Supplementary material: A stochastic model for topographically influenced cell migration

AJ Mitchinson[a,*], M Pogson[b], G Czanner[a], D Conway[c], RR Wilkinson[a], MF Murphy[d], I Siekmann[a], SD Webb[e]

[a] *School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool, L3 3AF, United Kingdom*
[b] *Department of Communication and Media, University of Liverpool, Liverpool, L69 7ZG, United Kingdom*
[c] *School of Biological Sciences, Queen's University Belfast, Belfast, BT9 5DL, United Kingdom*
[d] *School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, L3 3AF, United Kingdom*
[e] *Product Safety, Early Stage Research, Syngenta, Bracknell, RG42 6EY, United Kingdom*

1. Methods

1.1. Method of parameter estimation

We estimate model parameters using a grid search optimisation method. Let parameter space $P = (\alpha, \beta, \kappa) \in \mathbb{R}^3_+$. To reduce the search space, we use coarse grained grid searches and migration path length (derived from experimental measurements of cell speed, found here: [1]) to determine parameter boundaries within which migration lengths are, for this data-set, realistic. For each component of P, we define a lower bound $P_1 = (\alpha_1, \beta_1, \kappa_1)$ and an upper bound $P_n = (\alpha_n, \beta_n, \kappa_n)$, giving an interval within which to search e.g. $[\alpha_1, \alpha_n]$. We then discretise the interval into a finite set of m uniformly

*Corresponding author
Email address: A.J.Mitchinson@2013.ljmu.ac.uk (AJ Mitchinson)

spaced points producing a 'grid' of m^3 points over which to calculate an objective function, keeping m large whilst maintaining reasonable computation time [2].

We define our objective function as the nondimensional error function ϵ which, for a given parameter set $P_{a,b,c} = (\alpha_a, \beta_b, \kappa_c)$, is calculated by Eq. (1).

$$\epsilon = \sum_{i=1}^{N} \frac{(\theta_{i;sim} - \theta_{i;exp})^2}{\theta_{i;exp}^2}, \tag{1}$$

where $\theta_{i;exp}$ is the ith experimentally derived metric and $\theta_{i;sim}$ the ith simulated metric. N is total number of metrics. The goal is to minimise ϵ within our bounded discretised parameter space.

To illustrate the searching process algorithmically, for a given topography T_1 and parameter set e.g. $P_{a,b,c} = (\alpha_a, \beta_b, \kappa_c)$, we initiate a large number of model simulations (chosen arbitrarily but fixed, e.g. 1×10^3) and calculate migration metrics for the whole population of cells, each metric giving a distribution for the population. For this study, we use orientation angle, $\theta(°)$, and migration speed, s (μm/h) - for respective definitions see main article *Methods 2.2*, Eq. (4)-(6). We then use distribution statistics derived from experimental study and model simulations to obtain a value for ϵ, which is stored. We then alter the parameter set, e.g. $P_{a,c,c} = (\alpha_a, \beta_c, \kappa_c)$, and initiate a new simulation, iterating systematically in this manner through a predefined range for each parameter. A flow chart illustrating the algorithmic approach is given in Figure 1, where the lower and upper bounds for the parameter space are $P_1 = (\alpha_1, \beta_1, \kappa_1)$ and $P_n = (\alpha_n, \beta_n, \kappa_n)$ respectively.

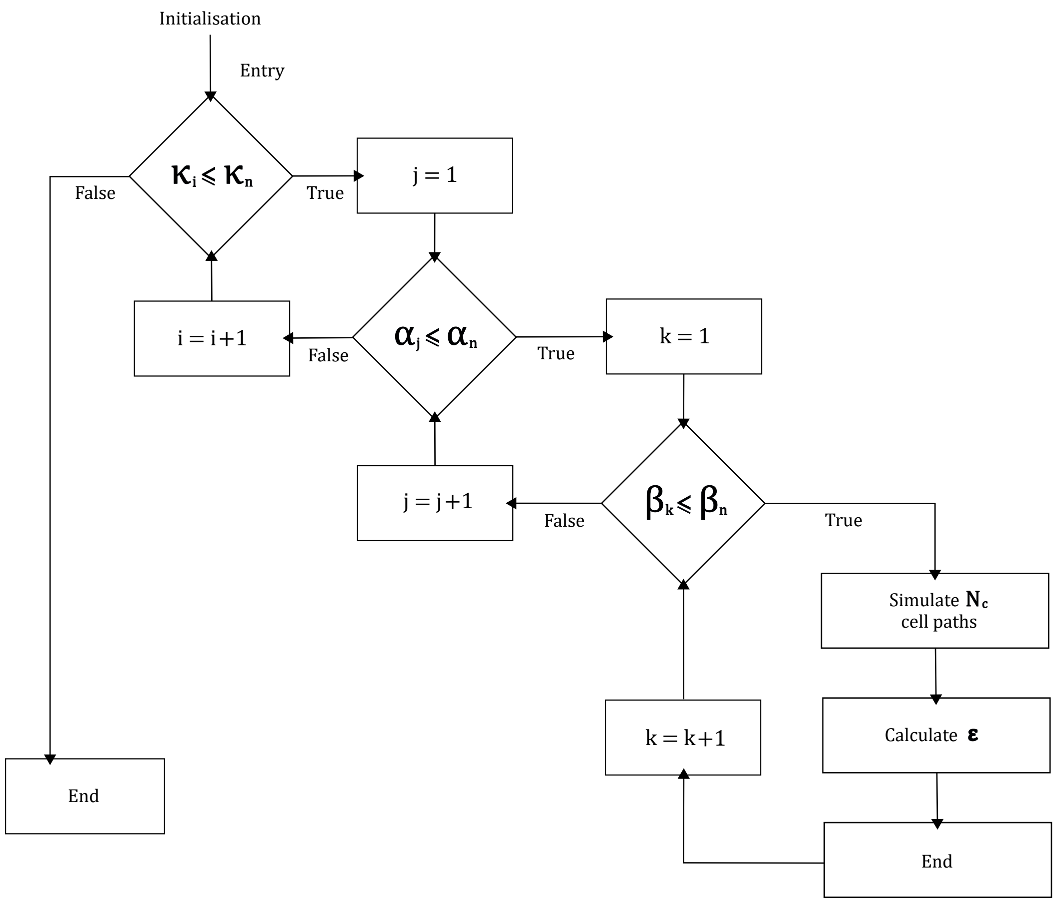


Figure 1: Flow chart to illustrate the algorithmic approach chosen to conduct the parameter grid search for the model. The general approach taken is to hold two parameters constant whilst iterating through one parameter range, simulating N_c cell paths and calculating ϵ for each individual parameter set, before incrementally adjusting the originally held parameters and repeating. To illustrate, in the initial state of the algorithm κ_1 and α_1 are held constant whilst migration paths and subsequently ϵ are simulated for $\beta_1, ..., \beta_n$. After the model simulation for β_n, the value for α changes, from α_1 to α_2, with κ_1 held constant whilst we again repeat simulations for $\beta_1, ..., \beta_n$, and so on until κ_n is reached and the algorithm ends. The result is a value for ϵ for each individual parameter combination resulting from discretisation of the interval $[P_1, P_n]$.

1.2. Experimental data

1.2.1. Fibroblast migration behaviour for a linearly ridged topography with variable ridge density (μm-scale)

To parametrise the model we use metric data extracted from an experimental study published in the journal *Biomaterials* (for details see [1]). The study probed NIH3T3 fibroblast migration on anisotropic substrata with precisely fabricated linear topographic features created using capillary force lithography (CFL). The method produced a surface pattern of alternating parallel ridges and grooves with constant depth (400nm) and ridge width (1μm) and variable ridge spacings from 1μm to 9.1μm, spaced in increasing 100nm increments from densely to sparsely spaced ridges. Cells were seeded at low density onto this surface topography to enable individual tracking. Fluorescent microscopic images were taken every 15 minutes over 12 hours to produce a time-lapse sequence within which to track the paths each cell would follow through time.

To quantify the orientation of cell movement compared to the direction of linear topographic surface features, at 14 hours post-culture the authors measured the acute angle between the longest axis of the cell, the 'polarised' cell direction, and groove direction, generating a distribution of 'polarisation angles' for cells across the variable groove widths on the substratum. The authors also calculated migration speed between increments for each cell. Speeds for each cell were averaged, these average cell speeds were then averaged by substratum position to give a single average speed for the local population.

The authors found groove-oriented migration was more pronounced in substratum regions of higher ridge density, prompting more linear migration paths and smaller standard deviations for polarisation angle distributions. Migration also showed a discernible preference towards intermediate groove widths, where average migration speed was highest.

We use metric statistics from the study, polarisation angle standard deviation, θ^*_σ ($^\circ$), and average migration speed, s^*_μ (μm/h), in the grid search optimisation calculation to parametrise our model. In lieu of explicit data values, we estimate values directly from the study figures which display θ^*_σ ($^\circ$) and s^*_μ (μm/h) for flat and grooved areas of the topography using a pixel measurement tool. We present estimated values in Table 1.

Average groove width	θ^*_σ ($^\circ$)	s^*_μ (μm/h)
Flat (no gradient)	47	28
8.6μm	38	29
6.3μm	20	40
2.6μm	12	34

Table 1: Estimated migration metric data, θ^*_σ and s^*_μ, for flat and linearly ridged/grooved (with average groove widths: 8.6μm, 6.3μm and 2.6μm) topographies from Kim et al. [1].

1.3. Migration metrics

(i) **Orientation angle**. In the study by Kim et al., the authors approximate cell direction by measuring the acute angle between the long axis of a cell and groove direction, taken as a single measurement for each cell at the end of a time course, this termed the 'polarisation angle', θ^*. The values θ^* for every cell in a given locale were accumulated to give a distribution of polarisation angles for different regions of the substrate, from which a distribution mean, θ^*_μ, and standard deviation, θ^*_σ, were calculated.

We replicate this for the migration model by introducing an analogous angle metric defined as the argument between cell velocity direction and groove direction, termed 'orientation angle', θ. In-keeping with the computation of θ^*, we set the calculation symmetric about directions orthogonal to groove direction $\boldsymbol{L}$ and determine the position of $0°$ to be at both opposing groove directions $\boldsymbol{L}$ and $-\boldsymbol{L}$. We measure θ with positive angles clockwise from the groove direction, keeping the angle range acute, $-90° \leq \theta \leq 90°$ (see Figure 2).

(ii) **Migration speed**. The authors of the Kim et al. study calculated migration speed from point-to-point cell trajectories tracked through a time-lapse sequence, giving a sequence of point-to-point speeds for each cell over time (9 hours at 15 minute intervals). The sequence of speeds for each cell was then averaged, and cells grouped by substratum position (average groove width) to give distributions by 'average groove width' from which an average migration speed, s^*_μ, was calculated for each.

To replicate the calculation of s^*_μ, we compute migration speed s from individual cell displacements as with orientation angles, between increments j and $j+1$ for every increment for each cell i in a given simulation, to give a distribution of migration speeds, $s_{ij,j+1}$, from which we calculate the mean migration speed, s_μ.

(iii) **Mean-squared displacement (MSD)**. We define the MSD $< D^2 >$ as the squared distance travelled by each cell during time interval t, summed and averaged over the total number of cells N_c to migrate during that time interval, given by Eq. (2).

$$< D^2 >= \frac{1}{N_c}\sum_{i=1}^{N_c}[\boldsymbol{x}_i(t) - \boldsymbol{x}_i(0)]^2, \tag{2}$$

where $\boldsymbol{x}_i(t)$ is the position of the cell i at time t and $\boldsymbol{x}_i(0)$ is the position at the start of displacement.

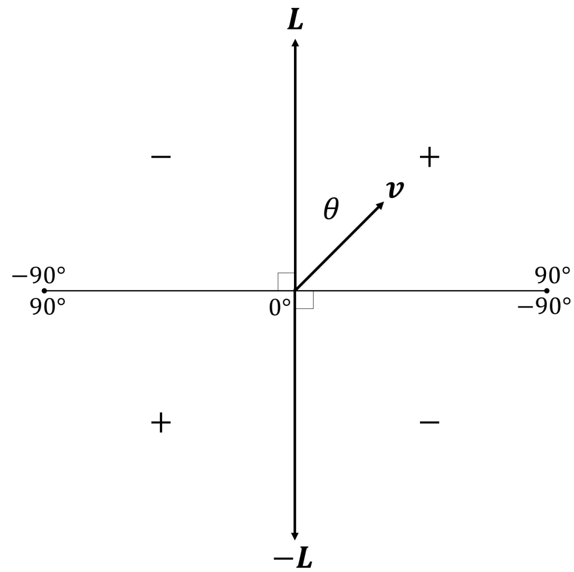


Figure 2: Schematic diagram to illustrate the measurement of 'orientation angle', θ, for a sample time increment. θ is measured as the argument between cell velocity $\boldsymbol{v}$ and groove direction $\boldsymbol{L}$ or $-\boldsymbol{L}$, dependent on the sign of the 'vertical' component $\boldsymbol{v}$. The calculation is symmetric about the directions orthogonal to groove direction $\boldsymbol{L}$, $0°$ at both opposing groove directions $\boldsymbol{L}$ and $-\boldsymbol{L}$. θ is measured with positive angles clockwise from the groove direction, in the range $-90° \leq \theta \leq 90°$.

1.4. Numerical implementation

1.4.1. Topography generation

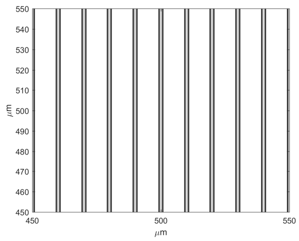
To create linear topographies with topographic features comparable to those featured in the experimental study from which we extrapolate data (see *1.2.1*), we generate simulated approximations with matched linear feature densities using MATLAB. We then test how closely each simulated topography can approximate corresponding migration metrics compared to their matched experimental topographies during the fitting procedure.

The general numerical approach we take to generate the topographies is grid-based. We define a 'substrate' matrix, tracing the domain boundaries for the topography, and assign a 'depth' value to relevant indices in the matrix corresponding to ridge height. To approximate a 'flat' topography (i.e. to cell-scale, no significant physical gradients present) we idealise and assume no physical gradients are present on the surface, using simply a matrix with homogeneous depth values. To approximate linear topographies we use the approximation for the flat topography, and, at uniform intervals across all columns in the matrix, assign depth values for all indices in each selected column, generating linear topographic features up to the domain boundary. To mimic the dimensions of the experimentally produced topographies in the study by Kim et al. (see *1.2.1*), and for simplicity, we choose only one depth value thereby producing a binary matrix in which one number represents ridge features and the other number groove features each of uniform height and depth respectively. To create each of the different linear topographies, we vary only ridge density by adjusting their spacing within the matrix.

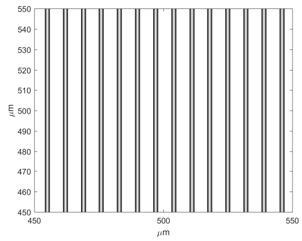
We assign spatial units based on fitting the model migration trajectory metrics on trial surfaces to metrics from [1], adjusting dimensions as necessary and ensuring boundaries are large enough to accommodate the trajectory range. Spatial units were assigned $1 \times 1\mu\text{m}^2$ to one matrix index. To approximate topography dimensions in the experimental study, we set ridge spacings on three separate topographies to two, six and nine matrix indices (corresponding to 2μm, 6μm and 9μm), ridges to one matrix index width (1μm), assigning a uniform depth of 0.4μm and matrix dimensions 1000×1000 indices ($1000 \times 1000\mu\text{m}^2$). The result, presented in Figure 3 (a)-(c), is a set of three linearly organised topographies with (a) sparse, (b) intermediate and (c) high density linear features, with uniform width and depth features.

To probe in a general manner how imprecise surface processing might affect migration behaviour, we devise a method to progressively introduce 'noise' to already generated linear topographic features. Our approach is to incrementally perturb with additive noise the linear feature in the plane orthogonal to its long axis direction. The method we use is to draw new index locations for the ridge feature from a Gaussian distribution, the mean centred on the axis of the linear feature. We vary the level of 'randomness' around the linear feature with the distribution variance, ρ. When $\rho = 0$ the arrangement of surface gradients are perfectly linear without noise, increasing ρ introduces higher levels of randomness to the feature, making it more 'distorted' and less linear.

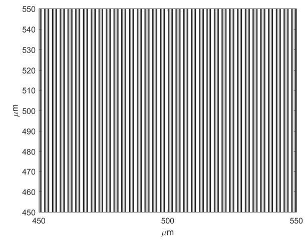
We do this numerically using MATLAB's pseudo-random number generator '*randn*' and round to the nearest integer for index values to assign a depth value. We use the same method across the three topographies shown in Figure 3 (a)-(c) to incrementally distort their linear features, keeping the range for ρ consistent (rather than dependent on feature density). The result, presented in Figure 3 (a)-(o), is a set of topographies with sparse (left column), intermediate (middle column) and high feature densities (right column) which range in organisation from parallel uniform linear features (a)-(c) ($\rho = 0$) to disordered randomly arranged features (m)-(o) ($\rho = 10$).



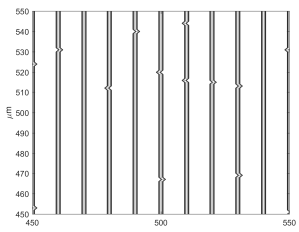
(a) 9μm spacing. $\rho = 0$.



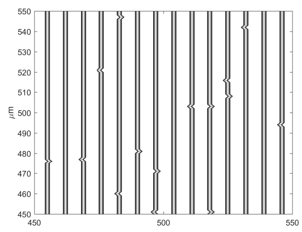
(b) 6μm spacing. $\rho = 0$.



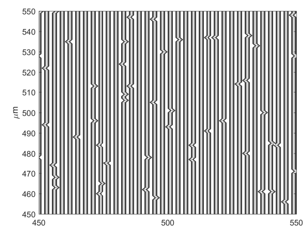
(c) 2μm spacing. $\rho = 0$.



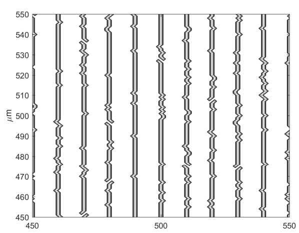
(d) 9μm spacing. $\rho = 0.2$.



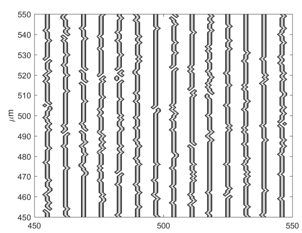
(e) 6μm spacing. $\rho = 0.2$.



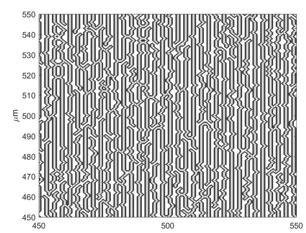
(f) 2μm spacing. $\rho = 0.2$.



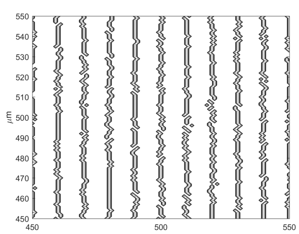
(g) 9μm spacing. $\rho = 0.35$.



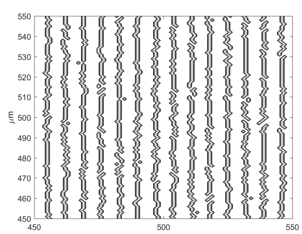
(h) 6μm spacing. $\rho = 0.35$.



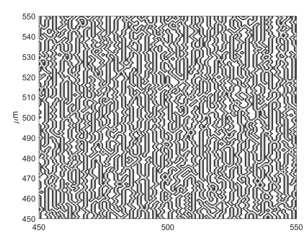
(i) 2μm spacing. $\rho = 0.35$.



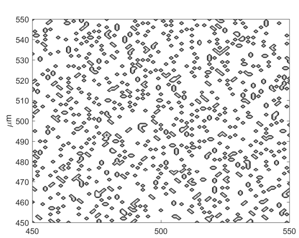
(j) 9μm spacing. $\rho = 0.5$.



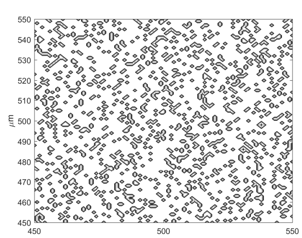
(k) 6μm spacing. $\rho = 0.5$.



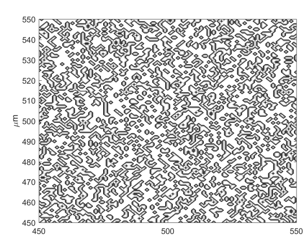
(l) 2μm spacing. $\rho = 0.5$.



(m) 9μm spacing. $\rho = 10$.



(n) 6μm spacing. $\rho = 10$.



(o) 2μm spacing. $\rho = 10$.

Figure 3: Surface topographies generated using MATLAB, featuring sparse (left), intermediate (middle) and high feature densities (right) ranging in organisation from parallel uniform linear features (a)-(c) through increasing levels of additive feature noise, determined by distortion parameter ρ (rows), to disordered randomly arranged features (m)-(o). Each domain is a $100 \times 100\mu\text{m}^2$ perspective of the topography.

2. Results

2.1. Parameter estimation

To estimate individual parameter combinations for the flat topography we fit a polynomial function to an identified region of minima and use the fitted function to approximate values for β and α. We first define a region of minima as that where the error ϵ is sufficiently small (we choose the threshold $\epsilon \leq 0.03$ to constrain each parameter space, the choice being otherwise arbitrary). We then identify mid-point locations of the region across β, which we interpret as approximate minima (blue regions in main article Figure 2 (a)) and, excluding clear outliers, fit an appropriate polynomial function (quartic) to the set of approximate minima using a numerical fitting tool, (MATLAB's *fit* function). We see in the main article Figure 2 (a) (blue line), the method captures the major β-α relationship present at the region of minima and yields reasonable approximate parametrisations for model output. We present the polynomial function through minima in main article Figure 2 (a), $\hat{f}$, in Eq. (3).

$$\hat{f}(\beta) = -6.102\beta^4 + 11.79\beta^3 - 5.577\beta^2 + 2.189\beta - 0.1469, \tag{3}$$

where β is a model parameter, and the range $\hat{f}(\beta)$ gives an approximation for α at minima, over the approximate domain $0.06 < \beta < 1.24$.

By contrast, in main article Figure 2 (b)-(d), we see clearly identifiable parameter combinations for β and α at given κ values for each of the linear topographies. This persists through ranges for κ for each of the linear topographies (results not shown). Generally, the ranges for β and α over which these minima occur through κ for the linear topographies are significantly smaller than those for the flat topography. To estimate individual parameter combinations for these topographies, we constrain $\epsilon \leq 0.025$ and take the median β value, β_η, over the resulting region of minima, choosing α at an arbitrary minimum for β_η.

In main article Figure 2 (b), we see minima (for which $\epsilon \leq 0.03$, arbitrarily) occur over the approximate ranges $0.07 < \beta < 0.23$ and $0.003 < \alpha < 0.1$ for the 9μm groove width topography at $\kappa = 1$. Minima persist through an approximate range $0.02 < \kappa < 10$ (results not shown). In main article Figure 2 (c) we see minima occur over the approximate ranges $0.03 < \beta < 0.079$ and $7.58 \times 10^{-4} < \alpha < 0.01$ for the 6μm groove width topography at $\kappa = 0.75$. Minima persist through an approximate range $0.15 < \kappa < 5$ (results not shown). In main article Figure 2 (d), we see minima occur over the approximate ranges $0.07 < \beta < 0.13$ and $2 \times 10^{-3} < \alpha < 0.01$ for the 2μm groove width topography at $\kappa = 0.5$. Minima persist through an approximate range $0.15 < \kappa < 5$ (results not shown).

References

[1] D.-H. Kim, K. Han, K. Gupta, K. W. Kwon, K.-Y. Suh, A. Levchenko, Mechanosensitivity of fibroblast cell shape and movement to anisotropic substratum topography gradients, Biomaterials 30 (29) (2009) 5433–5444.

[2] K. B. Ensor, P. W. Glynn, Stochastic optimization via grid search, Lectures in Applied Mathematics-American Mathematical Society 33 (1997) 89–100.