solved using classical system identification tools.

Chemotaxis is a directed migration of cells driven by external attractant concentrations. This type of movement facilitates key processes in living organisms, both physiological, such as tissue generation or immune system response, and pathological, such as cancer metastasis. Chemotaxis is performed by a wide range of cell types, from bacteria to eukaryotes, including neutrophils - a major class of mammalian white blood cells that respond to pathogen-associated stimuli. Their ability to kill invading microorganisms makes neutrophils the main mediators of innate immune system response to external threats. Therapeutic importance of neutrophil regulation is obvious, but it requires the quantitative evaluation of multiple aspects of cell migration that are experimentally inaccessible.

Knowledge of the underlying chemoattractant environment is crucial for model-based analysis of cell motility. The investigation of the cell-environment interaction has been greatly facilitated by biological experiments, most of which are conducted *in vitro* (Latin for "in the glass") within a regulated environment that can be easily quantified. Unfortunately, artificial environments designed during *in vitro* experiments do not accurately represent natural attractant generation process. Recent achievements in image processing led to the development of *in vivo* (Latin for "in the living") microscopy in various animal models such as flies, zebrafish, and mice. For instance, transparency of zebrafish early embryo and excessive genomic resources available for its genetic modification are utilised in experiments where the cell migration is observed directly in the living organism (Renshaw and Loynes, 2006). Neutrophil recruitment is triggered either by full tail fin transection or by a tail fin

⋆ The research is supported by the PhD Scholarship, Department of Automatic Control and Systems Engineering, the University of Sheffield to AK and with the partial support of BBSRC project BB/L000830/1 to HI, SA and VK.

10.1016/j.ifacol.2018.09.161

nick injury. Although this type of experiments fostered multiple experimental studies, there are very few models available to date that would describe the chemoattractant field at the whole organism scale.

Processing of *in vivo* data presents a new challenge in quantifying the unobserved environment. Experiments are conducted with living organisms, where it is impossible to generate controllable excitation signals. That is why, while migration is actively triggered by a tail fin injury, the concentration of attractant cannot be measured directly in the body. This poses a problem of chemoattractant field estimation based on cell migration patterns observed *in vivo*.

The problem we address here is similar to the ones solved by Liepe et al. (2012) and Kadirkamanathan et al. (2012). Authors of the former tackle environment estimation problem by introducing three different phenomenological models of the hidden field and discriminating between them within the Bayesian framework. The proposed solution requires making prior assumptions about the shape of the field, whose description becomes constrained to several generic models. The latter introduces a more flexible methodology in terms of field representation that relies on basis function decomposition of the chemoattractant environment. However, the authors solve the problem deterministically by decoupling field parameter learning and cell state estimation.

In this paper, we introduce a maximum likelihood framework that can estimate the unobserved chemoattractant concentration field and recover cell states simultaneously from spatiotemporally resolved *in vivo* data. Our solution employs the potential field method of modelling object-environment interaction, popular in simultaneous localisation and mapping (Murphy and Godsill, 2012) and swarming robotics (Mangion et al., 2011). In this approach, the underlying environment is modelled as a static potential field $\mathcal{U}$ driving object movement. This translates to the cell migration as follows: a cell that moves through the chemoattractant concentration field is subject to the attractive force that is proportional to the field gradient at the current cell position. The underlying field is parametrised via basis function decomposition which provides the flexibility necessary for inference of complex shapes without changing the model structure.

The framework is applied to six zebrafish larvae to test whether the model will confirm the experimental results that show the highest attractant concentration near the wound (Niethammer et al., 2009). Estimation results suggest that the proposed framework can serve as a foundation for model-based investigations of underlying mechanisms regulating the cell-environment interaction during chemotaxis and other cell signal mediated processes.

2. PRELIMINARIES

The estimation of the chemoattractant field relies on the assumed model of cell-environment interaction. This section describes the model of cell dynamics that incorporates both random aspect of Brownian-like cell motion and the influence of the underlying field as a deterministic term. Combined with the basis function representation of the field structure, proposed model facilitates the formulation of the inference problem within the well-known EM setting.

2.1 Cell dynamics

The dynamics of the k-th cell is described by a discrete-time state-space model

$$\mathbf{x}_{t+1}^k = A\mathbf{x}_t^k + B\mathbf{u}_t^k + G\mathbf{w}_t^k, \tag{1}$$

$$\mathbf{y}_t^k = C\mathbf{x}_t^k + \mathbf{v}_t^k, \tag{2}$$

where the state vector at sample time t contains position and velocity of the cell centroid on the two-dimensional image

$$\mathbf{x}_t^k = [x_1, x_2, \dot{x}_1, \dot{x}_2]_t^\top, \tag{3}$$

and where the state noise $\mathbf{w}_t \sim \mathcal{N}(0, Q_w)$, $Q_w \in \mathbb{R}^{2\times 2}$ and the measurement noise $\mathbf{v}_t \sim \mathcal{N}(0, R_v)$, $R_v \in \mathbb{R}^{2\times 2}$ are i.i.d. Gaussian sequences. System matrices A,B,G, and C are defined as follows

$$A = \begin{bmatrix} \mathbb{I} & T\times\mathbb{I} \\ \mathbb{O} & \mathbb{I} \end{bmatrix}, \tag{4a}$$

$$B = \left[\tfrac{T^2}{2}\times\mathbb{I}\ T\times\mathbb{I}\right]^\top, \tag{4b}$$

$$G = \left[\tfrac{T^2}{2}\times\mathbb{I}\ T\times\mathbb{I}\right]^\top, \tag{4c}$$

$$C = [\mathbb{I}\ \mathbb{O}]^\top, \tag{4d}$$

where T is the time increment, $\mathbb{I}$ is an identity matrix and $\mathbb{O}$ is a zero matrix of size 2×2.

The input vector $\mathbf{u}_t$ in (1) incorporates the influence of the environment on the moving cell. In the adopted potential filed paradigm, the force acting on a cell that moves through the chemoattractant field is given by the gradient of the field at the current cell location $x = [x_1, x_2]$.

Assumption 1. The input of the SSM (1) is proportional to the steepest gradient of the chemoattractant concentration field $\mathcal{U}$

$$\mathbf{u}_t = \mu\nabla\mathcal{U}, \tag{5}$$

where μ is a proportionality constant arbitrary set to 1.

2.2 The environment

A parametrised model of the functional form of the gradient is required for solving the field inference problem. Since no prior information about the environment is available, it is desirable to make minimum assumptions about its shape. This is achieved via basis function decomposition

$$\mathcal{U}(x) = \sum_{h=1}^{N_b} \beta_h(x)\theta_h, \tag{6}$$

where θ_h is a scaling parameter corresponding to a basis function $\beta_n(x)$ at the coordinate x, and where N_b is the number of basis functions. This representation of the potential field provides a parametric description of input at the position x

$$u = \mu\nabla\mathcal{B}(x)\Theta, \tag{7}$$

where

$$\mathcal{B}(x) = [\beta_1(x), \beta_2(x), \dots, \beta_{N_b}(x)], \tag{8}$$

$$\Theta = [\theta_1, \theta_2, .., \theta_{N_b}]^\top. \tag{9}$$

Combining (1) and (7) leads to a parametrised state-space model of cell-environment interaction that is linear with respect to the unknown parameter vector Θ.

2.3 Problem formulation

Consider K cells whose dynamics described by the discrete time SSM in the form

$$\mathbf{x}_{t+1}^k = A\mathbf{x}_t^k + \phi_t^k\Theta + G\mathbf{w}_t^k, \tag{10}$$

where $\phi_t^k = B\mu\mathcal{B}(x_t)$. The measurement process is described by (2).

Assumption 2. The full state $\mathbf{x}_t^k$ of the k-th cell at sampling time t is unobserved. A sequence states of all cells at all sampling times constitutes the hidden data set

$$\mathcal{X} = \left\{\mathbf{x}^k\right\}_{k=1}^K, \tag{11}$$

where

$$\mathbf{x}^k = \left\{\mathbf{x}_t^k\right\}_{t=1}^{T_k} \tag{12}$$

is a sequence of true states of an individual cell that is observed from until the sampling time T_k.

Assumption 3. A set of cell tracks is available from the tracking algorithm

$$\mathcal{Y} = \left\{\mathbf{y}^k\right\}_{k=1}^K, \tag{13}$$

where each track is a set of measured cell positions at each sampling time

$$\mathbf{y}^k = \left\{\mathbf{y}_t^k\right\}_{t=1}^{T_k}, \tag{14}$$

where a single measurement corresponds to the observed position of the cell centroid

$$\mathbf{y}_t^k = \left[y_{x_1}^k, y_{x_2}^k\right]_t^\top. \tag{15}$$

Tracking results for six fish are presented in Fig. 1.

Given the model defined above and a set of observations $\mathcal{Y}$, we can address following estimation objectives:

- *Cell state estimation*, where each hidden state in the set $\mathcal{X}$ is estimated given the estimate of parameters Θ and the full set of observations $\mathcal{Y}$.
- *Field parameter inference*, where unknown parameters Θ are estimated given the complete data set $\mathcal{Z} = \{\mathcal{X}, \mathcal{Y}\}$.

3. POTENTIAL FIELD INFERENCE

This section presents the formulation of the chemoattractant field inference problem within a maximum likelihood framework. Since the full set of observations $\mathcal{Y}$ is available from the tracking procedure, unknown parameters can be estimated via the expectation maximisation (EM) algorithm (Dempster et al., 1977). It is a popular approach for the joint state-parameter estimation of generic linear SSMs (Gibson and Ninness, 2005). The algorithm efficiently separates two estimation problems and solves them iteratively until convergence. The fact that the EM method does not require any *a priori* knowledge about unobserved data makes it particularly applicable to biological systems, where it is desirable to minimise the number of model assumptions (Dewar and Kadirkamanathan, 2007).

3.1 The likelihood function

In the EM algorithm, maximum likelihood estimates (MLEs) of unknown parameters can be obtained by maximising the complete data log-likelihood function

$$\hat{\Theta}_{\mathbf{ML}} = \arg\max_{\Theta} L(\mathcal{Z} \mid \Theta), \tag{16}$$

where

$$L(\mathcal{Z} \mid \Theta) = \log p(\mathcal{Z} \mid \Theta). \tag{17}$$

The fact that part of the data in the set Z is hidden prevents the computation of the full likelihood. Instead, the EM algorithm evaluates the expectation of $p(\mathcal{Z} \mid \Theta)$ given current parameter estimates and conditioned on the available measurement set $\mathcal{Y}$. The expected value of the complete data log-likelihood taken with respect to the hidden data distribution $p(\mathcal{X} \mid \mathcal{Y}, \hat{\Theta}^i)$ is termed the $\mathcal{Q}$-function and defined as follows

$$\mathcal{Q}(\Theta, \hat{\Theta}^i) = \mathbb{E}\left[\log p(\mathcal{Z} \mid \Theta) \mid \mathcal{Y}, \hat{\Theta}^i\right]. \tag{18}$$

The process of evaluating (18) constitutes the expectation step of the algorithm. The maximisation step updates parameter estimates by maximising the $\mathcal{Q}$-function

$$\hat{\Theta}^{i+1} = \arg\max_{\Theta} \mathcal{Q}(\Theta, \hat{\Theta}^i). \tag{19}$$

3.2 The maximisation step

For K cells observed across all sampling times there exists complete data set $\mathcal{Z} = \{\mathcal{X}, \mathcal{Y}\}$ with joint probability defined as follows

$$p(\mathcal{Z} \mid \Theta) = \prod_{k=1}^K \left[p(\mathbf{x}_1^k) \prod_{t=2}^{T_k} p(\mathbf{x}_t^k \mid \mathbf{x}_{t-1}^k, \Theta) \prod_{t=1}^{T_k} p(\mathbf{y}_t^k \mid \mathbf{x}_t^k) \right], \tag{20}$$

where individual probability densities are

$$p(\mathbf{x}_1^k) = \mathcal{N}(\mu_1, \Pi), \tag{21a}$$

$$p(\mathbf{x}_t^k \mid \mathbf{x}_{t-1}^k, \Theta) = \mathcal{N}(A\mathbf{x}_t^k + B\mathbf{u}_t^k, Q_w), \tag{21b}$$

$$p(y_t^k \mid \mathbf{x}_t^k) = \mathcal{N}(C\mathbf{x}_t^k, R_v). \tag{21c}$$

We note that only the term (21b) is a function of Θ.

Lemma 4. Suppose that all constituent terms of (20) are estimated during the expectation step. Then the MLE of the unknown parameter vector is given by

$$\begin{aligned}\hat{\Theta} = &\left(\sum_{k=1}^K \sum_{t=2}^{T_k} (\phi_{t-1}^k)^\top \phi_{t-1}^k \right)^{-1} \times \\ &\times \left[\sum_{k=1}^K \sum_{t=2}^{T_k} (\phi_{t-1}^k)^\top \left(\breve{\mathbb{E}}\left[\mathbf{x}_t^k\right] - A\breve{\mathbb{E}}\left[\mathbf{x}_{t-1}^k\right] \right) \right],\end{aligned} \tag{22}$$

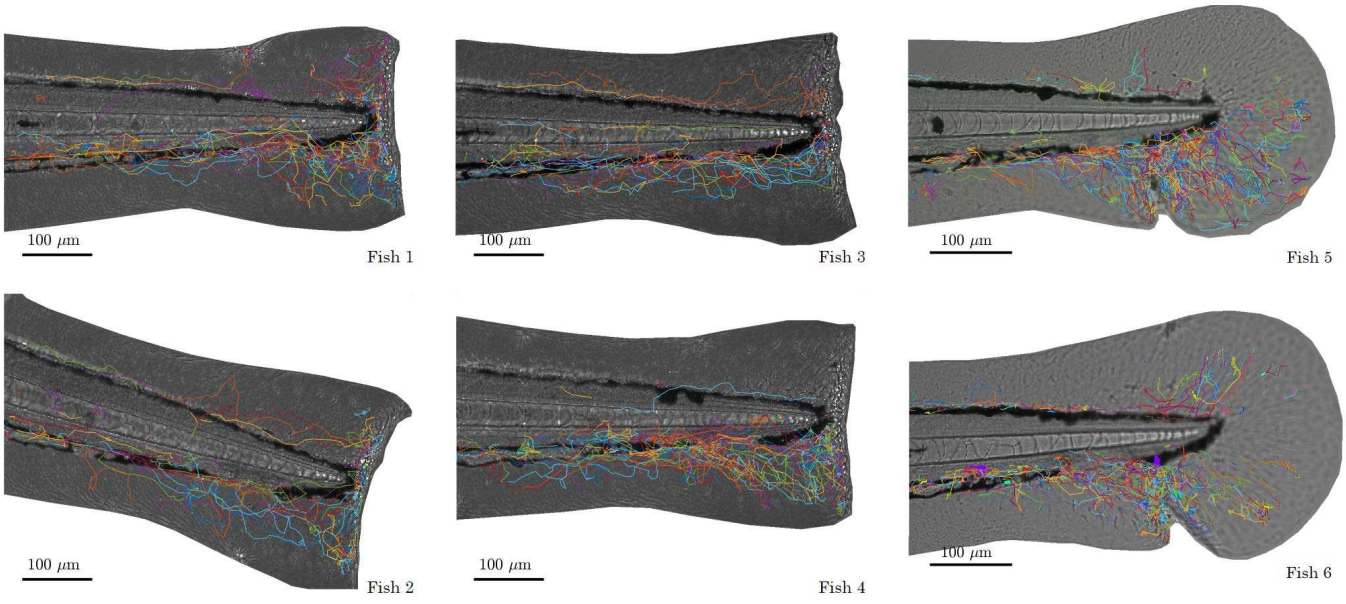
where

$$\breve{\mathbb{E}}\left[\mathbf{x}_{t-1}^k\right] \triangleq \mathbb{E}\left[\mathbf{x}_{t-1}^k \mid \mathbf{y}^k, \hat{\Theta}^i\right], \tag{23a}$$

$$\breve{\mathbb{E}}\left[\mathbf{x}_t^k\right] \triangleq \mathbb{E}\left[\mathbf{x}_t^k \mid \mathbf{y}^k, \hat{\Theta}^i\right] \tag{23b}$$

are expectations of cell full states at two subsequent times conditioned available measurements and the current parameter estimate $\hat{\Theta}^i$.

Proof. Expanding (18) using the definition of probabilities (21) leads to



(a) Full tail transection (b) Tail fin nick

Fig. 1. Tracking results plotted against greyscale images of zebrafish larvae. Figure (a) shows results for full transection injury. Figure (b) presents results obtained from a tail fin nick.

$$\mathcal{Q}(\Theta,\hat{\Theta}^i)=\mathbb{E}\Big[\text{const}-\frac{1}{2}\sum_{k=1}^{K}\sum_{t=2}^{T_k}Q_w^{-1}\times \times\big(\mathbf{x}_t^k-A\mathbf{x}_{t-1}^k-\phi_{t-1}^k\Theta\big)^2\Big], \tag{24}$$

where the constant includes all terms independent of Θ. We can expand (24) further and rearrange it using the properties of the expectation to produce

$$\mathcal{Q}(\Theta,\hat{\Theta}^i)=-\frac{1}{2}\sum_{k=1}^{K}\sum_{t=2}^{T_k}\Big\{\Theta^\top(\phi_{t-1})^\top Q_\omega^{-1}\phi_{t-1}\Theta+ +2\Theta^\top(\phi_{t-1})^\top Q_\omega^{-1}A^j\breve{\mathbb{E}}\left[\mathbf{x}_{t-1}^k\right]- -2\Theta^\top(\phi_{t-1})^\top Q_\omega^{-1}\breve{\mathbb{E}}\left[\mathbf{x}_t^k\right]+\text{const}\Big\}, \tag{25}$$

where the constant is extended to include more non-dependent terms.

The estimate of the parameter vector corresponding to the maximum of the log-likelihood function can be obtained by taking the partial derivative of (25) with regard to Θ and setting it equal to zero,

$$\frac{\partial\mathcal{Q}(\Theta,\hat{\Theta}^i)}{\partial\Theta}=-\sum_{k=1}^{K}\sum_{t=2}^{T_k}\Big\{(\phi_{t-1}^k)^\top Q_\omega^{-1}\phi_{t-1}^k\Theta+ +(\phi_{t-1}^k)^\top Q_\omega^{-1}\big(A\breve{\mathbb{E}}\left[\mathbf{x}_{t-1}^k\right]-\breve{\mathbb{E}}\left[\mathbf{x}_t^k\right]\big)\Big\}=0. \tag{26}$$

Solving (26) for Θ leads directly to (22). Furthermore, the 2^{nd} partial derivative of the $\mathcal{Q}$-function defined by

$$\frac{\partial^2\mathcal{Q}(\Theta,\hat{\Theta}^i)}{\partial\Theta^2}=-\sum_{k=1}^{K}\sum_{t=2}^{T_k}\Big\{(\phi_{t-1}^k)^\top Q_\omega^{-1}\phi_{t-1}^k\Big\} \tag{27}$$

is negative definite, which verifies that the new parameter estimate $\hat{\Theta}^{i+1}$ is located at a local maximum of the likelihood function, completing the proof.

3.3 The expectation step

Prior to the maximisation step we must evaluate the $\mathcal{Q}$-function conditioned on the current parameter estimates and available measurements. This process involves estimation of full cell states at two successive sampling times:

$$\breve{\mathbb{E}}\big[\mathbf{x}_{t-1}^k\big]=\mathbb{E}\big[\mathbf{x}_{t-1}^k\mid\mathbf{y}^k,\hat{\Theta}^i\big]=\hat{\mathbf{x}}_{t-1|T_k}^k, \tag{28}$$

$$\breve{\mathbb{E}}\big[\mathbf{x}_t^k\big]=\mathbb{E}\big[\mathbf{x}_t^k\mid\mathbf{y}^k,\hat{\Theta}^i\big]=\hat{\mathbf{x}}_{t|T_k}^k, \tag{29}$$

where $\hat{\mathbf{x}}_{t-1|T_k}^k$ and $\hat{\mathbf{x}}_{t|T_k}^k$ are smoothed states. Since the set of full tracks $\mathcal{Y}$ is available *a priori*, the state space model described by (10) and (2) is used in a fixed-interval smoother to recover full cell states (Rauch et al., 1965).

Remark 5. Since the basis functions $\mathcal{B}$ depend on the current position x_t^k of the k-th cell at time t, the matrix ϕ_t^k may be a nonlinear function of the cell hidden state. In order to solve the state-parameter estimation as a linear problem, all matrices ϕ are calculated at the beginning of the expectation step on each iteration of the EM algorithm by replacing hidden x_t^k with its estimate from the previous iteration $(\hat{x}_t^k)^{(i-1)}$.

3.4 The estimation algorithm

The algorithm is initialised by calculating values of ϕ_t^k for all measured cell positions $\hat{\phi}_t^k=B\mu\mathcal{B}(\mathbf{y}_t^k)$. These matrices are then utilised in the estimation of full cell states with the assumption that there is no field acting on moving cells, *i.e.* all magnitude coefficients are set equal to zero $\hat{\Theta}^0=0$. The maximisation step is performed afterwards to obtain an initial parameter estimate. The two steps defined above are repeated until convergence. It has been established that the likelihood function increases at each iteration (Wu, 1983). However, assessing convergence of the likelihood function comes at a high computational cost as it requires

computation of all terms included in the constant in (25) at each iteration of the algorithm. Instead, use the following stopping criteria

$$\Delta\Theta^i = \frac{(\Theta^i - \Theta^{i-1})^\top (\Theta^i - \Theta^{i-1})}{(\Theta^i)^\top \Theta^i} < \epsilon. \qquad (30)$$

The rate of convergence is highly depended on the selection of the initial parameter vector estimate.

Remark 6. As an EM solution, the algorithm described in this section is liable to converge to local maxima of the likelihood function.

4. RESULTS

The estimation is performed with the same settings across all fish larvae. The time increment $T = 0.5$ min used in the model matches settings of the the microscopy procedure. The process noise covariance is set to $Q_w = 2\mathbb{I}\mu m^4/\text{min}^4$ and the measurement noise covariance is $R_v = \mathbb{I}\mu m^2$, where $\mathbb{I}$ is a 2×2 identity matrix. Initial cell speed estimates for forward and backward passes of the smoother are arbitrary assumed to be zero with the covariance $0.5\mathbb{I}\ \mu m^2/min^2$.

In this study, the decomposition of chemoattractant field is performed using tensor spline products

$$\beta_h(x) = \sum_{j=1}^{n_1} \sum_{l=1}^{n_2} c_{j,l} \psi_j(x_1) \varphi_l(x_2), \qquad (31)$$

where $\psi_j(x_1)$ and $\varphi_l(x_2)$ are cardinal cubic B-splines constructed for a uniform knot sequence in each spatial dimension, and where $c_{j,l}$ is a scaling coefficient arbitrary set equal to 1. B-splines are selected over Gaussian bases used in (Kadirkamanathan et al., 2012), as they provide a better approximation of first order functions.

A 5×3 grid of basis functions overlapping at one third of their base width is placed over each fish image. The node spacing is changed automatically for different image sizes. The boundary conditions are introduced for the visualisation of the estimated field by masking the area of the image outside of fish boundary.

4.1 Estimation results

The estimated chemoattractant fields for six zebrafish larvae are presented in Fig. 2. Each result demonstrates the correlation between the total number of cells passing through the area and the estimated chemoattractant field magnitude of that area.

Remark 7. The solution provided by the algorithm is unique for the selected initial estimate $\hat{\Theta}^0$, but the shape of the estimated chemoattractant field is only unique up to an additive constant since it is not the magnitude of the field that drives the cell migration but its gradient.

As illustrated in Fig. 2, estimated fields for Fish 1 and 3 conform to the well-established hypothesis that the chemoattractant is uniformly distributed along the axis parallel to the injury with the highest concentration located at the injury site. Fish 4, however, demonstrates a different type of neutrophil behaviour. Contrary to the expected pattern, cells "swarm" towards the lower region of the injury site. The swarming may be caused by the self-generated gradient recently observed in (Lammermann et al., 2013), that may be modulated predominantly by the LTB4 mediator released by neutrophils themselves. In Fish 2, for which the lowest number of cell tracks is available, the
the most persistent cell tracks.

For the nicked tail fin injury the range of chemoattractant concentrations is narrower, which is expected for a minor injury. The highest field magnitude is also smaller compared to models with the full tail fin transection, which is consistent with the fact that smaller injury attracts fewer cells. While in Fish 5 the estimated peak of chemoattractant concentration is located near the wound, Fish 6 has an outlier peak that matches the area with several persistent cell tracks directed away from the wound. Cells may be driven by another neutrophil-generated gradient. Results for Fish 4 and Fish 6 indicate that the future modelling and estimation work on population dynamics should take into account cell-to-cell interaction.

5. CONCLUSION

In this paper, we introduce a maximum-likelihood framework for the chemoattractant environment inference from observed migration patterns of eukaryotic cells. The proposed expectation-maximisation algorithm builds on the potential field model of interaction between migrating cells and the hidden underlying environment. Based on the assumption that cell movement is driven exclusively by the field gradient, we utilise cell positions tracked *in vivo* to infer the chemoattractant environment. The performance of the developed framework is demonstrated on several data sets of neutrophil recruitment to the tail injury site observed in a transgenic zebrafish larvae. Presented inference results for two types of tail fin injury are obtained with the same settings and support our claim that the framework does not require any *a priori* knowledge about the shape of the environment.

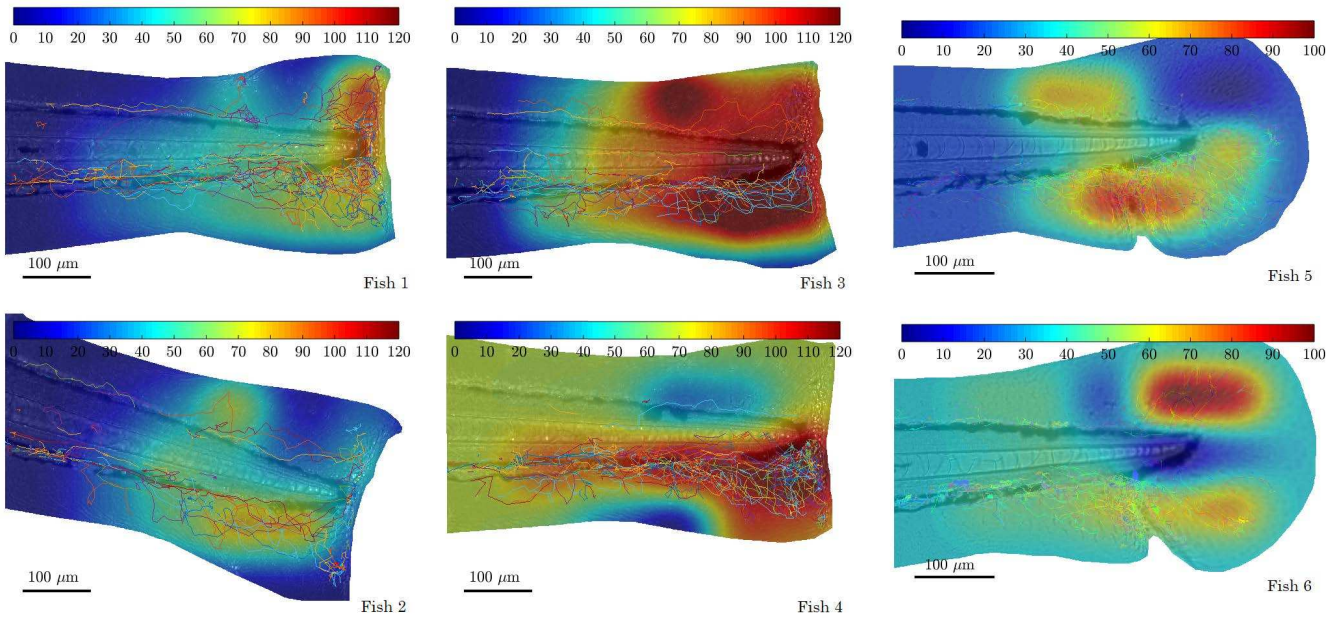
A number of extensions for the proposed solution can be considered. For instance, changes in the behaviour of an individual migrating cell fall beyond the scope of the linear state space representation. A more complex hybrid model can be employed to represent heterogeneous dynamics of moving cells, although lack of exact methods for state estimation of hybrid systems would significantly increase the complexity of the resulting algorithm. Furthermore, we consider a model based on Newton's second law, where all dynamic matrices are known. Including model selection algorithm to discriminate between various representations of cell dynamics in the existing framework will extend its applicability to a broader range of models.

REFERENCES

Bekey, G.A. and Beneken, J.E. (1978). Identification of biological systems: a survey. *Automatica*, 14(1), 41 – 47.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society, Series B*, 39(1), 1–38.

Dewar, M. and Kadirkamanathan, V. (2007). A canonical space-time state space model: State and parameter estimation. *IEEE Transactions on Signal Processing*, 55(10), 4862–4870.



(a) Full tail transection (b) Tail fin nick

Fig. 2. Estimated chemoattractant fields plotted against images of zebrafish larvae. Figure (a) shows results for full transection injury. Figure (b) presents results obtained from a tail fin nick.

Gibson, S. and Ninness, B. (2005). Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41(10), 1667–1682.

Kadirkamanathan, V., Anderson, S., Billings, S., Zhang, X., Holmes, G., Reyes-Aldasoro, C., Elks, P., and Renshaw, S. (2012). The neutrophil's eye-view: inference and visualisation of the chemoattractant field driving cell chemotaxis in vivo. *PLoS ONE*, 7(4), e35182.

Lammermann, T., Afonso, P., Angermann, B., Wang, J., Kastenmuller, W., Parent, C., and Germain, R. (2013). Neutrophil swarms require ltb4 and integrins at sites of cell death in vivo. *Nature*, 498(5), 371–5.

Liepe, J., Taylor, H., Barnes, C.P., Huvet, M., Bugeon, L., Thorne, T., Lamb, J.R., Dallman, M.J., and Stumpf, M.P.H. (2012). Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation. *Integrative Biology*, 4, 335–345.

Mangion, A.Z., Anderson, S., and Kadirkamanathan, V. (2011). Exploration and control of stochastic spatiotemporal systems with mobile agents. *IFAC Proceedings Volumes*, 44(1), 4489 – 4494. 18th IFAC World Congress.

Murphy, J. and Godsill, S. (2012). Simultaneous localization and mapping for non-parametric potential field environments. In *2012 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 1–6.

Niethammer, P., Grabher, C., Look, A.T., and Mitchison, T.J. (2009). A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish. *Nature*, 459, 996–999.

Nusslein-Volhard, C. and Dahm, R. (2002). *Zebrafish: A Practical Approach*. [Oxford University Press], 1st edition.

Rauch, H.E., Tung, F., and Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8), 1445–1450.

Renshaw, S.A. and Loynes, C.A. (2006). A transgenic zebrafish model of neutrophilic inflammation. *Blood*, 108(13), 3976–3978.

Wu, J.C.F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1), 95–103.

Appendix A. DATA ACQUISITION

Ethics statement. All animal experiments were performed according to legislation and guidelines detailed in the Animals (Scientific Procedures) Act 1986. Ethical approval was given by the University of Sheffield Local Ethical Review Panel and experiments were fully approved by the Home Office (Project license PPL 70/8178).

Zebrafish husbandry. The neutrophil-specific fluorescent Tg(mpx:GFP)i114 zebrafish line, also referred to as mpx:GFP, was used for all experiments. Adult zebrafish were raised in the Bateson Centre at The University of Sheffield in UK Home-Office approved aquaria. All zebrafish were maintained according to standard protocols in (Nusslein-Volhard and Dahm, 2002).

Microscopy. All experiments were conducted on 3 days post fertilisation (dpf) mpx:GFP embryos anaesthetised by immersion in E3 containing 4.2% tricaine (Sigma-Aldrich). The inflammatory response was activated by tail fin transection using a sterile scalpel blade as described previously in (Renshaw and Loynes, 2006). 3dpf mpx:GFP embryos were mounted in 0.7% low melting point agarose (Sigma-Aldrich) containing 4.2% tricaine for imaging immediately after tail transection. Time lapse imaging was performed from 0-5 hours post injury with 30 second intervals between time points using an Eclipse TE2000-U fluorescence microscope with a Andor Zyla 5.5 camera. The neutrophil trajectory data was extracted from time lapse images via the tracking tool within the NIS Elements software.