Statistical modelling of cell movement

Diana Giurghita[1], Dirk Husmeier[1]

[1] University of Glasgow, United Kingdom

E-mail for correspondence: d.giurghita.1@research.gla.ac.uk

Abstract: In this paper we demonstrate an application of the unscented Kalman filter in the context of cell movement, using a model defined in terms of stochastic differential equations (SDEs).

Keywords: Chemotaxis, Cancer, Stochastic Differential Equations, Unscented Kalman Filter.

1 Introduction

Many important biological processes, such as wound healing, tissue development and cancer cell invasion, are based on the collective movement of cells. One of the main mechanisms for directed cell movement is chemotaxis, where cells follow chemical gradients (chemoattractants) present in their environment. These gradients might arise from the presence of a local source of chemoattractant or due to local depletion of the chemical in the environment (Tweedy et al. 2016). An example of the former scenario is the migration of breast tumour cells that respond to the epidermal growth factor released by macrophages. In an attempt to acquire a deeper understanding of the mechanisms behind cell movement many population based models have been formulated using partial differential equations, with very few of them attempting to fit these models to actual data.
In this paper, we propose a model that describes the movement of any individual cell being driven by an external resource gradient using SDEs of the form:

$$dX_t = \sigma dB_t^X, \qquad dY_t = \frac{\alpha\beta \exp[-\beta(Y_t - \gamma t)]}{\{1 + \exp[-\beta(Y_t - \gamma t)]\}^2} dt + \sigma dB_t^Y \tag{1}$$

Equations (1) describe the evolution in time of the x and y coordinates of a cell in 2D space. σdB_t^X and σdB_t^X are Brownian motion terms which in

This paper was published as a part of the proceedings of the 32nd International Workshop on Statistical Modelling (IWSM), Johann Bernoulli Institute, Rijksuniversiteit Groningen, Netherlands, 3–7 July 2017.

this model represent the intrinsic randomness in a cell's movement. The coordinate in the y direction has a drift term that is described by three parameters: α - the amplitude of the resource gradient, β - the steepness of the gradient and γ which indicates how fast the gradient changes over time. The strength of the random component in the cell movement equations is indicated by the diffusion coefficient σ.

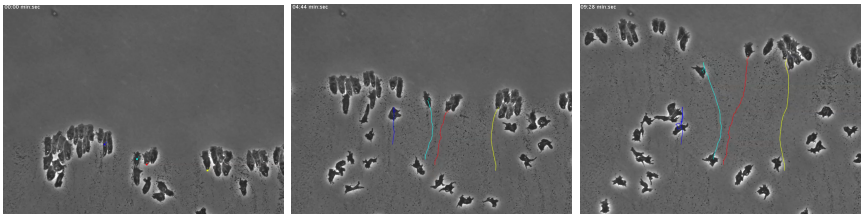

FIGURE 1. Three frames from the video recording *Dictyostelium* cells movement.

In this paper, we present our approach to fitting this model to cell movement data using a non-linear Bayesian filter. We provide some insight into the particularities of this model using simulated data and we discuss the results of this analysis from a real data set, describing the movement of *Dictyostelium* cells (see Figure 1).

2 Methods

Inference in non-linear dynamical systems poses numerous challenges due to the stochastic nature of the data, intractable likelihoods and unidentifiable parameters. Recent developments have tackled this problem using likelihood-free methods (sequential Monte Carlo ABC) or computational methods (particle Markov Chain Monte Carlo) (Golightly & Wilkinson, 2011), however these can become too computationally expensive as the number of time points or parameters increases. The unscented Kalman filter (UKF) is an online Bayesian filtering method that can easily be scaled up to higher dimensions (Julier & Uhlmann, 1997). Intuitively, the UKF starts from the initial distribution of the state vector, drawn from a multivariate normal distribution, which is then iterated through a prediction and updating step for each measurement available using the transition and observation models.
We introduce the UKF by referring to a general state-space model:

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \boldsymbol{\epsilon}_t), \quad \mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \boldsymbol{\nu}_t) \tag{2}$$

where $\mathbf{x}_t$ represents the vector of the hidden states, $\mathbf{y}_t$ are the measurements, $\boldsymbol{\epsilon}_t$ is the process noise at time t, $\boldsymbol{\nu}_t$ is the observation noise at time t and the functions $\mathbf{f}$ and $\mathbf{g}$ represent the transition and, respectively, the

observation models. The model parameters $\boldsymbol{\theta}$ can be included as dynamical variables in the hidden states vector $\mathbf{x}_t$, which means they will be estimated at every time point along with the observed system states (Sitz et al., 2002). The advantage of the method comes from the fact that the probability distribution of the predictor step: $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and the probability distribution of the updating step $p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{y}_{1:t-1})$ can be obtained in closed form using properties of the Gaussian distribution (Julier & Uhlmann, 1997):

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{N}(\mathbf{x}_t|\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t) \tag{3}$$

$$p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{y}_{1:t-1}) \approx \mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \tag{4}$$

where $\bar{\boldsymbol{\mu}}_t$ and $\bar{\boldsymbol{\Sigma}}_t$ are the prediction mean and covariance at time t and $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are the update mean and covariance at time t (see Julier & Uhlmann, 1997 for full derivations). Therefore, the algorithm essentially updates the mean and covariance of the Gaussian distribution of the state vector at each iteration. The approximation of the Gaussian distribution is made using the unscented transform, which consists of a set of deterministically chosen sigma-points that are passed through the non-linear function and weighted to obtain the mean and covariance of the Gaussian. The unscented transformation is used twice for each iteration of the algorithm: in the prediction and respectively in the update step.

3 Simulation results

We apply the Euler-Maruyama discretisation to bring the system in Equation (1) into the standard state-space model described in Section 2:

$$X_t = X_{t-1} + \sigma \Delta B_t^X \tag{5}$$

$$Y_t = Y_{t-1} + \frac{\alpha\beta \exp[-\beta(Y_{t-1} - \gamma t)]}{\{1 + \exp[-\beta(Y_{t-1} - \gamma t)]\}^2} dt + \sigma \Delta B_t^Y \tag{6}$$

Here ΔB_t^X and ΔB_t^Y are just sums of random normal increments between time $t-1$ and t. Equations (5) and (6) thus define the transition function $\mathbf{f}$ from Equation 2. In this scenario, we assume the process is observed with a small amount of Gaussian noise $\nu_t \sim \mathcal{N}(0, 0.1^2)$, so $\mathbf{g}$ from Equation 2 is just the identity function.

We then fit the UKF to a synthetic data set using the following parameters: $\alpha = 5, \beta = 2, \gamma = 0.5, \sigma = 0.1$. The results summarised in Figure 2 indicate good agreement between the estimated UKF path and the true cell path. The UKF also provides good estimates for the parameters: $\hat{\beta} = 1.88, \hat{\gamma} = 0.51, \hat{\sigma} = 0.04$ with relatively small standard errors: $0.17, 0.83, 0.06$, except for $\hat{\alpha}$ where the estimates indicate a more substantial deviation from the true parameter (bias: 0.44 and standard error is 0.35).

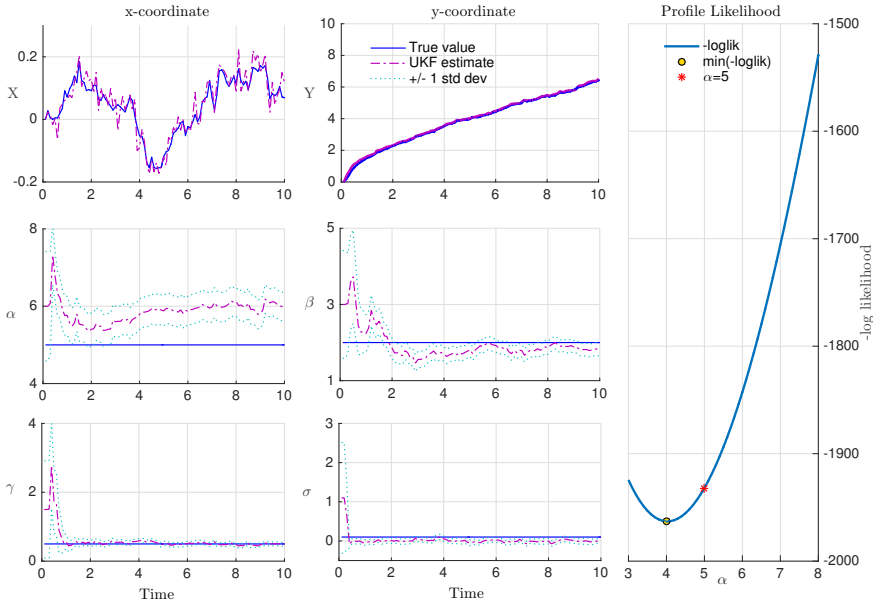

FIGURE 2. Simulation results: UKF tracking of cell coordinates: x and y, and parameters: $\alpha, \beta, \gamma, \sigma$ for time interval $[0, 10]$. Parameter estimates include ± 1 standard error bounds. On the right hand side, negative log profile likelihood plot for α, obtained by fixing the other three parameters at their true values.

A potential source of bias as the one observed in Figure 2 can be investigated by looking at the likelihood i.e.: marginal likelihood with respect to the hidden states. In order to do that, we first derive the probability of the observed system at time t conditional on the state of the system at time $t-1$ by integrating out the latent variable $\mathbf{x}_t$:

$$p(\mathbf{y}_t|\mathbf{y}_{t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{t-1})d\mathbf{x}_t \tag{7}$$

$$= \int \mathcal{N}(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t|\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t)d\mathbf{x}_t \tag{8}$$

Using the Gaussian convolution integral results (Bishop, 2006) we simplify (8) to $p(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_t|\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t + \mathbf{R}_t)$, where $\bar{\boldsymbol{\mu}}_t$ and $\bar{\boldsymbol{\Sigma}}_t$ are the predicted mean and covariance at time t, and $\mathbf{R}_t$ is the measurement noise covariance matrix at time t. The log likelihood is then:

$$\mathcal{L} = \log \prod_t p(\mathbf{y}_t|\mathbf{y}_{t-1}) = \sum_t \log \mathcal{N}(\mathbf{y}_t|\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t + \mathbf{R}_t) \tag{9}$$

$$\propto \sum_t \{\log \det(2\pi\boldsymbol{\Sigma}_t) + 0.5(\mathbf{y}_t - \bar{\boldsymbol{\mu}}_t)^T \boldsymbol{\Sigma}_t^{-1}(\mathbf{y}_t - \bar{\boldsymbol{\mu}}_t)\}, \tag{10}$$

Where $\bar{\mathbf{\Sigma}}_t + \mathbf{R}_t = \mathbf{\Sigma}_t$. We evaluate the marginal log likelihood in (10) by considering a grid of values for each parameter in the model and fitting the UKF with each parameter combination. The results summarising the profile likelihood for the α parameter in Figure 2 can be used to calculate the Cramer-Rao lower bound, which provides an indication of the intrinsic uncertainty specific to the problem. In this case, the minimum standard deviation attainable by an estimator of α is 0.14. Considering the standard error obtained from the UKF estimation for α is 0.35, this then indicates that the estimated value of the parameter is reasonably close to the true value.

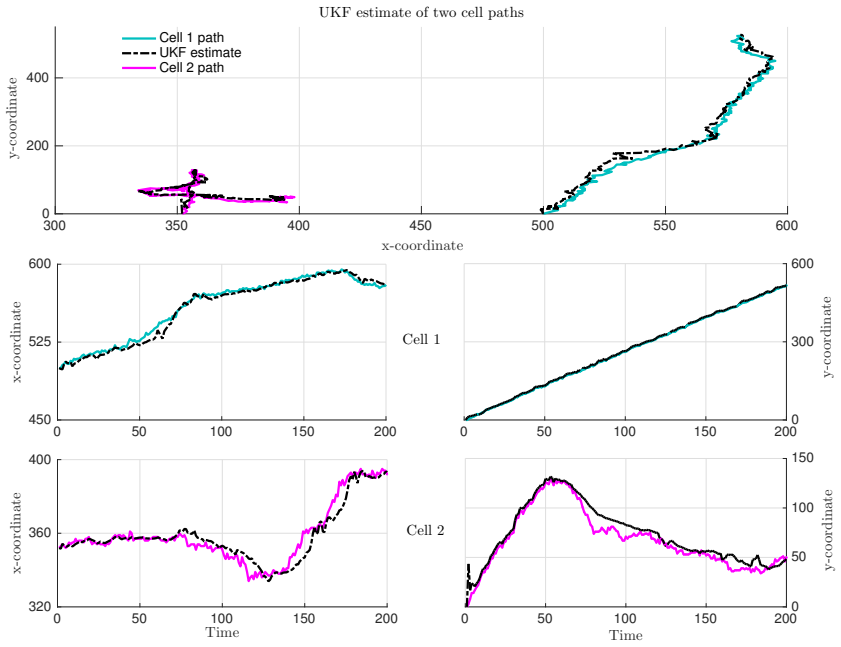

FIGURE 3. UKF tracking of two cells with different movement patterns.

4 Real data application

Dictyostelium cells are widely used in experiments as proxies for understanding the mechanisms of human disease because of their similarities to important human cells (leukocytes and cancer cells) in terms of biology and response to chemotaxis (Tweedy et al., 2016). The data consists of two cell paths corresponding to Dictyostelium cells locations extracted from a time series of high resolution microscopy images (see Figure 1). We emphasise that the main interest for biological applications is the inference of

the parameters. However, since the true parameters for the real data are unknown, we use the tracking of the cell trajectories as a proxy for assessing the accuracy of inference (see Figure 3). As can be seen from the left panel of Figure 3, we have picked two cells with very different behaviour (one dominated by drift, the other dominated by diffusion). In both cases, the path reconstructed with the UKF is very accurate.

5 Conclusions and future work

In this paper, we demonstrate the application of the UKF, a Bayesian filtering technique that adequately trades off accuracy versus computational efficiency, to a real-world problem potentially relevant to cancer research: the movement of *Dictyostelium* cells, which has not been tackled at individual cell level before. Our results indicate that the UKF can be successfully used for parameter inference and tracking cells displaying various movement patterns. Future work will extend this work by applying the UKF to a population of cells. Additionally, we plan to fit models describing alternative movement mechanisms, such as the self-induced gradient model described by Tweedy et al. (2016) and employ model selection criteria to choose the best model.

Acknowledgments: The research described in this article is part of the research programme of SoftMech, the Centre for multiscale soft tissue mechanics with application to heart & cancer, funded by the Engineering and Physical Sciences Research Council (EPSRC) of the UK, grant reference number EP/N014642/1.

References

Bishop, C. (2006) Pattern Recognition and Machine Learning. *Springer*

Golightly, A. & Wilkinson, D. J. (2011) Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface focus*, **1(6)**, 807 – 820

Julier, S. J. & Uhlmann, J. K. (1997) A New Extension of the Kalman Filter to Nonlinear Systems. *AeroSense'97*, 182 – 193

Sitz, A., Schwarz, U., Kurths, J. & Voss, H. U. (2002) Estimation of parameters and unobserved components for nonlinear systems from noisy time series *Phys. Rev. E*, **66(1)**, 016 – 210

Tweedy L., Knecht D.A., Mackay G.M. & Insall R.H. (2016) Self-Generated Chemoattractant Gradients: Attractant Depletion Extends the Range and Robustness of Chemotaxis. *PLoS Biol*, **14(3)**