

2 **Article**

3 **Quantifying the impact of electric fields on single-cell** 4 **motility**

5 TP Prescott, K Zhu, M Zhao, and RE Baker

6 Affiliation list placeholder

7 **ABSTRACT** Cell motility in response to environmental cues forms the basis of many developmental processes in multicellular
8 organisms. One such environmental cue is an electric field (EF), which induces a form of motility known as electrotaxis.
9 Electrotaxis has evolved in a number of cell types to guide wound healing, and has been associated with different cellular
10 processes, suggesting that observed electrotactic behaviour is likely a combination of multiple distinct effects arising from the
11 presence of an EF. In order to determine the different mechanisms by which observed electrotactic behaviour emerges, and
12 thus to design EFs that can be applied to direct and control electrotaxis, researchers require accurate quantitative predictions of
13 cellular responses to externally-applied fields. Here, we use mathematical modelling to formulate and parametrise a variety of
14 hypothetical descriptions of how cell motility may change in response to an EF. We calibrate our model to observed data using
15 synthetic likelihoods and Bayesian sequential learning techniques, and demonstrate that EFs bias cellular motility through only
16 one of a selection of hypothetical mechanisms. We also demonstrate how the model allows us to make predictions about cellular
17 motility under different EFs. The resulting model and calibration methodology will thus form the basis for future data-driven and
18 model-based feedback control strategies based on electric actuation.

SIGNIFICANCE Electrotaxis is attracting much interest and development as a technique to control cell migration due to the precision of electric fields as actuation signals. However, precise control of electrotactic migration relies on an accurate model of how cell motility changes in response to applied electric fields. We present and calibrate a parametrised stochastic model that accurately replicates experimental single-cell data and enables the prediction of input–output behaviour while quantifying uncertainty and stochasticity. The model allows us to elucidate and quantify how electric fields perturb the motile behaviour of the cell. This model and the associated simulation-based calibration methodology will be central to future developments in the control of electrotaxis.

19 INTRODUCTION

20 Cell migration underpins key physiological processes central to developmental biology, as well as wound healing and tissue
 21 regeneration, and it plays a crucial role in invasive, metastatic cancers. There are ongoing efforts to intervene in and influence
 22 these phenomena to, for example, inhibit metastasis (1) or accelerate wound healing (2). However, the cellular processes driving
 23 collective migration are complex and multifaceted, deriving from diverse physical mechanisms and various external stimuli (3),
 24 making it challenging for researchers to accurately and robustly direct cell motility. Due to the ease with which electric fields
 25 can be controlled and applied to cells, research into the control of cell motility has recently focused on exploiting *electrotaxis*
 26 (also known as galvanotaxis) (3–5). However, the precise effects of electric fields on intracellular processes and thus on cell
 27 motility are not fully understood, making quantitative predictions and control policy design impractical.

28 Electrotactic cells have been observed to change their motile behaviour in response to the presence of a direct current (DC)
 29 electric field (EF) (3–7). Researchers seeking to control cell motility exploit this phenomenon by applying external electrical
 30 cues to cell populations (2, 4–9). The key advantages of using electrical cues to guide cell migration include the ability to exploit
 31 endogenous, evolved biological functionality to respond to precisely controllable DC EFs. This compares favourably to using
 32 chemoattractants to guide motility, since chemical signals experienced by the cell cannot be so precisely or flexibly controlled,
 33 especially dynamically, and chemoattractants are usually highly cell-specific. In contrast, light-directed motility allows for
 34 precise actuation signals. However, it requires sophisticated optogenetic manipulations of the cell population under control (10).
 35 As such, EFs provide a relatively precise and simply implemented actuation signal to achieve specified motile behaviours.

36 While an important strength of electrotactic cell control is that applying an EF for actuation is flexible enough to apply to
 37 any electrotactic cell type, the precise signal to be applied in order to achieve any specified goal needs to be carefully calibrated.
 38 At the most basic level, even the direction of migration within the same DC field has been shown to vary across different cell
 39 types, and within one cell type under different experimental conditions (5, 11). More broadly, a large number of biochemical
 40 and biophysical mechanisms have been implicated in the electrotactic response across different cell types (3). Each electrotactic
 41 mechanism, which may co-exist in combination at unknown relative strengths, may induce distinct observable effects on the
 42 dynamics of cellular motility. Overcoming this uncertainty in the observable electrotactic response is a fundamental challenge
 43 for designing EFs to control cell motility.

44 Mathematical models are a vital tool for quantifying the different ways in which cells can change their motility in response
 45 to EFs (12–14). In this paper we describe a parametrised stochastic model of the motile behaviour of a single human corneal
 46 epithelial cell, in which the cell's motility is driven by an internal polarity, in combination with the external influence of a DC
 47 EF. We assume that the cell can undergo both spontaneous and electrotactic polarisation. The model allows us to describe
 48 mathematically four distinct ways an EF may influence motility. We use experimentally observed trajectories of single cells,
 49 both with and without applied EFs, to calibrate the parameters of this model, thereby quantifying the extent to which different
 50 aspects of cell motility are impacted by the EF. The resulting calibrated model provides a vital first step towards being able to

design feedback control policies and provide robustness guarantees, which are necessary if electrotaxis is to be used to control cell motility in practical applications such as wound healing or tissue engineering.

Single-cell modelling

The agent-based modelling framework used in this work follows standard modelling assumptions outlined in (13). Specifically, we model the evolution of the velocity of a single cell in the overdamped regime, so that cell velocity is proportional to the sum of non-frictional forces on the cell. We provide full details on the mathematical model in *Materials and Methods* and in the *Supplementary Material*.

In the absence of any EF, the only non-frictional force acting on the cell is assumed to be an active force arising from the internal *polarity* of the cell. Thus, the cell velocity, $\mathbf{v} = \mathbf{v}_{\text{cell}}$, is comprised of a single component. A preliminary analysis of single-cell motility data, described more fully in the *Supplementary Material*, suggests that the cell velocity arises from a cell being *polarised* in a particular direction, and that the direction of polarisation drifts stochastically over time. A polarised cell has a positive speed, parametrised by a modal value, $\|\mathbf{v}_{\text{cell}}\| \approx v$, where the scalar-valued parameter $v > 0$ has dimensions $\mu\text{m min}^{-1}$. In addition to random changes in cell speed, preliminary analysis also suggests that the direction of cell motion stochastically evolves according to a persistent random walk, such that the autocorrelation between displacement directions decays as the time lag increases. Thus, the direction of cell motion (in the absence of an EF) is assumed to vary according to an unbiased random walk with positive timescale constant $D > 0$, with dimensions min^{-1} , which characterises the rate of decay in the autocorrelation of the polarisation direction over time. Eq. (2) in *Materials and Methods* provides the mathematical formulation of this model.

We hypothesise that a vector-valued DC EF, $\mathbf{u}$, can affect cell motility in a variety of ways. We use a number of extensions of the model in order to implement different ways in which motility may be impacted by the EF, specifying, in particular, four distinct ways in which it may affect the dynamics of a motile cell. We parametrise the magnitude of each hypothesised electrotactic effect, observed at a reference EF strength of 200 mV mm^{-1} , by the parameters γ_1 , γ_2 , γ_3 and γ_4 , such that if $\gamma_i = 0$ then the corresponding hypothesised effect is not included in the model. Eq. (3) in *Materials and Methods* provides the mathematical formulation of this model.

The four means by which we model cell motility to be perturbed by the EF are:

Velocity bias (γ_1) The EF imparts an additional component of force on the cell. The resulting velocity, $\mathbf{v} = \mathbf{v}_{\text{cell}} + \mathbf{v}_{\text{EF}}$, is thus the sum of two components: the original polarity component, $\mathbf{v}_{\text{cell}}$, and an EF component, $\mathbf{v}_{\text{EF}}$. The EF velocity component acts in the direction of the field with magnitude $\gamma_1 v$.

Speed increase (γ_2) Polarised cells travel more quickly under the influence of an EF in the direction in which they are polarised. The modal magnitude of $\mathbf{v}_{\text{cell}}$ for polarised cells is increased by $\gamma_2 v$.

Speed alignment (γ_3) Polarised cells travel more quickly when the direction of their polarisation aligns with the EF, but slower if opposed to the EF. The modal magnitude of $\mathbf{v}_{\text{cell}}$ for polarised cells is increased by $\gamma_3 v \cos(\theta)$, where θ is the angle between $\mathbf{v}_{\text{cell}}$ (i.e. the polarity direction) and the EF direction.

Polarity bias (γ_4) The random walk determining cell polarity is biased so that cells preferentially polarise in the direction of the EF. The strength of this bias is parametrised by γ_4 .

Two models can be distinguished: the *autonomous* model, where no EF is applied, and the *electrotactic* model, where a reference strength EF is applied. In each of these models, the cell velocity at time t , denoted $\mathbf{v}(t)$, undergoes a random walk. Figure 1 characterises each of these models by depicting the stationary probability distribution of this random walk. The top plot shows that without an applied EF the modal cell speed is near v , with direction chosen uniformly at random. The bottom plot of this figure demonstrates how each electrotactic effect, quantified by the value of γ_i for $i = 1, 2, 3, 4$, can be interpreted in terms of the probability distribution of the cell velocity: γ_1 translates the velocity distribution uniformly in the direction of the field; γ_2 rescales the domain of the distribution; γ_3 parametrises asymmetry in the shape of the velocity distribution; and γ_4 parametrises asymmetry in the density of the velocity distribution.

[Figure 1 about here; moved to end of manuscript by endfloat.]

Outline

The primary goal of this work is to use single-cell experimental data to calibrate the parametrised mathematical model of spontaneous polarisation and electrotaxis. The model calibration process enables the identification of which of the four hypothesised electrotactic effects of EFs on cell motility can be observed in the experimental data. Importantly, the calibrated model also quantifies the relative contribution of each of these identified effects. The level at which we model the system allows us to subsequently use the calibrated model to simulate and predict the single-cell response to dynamic EFs.

The data used for model calibration is gathered from two assays in which the trajectories of motile human corneal epithelial cells are recorded for five hours: (a) without any EF applied for the entire experiment, and (b) with a DC EF at a reference strength of 200 mV mm^{-1} , applied from left to right over hours 2–3 and from right to left over hours 4–5. These assays are termed the *autonomous* and *electrotactic* experiments, respectively. We use all five hours of the autonomous experiment and the first three hours of the electrotactic experiment as training data to calibrate the parameters of the autonomous and the electrotactic models. To calibrate the electrotactic model, we first identify which combination of the four hypothesised electrotactic effects is best supported by the data. After identifying which of the electrotactic effects are present in the model, we can then proceed to quantify the relative contribution of each of them to the observed electrotaxis induced by the EF.

After formulating and calibrating the extended model of electrotaxis, we use simulations of the calibrated model to predict how the cell trajectories evolve over the final two hours of the electrotactic experiment, in which the EF input has changed

direction. We compare these predictions to the cell trajectories observed over the final two hours of the electrotactic experiment, held back to be used as test data, and thus validate the predictive capability of the model for dynamic EF inputs. The ability to make predictions of cellular motility using a calibrated, stochastic, uncertain model is a first step towards the future goal of model-based policy design for the electrotactic control of single-cell and population-level motility.

MATERIALS AND METHODS

Data collection

Two experiments were carried out, which we call the *autonomous* and *electrotactic* experiments. In both experiments, time-lapse images of human corneal epithelial cells, seeded at a low density, were acquired at 5 min intervals over 5 h. In the autonomous experiment, no EF was applied. In the electrotactic experiment, the cells were subjected to a DC EF at a reference strength, 200 mV mm⁻¹, applied across the medium from $t = 60$ min to the end of the experiment. The EF was directed from left to right from 60 min to 180 min, at which point the field direction was reversed from right to left for 180 min to 300 min. Two replicates of each experiment were performed, with 26 and 27 cell centroids tracked for each of the autonomous assay replicates, and with 26 and 30 cell centroids tracked for each of the electrotactic assay replicates, all over the entire time horizon. Visual confirmation from the raw experimental output confirms that cell collisions were rare, due to the low density (100 cells cm⁻²) at which cells are initially seeded. We thus assume that cell-cell interactions can be neglected in the current model. The data used in this work is shared online at [DOI:10.5281/zenodo.4749429](https://doi.org/10.5281/zenodo.4749429).

We denote the resulting cell trajectory data $\mathbf{x}_{\text{NoEF},i}(t_k)$ and $\mathbf{x}_{\text{EF},j}(t_k)$ for the autonomous and electrotactic experiments, respectively, where each trajectory is translated to begin at the origin, such that $\mathbf{x}_{\text{NoEF},i}(0) = \mathbf{x}_{\text{EF},j}(0) = 0$ for all i and j . For these experiments, the indices $i = 1, \dots, 53$ and $j = 1, \dots, 56$ refer to the cell being traced, while $t_k = 5k$ min refers to the snapshot time points for $k = 0, \dots, 60$. We hold back $\mathbf{x}_{\text{EF},j}(t_k)$ for $j = 1, \dots, 56$ and $k = 36, \dots, 60$ as test data, for the purposes of validating model predictions. The remaining data is used as training data, from which the model is calibrated. Thus, the training data consists of the trajectories from the autonomous experiment over the entire time horizon, and the trajectories from the electrotactic experiment over 0 min to 180 min, which are denoted by $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, respectively. The test data, denoted $\mathbf{x}_{\text{test}}$, consists of all trajectories from the electrotactic experiment over 180 min to 300 min, where the input EF has switched direction.

Materials

EpiLife culture medium with Ca²⁺ (60 μM), EpiLife defined growth supplement, and penicillin/streptomycin were purchased from ThermoFisher Scientific (Waltham, MA, USA). FNC Coating Mix was purchased from Athena Enzyme Systems (Baltimore, MD, USA). Dow Corning high-vacuum grease was purchased from ThermoFisher. Agar was purchased from MilliporeSigma (Burlington, MA, USA). Silver wires with 99.999% purity were purchased from Advent Research Materials

140 Ltd. (Oxford, United Kingdom).

141 Cell culture

142 Telomerase-immortalized human corneal epithelial cells (hTCEpi) were routinely cultured in EpiLife medium supplemented
143 with EpiLife defined growth supplement and 1% (v/v) penicillin/streptomycin. Cells were incubated at 37 °C with 5% CO₂ until
144 they reached ~70% confluence and were used between passages 55 and 65 for all cell migration assays.

145 Electrotaxis assay

146 Electrotaxis experiments were performed as previously described (15, 16) with minor changes. Briefly, the electrotaxis
147 chambers (20 mm × 10 mm × 0.2 mm) were constructed in 100 mm petri dishes with glass strips and high-vacuum grease.
148 The dimensions of the chambers were defined by the thickness and length of the glass slides, respectively. Chambers were
149 coated with FNC Coating Mix, following the manufacturer's instructions to facilitate cell attachment. Cells were seeded at a
150 low density (100 cells cm⁻²) and cultured overnight (12 h to 18 h) in the chambers to allow sufficient attachment. Chambers
151 were covered with glass coverslips and sealed with high-vacuum grease. Electric currents were applied to the chamber through
152 agar-salt bridges connecting with silver–silver chloride electrodes in Steinberg's solution (58 mM NaCl, 0.67 mM KCl and
153 0.44 mM Ca(NO₃)₂, 1.3 mM MgSO₄ and 4.6 mM Tris base, pH 7.4). Fresh cell culture medium (EpiLife) was added into
154 reservoirs to ensure good salt bridge contact and to support cell viability during electric stimulation. An EF strength of
155 200 mV mm⁻¹ was used unless otherwise noted. A pair of measuring electrodes was placed at the end of the electrotaxis
156 chamber and connected to the multimeter for real-time monitoring of EF strength. The applied voltages were confirmed at the
157 beginning of the experiment and every 30 min afterwards to ensure consistent EF application.

158 Time-lapse imaging and quantification of cell migration

159 Cell migration was monitored and recorded by phase-contrast microscopy using an inverted microscope (Carl Zeiss, Oberkochen,
160 Germany) equipped with a motorized stage and a regular 10× objective lens. Time-lapse images were acquired at 5 min intervals
161 using Metamorph NX imaging software (Molecular Device, Sunnyvale, CA, USA). To maintain standard cell culture conditions
162 (37 °C, 5% CO₂), a Carl Zeiss incubation system was used. Time-lapse images of cell migration were analyzed by using ImageJ
163 software from the National Institutes of Health (<http://rsbweb.nih.gov/ij/>). Adherent cells in the images were manually
164 tracked, and cells that divided, moved in and out of the field, or merged with other cells during the experiment were excluded
165 from analysis. The position of a cell was defined by its centroid.

166 Model construction

167 We constructed a mathematical model of single-cell dynamics. The model tracks the position of the cell centre in the plane,
 168 $\mathbf{x}(t) \in \mathbb{R}^2$, as a function of time, $t \geq 0$ min, with initial condition $\mathbf{x}(0) = 0$ at the origin. The position is a deterministic integral
 169 of cell velocity, $\mathbf{v}$, such that

$$\mathrm{d}\mathbf{x}(t) = \mathbf{v}(t) \, \mathrm{d}t, \quad (1)$$

170 and the stochastic dynamics of $\mathbf{v}$ are modelled. The key to this modelling task is the non-dimensional internal variable
 171 representing the cell polarity, $\mathbf{p}(t) \in \mathbb{R}^2$. We assume that the polarity imparts a force on the cell that corresponds to its active
 172 motility, resulting in a velocity component $\mathbf{v}_{\text{cell}}(t)$.

173 Modelling spontaneous polarisation and motility

174 We first describe the model of cellular motility with no biasing EF, which we will term the *autonomous* model. The only velocity
 175 component is that due to polarisation, so that we write the cell velocity as a single component,

$$\mathbf{v}(t) = \mathbf{v}_{\text{cell}}(t) = v\mathbf{p}(t), \quad (2a)$$

176 where the parameter $v \geq 0$, with dimensions $\mu\text{m min}^{-1}$, represents the modal magnitude of $\mathbf{v}_{\text{cell}}$ for a polarised cell. Note that
 177 Eq. (2a) implies that the polarity variable, $\mathbf{p}$, is a non-dimensionalisation of the velocity component $\mathbf{v}_{\text{cell}}$. We further assume
 178 that the polarity, $\mathbf{p}$, undergoes a random walk according to a Langevin diffusion, such that

$$\mathrm{d}\mathbf{p}(t) = -D\nabla W(\mathbf{p}(t)) \, \mathrm{d}t + \sqrt{2D} \, \mathrm{d}\mathbf{B}, \quad (2b)$$

179 where $\mathbf{B}(t) \in \mathbb{R}^2$ is a two-dimensional Wiener process, and the parameter D (in min^{-1}) quantifies the speed at which the random
 180 walk approaches stationarity. The initial polarity, denoted $\mathbf{p}_0 = \mathbf{p}(0)$, also needs to be specified.

181 The *potential function* $W(\mathbf{p})$ in Eq. (2b) is defined to capture the intended features of the autonomous model, namely that
 182 the magnitude of the cell velocity is randomly distributed around a modal value of v , and that the direction of the polarity is
 183 uniformly distributed at stationarity. It can be shown (17, 18) that the variability of the velocity around its modal value of v is
 184 determined by a non-dimensional *energy barrier*, denoted ΔW , that is sufficient to define the potential function, $W(\mathbf{p})$. For
 185 further details on the definition of W , see the *Supplementary Material*. We will calibrate the autonomous model in Eq. (2) by
 186 identifying the parameters v , D , and ΔW .

Modelling motility bias due to an EF

We use a vector-valued function, $\mathbf{u}(t)$, with non-dimensional magnitude $\|\mathbf{u}(t)\| = u(t)$ to describe a (time-varying) DC EF of strength $200u(t)$ mV mm⁻¹, directed parallel to $\mathbf{u}(t)$. In particular, the specific EF used in the electrotactic experiment, with magnitude 200 mV mm⁻¹ in the positive x direction (left to right) over 60 min to 180 min, and reversed over 180 min to 300 min, is represented using the constant canonical unit vector, $\mathbf{i}$, by the vector-valued function

$$\mathbf{u}_{\text{EF}}(t) = \begin{cases} 0 & t < 60, \\ \mathbf{i} & 60 \leq t < 180, \\ -\mathbf{i} & 180 \leq t. \end{cases}$$

Note that the function $\mathbf{u}_{\text{EF}}(t)$ represents the specific EF corresponding to the electrotactic experiment, while arbitrary EF inputs are modelled using the notation $\mathbf{u}(t)$. The autonomous model in Eq. (2) can be extended to include the four hypothesised effects of the EF. The *velocity bias* effect is accounted for by modelling the velocity using two components,

$$\mathbf{v}(t) = \mathbf{v}_{\text{cell}}(t) + \mathbf{v}_{\text{EF}}(t), \quad (3a)$$

where the EF induces a deterministic velocity component in the direction of the field,

$$\mathbf{v}_{\text{EF}}(t) = \gamma_1 v \mathbf{u}(t). \quad (3b)$$

The two hypothesised electrotactic effects of *speed increase* and *speed alignment* are both modelled through adapting the velocity component induced by the cell polarity, originally defined in Eq. (2a), into

$$\mathbf{v}_{\text{cell}}(t) = (1 + \gamma_2 u(t) + \gamma_3 \mathbf{u}(t) \cdot \hat{\mathbf{p}}(t)) v \mathbf{p}(t), \quad (3c)$$

where $\hat{\mathbf{p}}$ is the unit vector in the direction of the polarity, $\mathbf{p}$. Finally, the hypothetical *polarity bias* effect is modelled in the stochastic evolution of the polarity variable $\mathbf{p}$. We add a drift term proportional to the EF to the Langevin diffusion equation, such that

$$d\mathbf{p}(t) = -D [\nabla W(\mathbf{p}(t)) - \gamma_4 \mathbf{u}(t)] dt + \sqrt{2D} d\mathbf{B}, \quad (3d)$$

where $W(\mathbf{p})$ is the same potential function as used in Eq. (2b). As for the autonomous model, the initial value for the polarity, denoted $\mathbf{p}_0 = \mathbf{p}(0)$, is also required.

Note that substituting $\mathbf{u}(t) \equiv 0$ or setting $\gamma_i = 0$ for all $i = 1, 2, 3, 4$ into Eq. (3) recovers the dynamics of the autonomous model, in Eq. (2). We will term the extended model in Eq. (3) the *electrotactic* model. It is parametrised by the three parameters

v , ΔW and D , with the same meaning and dimensions as in the autonomous model, and also by γ_i for $i = 1, 2, 3, 4$, which, since $\mathbf{u}$ is non-dimensional, are all non-dimensional.

The models in Eq. (2) and Eq. (3) result in a stochastic path for the velocity, $\mathbf{v}(t)$, with a parametrically determined stationary distribution. Following Eq. (1), each path can be integrated to produce a stochastic trajectory of the cell position over time. The stationary distributions of $\mathbf{v}$ under the autonomous and electrotactic models are depicted in Figure 1. The effect of each of the parameters γ_i , and hence each of the hypothesised electrotactic effects, can be identified by comparing the position, scale, and asymmetries of the two stationary distributions.

Summarising simulations

For any given set of parameter values, $\theta = (v, \Delta W, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$, together with initial polarity, $\mathbf{p}_0$, and non-zero EF input, $\mathbf{u}(t)$, the stochastic model in Eq. (3) can be simulated. Note that, if the EF is zero, we simulate the autonomous model in Eq. (2). Each simulation produces a random trajectory, denoted $\omega = (\mathbf{p}(t), \mathbf{x}(t))_{t \geq 0}$. We will use *summary statistics* to analyse the model outputs by mapping each simulated trajectory, ω , to a number (or small set of numbers) that summarise the trajectory. More details of the summary statistics can be found in the *Supplementary Material*.

We define a set of summary statistics based on simulated cell positions at five-minute timepoints $t_j = 5j$ over any given time interval, $t_n < t_{n+m}$. We consider: (a) the net horizontal cell displacement over the entire interval, $(\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)) \cdot \mathbf{i}$, denoted by $Y_1(\omega)$; (b) the net absolute cell displacement over the entire interval, $\|\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)\|$, denoted $Y_2(\omega)$; (c) the path length, measured as the sum of displacements, $\sum_{r=1}^m \|\mathbf{x}(t_{n+r}) - \mathbf{x}(t_{n+r-1})\|$, between the five-minute sample points, denoted $Y_3(\omega)$, (d) the standard deviation of the displacements, $\|\mathbf{x}(t_{n+r}) - \mathbf{x}(t_{n+r-1})\|$, over $r = 1, \dots, m$, and denoted $Y_4(\omega)$. Note that the four summary statistics Y_1, Y_2, Y_3 and Y_4 can also be applied to the observed data, $\mathbf{x}_{\text{NoEF},i}$ and $\mathbf{x}_{\text{EF},i}$, in addition to any simulated trajectory, ω .

In the models in Eq. (2) and Eq. (3), the polarity, $\mathbf{p}(t)$, evolves randomly from initial value $\mathbf{p}_0$. We define a further three summary statistics based on the simulated polarity, using a threshold polarity magnitude, $\bar{p}$. First, we define the *time to polarise*, T_1 , as the average time at which a simulated cell polarity, from initial polarity $\mathbf{p}_0 = \mathbf{0}$, first has polarity $\|\mathbf{p}(t)\| \geq \bar{p}$. Conversely, we define the *time to depolarise*, T_0 , as the average time at which a simulated cell polarity, from initial polarity $\mathbf{p}_0 = \mathbf{i}$, first has polarity $\|\mathbf{p}(t)\| \leq \bar{p}$. Finally, the value $\mathbb{P}(\text{polarised})$ is defined as the probability that a simulated cell polarity at the end of an assay satisfies $\|\mathbf{p}(300)\| \geq \bar{p}$. Note that these summary statistics *cannot* be applied to the observed data, since the polarity is not observed, and can only be used to summarise simulated trajectories.

Model calibration and selection

Given the experimental training data sets, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, the autonomous and electrotactic models can be *calibrated* by identifying the values of the parameters,

$$\theta = (v, \Delta W, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4),$$

that are consistent with the observed behaviour. We employ a Bayesian approach to parameter inference, whereby prior beliefs about θ , encoded in a *prior distribution*, $\pi(\theta)$, are updated in the context of the experimental data according to Bayes's rule,

$$\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) = \frac{\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta) \pi(\theta)}{p(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})},$$

where $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)$ is the *likelihood* of observing the data under the models in Eq. (2) and Eq. (3) with the parameter value θ . The resulting *posterior distribution*, $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, represents the remaining uncertainty in the parameter values, given the training data (19).

The simulation and inference algorithms used in this work have been developed in Julia 1.5.1 (20). The code is publicly available at github.com/tpprescott/electro.

Bayesian synthetic likelihoods and sequential Monte Carlo

In practice, the likelihood cannot be calculated directly, and so we require a likelihood-free approach. We replace the true likelihood with a *synthetic likelihood*, where for each value of θ the likelihood is approximated by the likelihood of summarised data under an empirical Gaussian distribution. The empirical distribution is fit to the sample mean and covariance of a set of $n = 500$ summarised simulations (21–23). The summary statistics we use are Y_1, Y_2, Y_3 and Y_4 , as described in *Summarising simulations*, above. We summarise the interval 0 min to 300 min for the trajectories from the autonomous experiment, and the separate intervals 0 min to 60 min and 60 min to 180 min for the trajectories from the electrotactic experiment. To mitigate the computational burden of the large number of model simulations required for parameter inference, we combine a sequential Monte Carlo (SMC) algorithm with synthetic likelihoods (21, 24, 25). This approach is a popular strategy for efficiently sampling from a target distribution, and also allows the exploitation of parallelisation to speed inference (21, 25–27). We provide full details of the SMC inference approach using summary statistics and synthetic likelihoods in the *Supplementary Material*.

Prior specification and model selection

The space of possible parameter values is defined as the product of intervals,

$$\Theta = (0, 5]^2 \times (0, 0.5] \times [0, 2]^4,$$

where the interval bounds were chosen based on a preliminary qualitative, visual analysis of the simulation outputs in comparison to observed data. In order to identify which of the sixteen possible combinations of the four hypothesised electrotactic effects are best supported by the experimental data, we will define sixteen possible priors on Θ . For each of the sixteen subsets, $X \subseteq \{1, 2, 3, 4\}$, we define a uniform prior distribution $\pi_X(\theta)$ on Θ that takes a constant, positive value for parameter vectors θ if and only if $\gamma_i > 0$ for all $i \in X$, and $\gamma_i = 0$ otherwise. Thus, by performing Bayesian inference using the prior distribution, π_X , we constrain the electrotactic model in Eq. (3) to model only electrotactic effects included in the subset $X \subseteq \{1, 2, 3, 4\}$.

We define an optimisation problem that aims to prevent over-fitting, by balancing the closeness of the model fit to data while prioritising smaller parameter dimensions. The optimal subset, X , of electrotactic effects is defined as the maximiser of the objective function,

$$J_\mu(X) = \log p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) - \mu(3 + |X|), \quad (4)$$

where the regularisation parameter $\mu > 0$ controls the cost of over-fitting by penalising the total number of non-zero parameters. This number is three, corresponding to v , ΔW , and D , plus $|X|$, corresponding to the positive γ_i for $i \in X$. We use $\mu = 0$ and $\mu = 2$ in our analysis, though we note that the choice of μ is somewhat arbitrary. Choosing $\mu = 2$ imposes a penalty on the parameter dimension analogous to that used in the Akaike information criterion (19). One interpretation of the value of μ is that it effectively imposes a ‘prior’ on the subsets, $X \subseteq \{1, 2, 3, 4\}$, with probability mass proportional to $\exp(-\mu|X|)$.

The first term in $J_\mu(X)$ measures the closeness of fit between the data and the model, when constrained to only include the electrotactic effects in X . This fit is defined for each $X \subseteq \{1, 2, 3, 4\}$ by the value of the partition function,

$$p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) = \int \mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta) \pi_X(\theta) d\theta.$$

As the likelihoods $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)$ cannot be calculated directly, the partition functions $p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$ are estimated for each X by Monte Carlo sampling, where again the simulation-based synthetic likelihood is used in place of the true likelihood. More details of the specific sequential Monte Carlo sampling methodology used for this estimate are given in the *Supplementary Material*.

RESULTS

We initially calibrate the autonomous model, based on the portion of the training data set $\mathbf{x}_{\text{NoEF}}$ from the autonomous experiment alone, in order to confirm the principle of the modelling framework and its ability to replicate observed behaviours, and to check that the parameters are identifiable from the data. Then, we calibrate the full electrotactic model using the full training data set, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, in two stages. We first assess which subset of the four hypothesised electrotactic effects are best supported by the data. After choosing the optimal combination of electrotactic effects, we then calibrate the parameters of the selected electrotactic model.

Parameters of the autonomous model are identifiable

We begin by confirming that the chosen modelling and inference approaches appropriately capture the autonomous experimental behaviour, $\mathbf{x}_{\text{NoEF}}$, where no external EF is applied. The cell trajectories in this portion of the training data are depicted in Figure 2(a). This scenario is modelled by the autonomous model in Eq. (2), which depends on three parameters, $\theta_{\text{NoEF}} = (v, \Delta W, D)$. The Bayesian synthetic likelihood approach was used to generate posterior samples for: the characteristic speed of a polarised cell, $v \mu\text{m min}^{-1}$; the timescale constant, $D \text{ min}^{-1}$, which determines the characteristic timescale of the spontaneous polarisation dynamics; and the dimensionless parameter, ΔW , which determines the variability of the cell polarity around its modal value.

[Figure 2 about here; moved to end of manuscript by endfloat.]

Figure 2(b–d) depicts the marginals of the posterior distribution, $\pi(\theta_{\text{NoEF}} | \mathbf{x}_{\text{NoEF}})$, for each of the three calibrated parameters. The prior distribution used for Bayesian inference assumed that the parameters were independently uniformly distributed on the intervals $0 < v \leq 5 \mu\text{m min}^{-1}$, $0 < \Delta W \leq 5$, and $0 < D \leq 0.5 \text{ min}^{-1}$. Each plot in Figure 2(b–d) demonstrates that the posteriors are concentrated within a small interval of the prior support, implying that the parameters of the autonomous model are identifiable from the experimental data, with quantifiable uncertainty.

The sample median parameter value, calculated from the sample in Figure 2(b–d), can be used as a point estimate for the parameter values: $v = 1.15 \mu\text{m min}^{-1}$, $\Delta W = 0.018$, and $D = 0.042 \text{ min}^{-1}$. In Figure 2(e), we depict a random sample of trajectories simulated from the autonomous model, with parameter values sampled from the empirical posterior, that compare well with the observed trajectories from the autonomous experiment, $\mathbf{x}_{\text{NoEF}}$. A comparison between these plots shows that parameter inference based only on the selected four-dimensional summary statistics produces a close match (for this point estimate) between the visual characteristics of simulations and experimental observations. A more detailed analysis of the fit of the calibrated autonomous model to the training data is given in the *Supplementary Material*, including a comparison of posteriors trained on the two replicates separately, and a cross-validation of the posterior predictive distribution of the four summary statistics.

Figure 2(b–d) quantifies the uncertainty in each parameter value resulting from the Bayesian approach to parameter inference. In order to make sense of this uncertainty in terms of the model outputs, simulations can be used to interpret how the uncertainty propagates to observable behaviour. Figure 2(f–h) depicts an estimate of the uncertainty in (f) the average time a simulated cell takes to polarise, T_1 , (g) the average time a simulated cell takes to depolarise, T_0 , and (h) the proportion of simulated cells that are polarised by the end of the experiment, $\mathbb{P}(\text{polarised})$. Each of these distributions are conditioned on the posterior parameter distribution in Figure 2(b–d). This procedure allows us to map quantified uncertainty in the parameter values to uncertainty in cell behaviour. The calibrated model suggests that the expected time for a cell to spontaneously polarise (i.e. without an EF applied) ranges from 2.8 min to 10 min (5% to 95% quantiles), with median value of 5.4 min. Similarly, the expected time for a cell to depolarise is 94 min to 174 min, with median value 124 min. Finally, the probability that a simulated cell is polarised (in any direction) at the end of the experiment is 0.9 to 0.97, with median value 0.94.

314 **One of the four proposed electrotactic effects is supported by the data**

315 Given that the autonomous model can be calibrated to the data set from the autonomous experiment, we now seek to calibrate
 316 the full electrotactic model to the entire training data set taken from both experiments. However, some or all of the hypothesised
 317 electrotactic effects used to define the model in Eq. (3) may not be supported by the experimental data. Thus, we first use the
 318 training data to select which of these proposed effects can be detected in the observed cell behaviours. Recall that the parameters
 319 γ_1 , γ_2 , γ_3 and γ_4 correspond to four distinct hypothesised electrotactic effects: velocity bias, speed increase, speed alignment,
 320 and polarity bias. Positive values of the parameters γ_i , for $i = 1, 2, 3, 4$, mean that the corresponding effect is included in the
 321 model. Conversely, setting any of these parameters to zero excludes the corresponding effect(s) from the model. There are
 322 a total of $2^4 = 16$ possible combinations of the four proposed electrotactic effects that the model in Eq. (3) can implement,
 323 through combinations of positive and zero parameter values.

324 [Figure 3 about here; moved to end of manuscript by endfloat.]

325 Each of the 16 possible combinations of the four electrotactic effects corresponds to a subset $X \subseteq \{1, 2, 3, 4\}$. We evaluate
 326 each combination of electrotactic effects, given by X , with respect to the objective function, $J_\mu(X)$, given in Eq. (4). This
 327 objective quantifies the trade-off between the model fit and the number of non-zero parameters in order to select a suitably
 328 accurate model while avoiding over-parametrisation. Figure 3 ranks each of the 16 possible combinations of electrotactic effects,
 329 $X \subseteq \{1, 2, 3, 4\}$, using two different objective functions. The top plot considers $\mu = 0$, such that the maximiser of J_0 is the
 330 combination that gives the best fit to data, with no consideration given to the dimension of parameter space. The bottom plot
 331 uses $\mu = 2$, which imposes a marginal cost on increasing the dimension of parameter space. Both objective functions are
 332 maximised by the singleton subset $X = \{4\}$, by a margin of over 10 from $X = \{1, 4\}$ in second-place. This margin implies a
 333 Bayes factor (19) greater than 10 in favour of a prior such that $\gamma_4 \in (0, 2]$, with $\gamma_1 = \gamma_2 = \gamma_3 = 0$, thus providing strong support
 334 for including only the polarity bias effect of the EF in our model, and neglecting all of the other hypothesised effects. Indeed,
 335 we can also conversely conclude from Figure 3 that any prior that sets $\gamma_4 = 0$ would induce a poor fit to the observed data.

336 **The electrotactic effects of the EF on motility can be quantified**

337 Recall that cell motility is modelled as the sum of an active force component, deriving from cell polarisation, and a component
 338 comprised of other external forces acting on the cell. In the selected model found in the preceding section, we found $\gamma_1 = 0$,
 339 meaning that the training data provides no evidence that the EF imparts an external force. Finding that $\gamma_2 = \gamma_3 = 0$ further
 340 implies that polarised cells do not travel any faster in the presence of a field, neither uniformly nor only if polarised in alignment
 341 with the field. Instead, the EF produces the observed bias in cell motility solely due to causing cells to preferentially polarise in
 342 the direction of the EF. In this section, we calibrate the electrotactic model by using the entire training data set, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$,
 343 to infer the posterior distribution of $\gamma_4 > 0$, while also refining the posterior distributions of v , ΔW , and D .

Bayesian synthetic likelihoods were used to calibrate the electrotactic model by inferring the posterior distribution,

$\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, for

$$\theta = (v, \Delta W, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4).$$

The chosen prior distribution, $\pi_{\{4\}}(\theta)$, is the product of independent uniform distributions on the intervals $0 < v \leq 5 \mu\text{m min}^{-1}$, $0 < \Delta W \leq 5$, and $0 < D \leq 0.5 \text{ min}^{-1}$, multiplied by an independent and uniformly distributed prior for the polarity bias parameter on the interval $0 < \gamma_4 \leq 2$. The remaining parameters in $\pi_X(\theta)$ are fixed at $\gamma_1 = \gamma_2 = \gamma_3 = 0$.

[Figure 4 about here; moved to end of manuscript by endfloat.]

Figure 4(a–d) shows the empirical marginals from the posterior sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, constructed using Bayesian synthetic likelihoods and SMC sampling. The marginals shown correspond to the four non-zero parameters of the model, v , ΔW , D , and γ_4 . The posterior marginal distributions of the previously inferred parameters, v , ΔW and D , closely match those in Figure 2(b–d), as depicted in Figure S7 of the *Supplementary Material*. Similarly to Figure 2, the posterior distribution is concentrated in a small region of the prior domain, providing evidence that each of the parameters is identifiable using the chosen summary statistics.

The median of the SMC sample depicted in Figure 4(a–d) can be used as a point estimate for the parameter values: $v = 1.11 \mu\text{m min}^{-1}$, $\Delta W = 0.020$, $D = 0.069 \text{ min}^{-1}$, $\gamma_4 = 0.60$. Figure 4(e–h) compares the training data against 50 simulations from both models, Eq. (2) and Eq. (3), using parameter values randomly sampled from the empirical posterior depicted in Figure 4(a–d). The observed bias in motility towards the direction of the EF is reflected in the stochastically simulated outputs. This provides visual confirmation that parameters inferred by Bayesian synthetic likelihood, based on the chosen summary statistics, produce simulated outputs that share observable characteristics with the experimental data. In the *Supplementary Material* we provide a more detailed validation of the fit of the calibrated model to the training data, including a cross-validation of the posterior predictive distributions of the summary statistics against the observed summary statistics, in Figure S8.

Validation against test data

Recall that the portion of the data collected from the electrotactic experiment corresponding to the time interval 180 min to 300 min was held back from the training set used to calibrate the model. The predictions of the calibrated electrotactic model can be validated against this test data set, as a prediction of the cell responses to a reversed EF input.

[Figure 5 about here; moved to end of manuscript by endfloat.]

In Figure 5, we compare the predictions to the test data in two ways. Figure 5 depicts the marginals of the empirical posterior predictive distribution for all four summary statistics over the interval 180 min to 300 min, overlaid against the corresponding summarised test data. These distributions show a good level of agreement, demonstrating that these characteristics of the

observed motility data can be predicted by the model. In Figure 5(e–f) we also compare the observed trajectories over 180 min to 300 min (translated to begin at the origin) against a set of simulated trajectories, using parameter values sampled randomly from the posterior in Figure 4(a–d). A visual comparison shows that the diversity of observed trajectory characteristics is well-captured by the stochastic simulations and parameter uncertainty. These comparisons provide evidence helping to validate the calibrated model, by demonstrating its ability to accurately predict the cellular response to dynamic EF inputs against unseen data.

DISCUSSION & CONCLUSION

The primary goal of this work has been to use mathematical modelling to quantitatively identify the contributions of multiple hypothesised means by which EFs induce electrotaxis in single cells. We have presented an empirical, parametrised, agent-based model of electrotactic cell motility, and shown that it can be calibrated to single-cell trajectory data using likelihood-free Bayesian inference. To our knowledge, although many models of single-cell and collective motility under environmental cues have been developed (13), there have been few mathematical models of electrotaxis (28, 29), and this work is the first use of detailed mathematical modelling at a single-cell level to quantify motility under electrotaxis. Moreover, the inferred parameter values of the calibrated model provide quantitative, mechanistic insights into experimentally-observed electrotaxis.

Specifically, by calibrating the model to experimental observations of electrotaxis in human corneal epithelial cells, we have concluded that the observed bias in motility is the result of *polarity bias*, where cells preferentially evolve the direction of their polarisation to align with the direction of the EF. The data does not support the hypothesis that an EF contributes an external force on the cell; nor that polarised cells travel at different speeds in the presence of a field. By carefully calibrating the parametrised mathematical model to experimental data, we have quantified the effect of the polarity bias on the electrotactic phenotype of this cell line.

A key strength of the model presented in Eq. (3) is its flexibility. The parametric design means that the Bayesian calibration methodology used in this work can be recapitulated to calibrate the same model to electrotaxis assays using other cell types or with different experimental conditions. Thus, observed differences in spontaneous and electrotactic motility between different cells and experimental conditions (3, 5) can be modelled and predicted within a common parametric framework. It is also important to acknowledge that we have chosen from only four hypothetical observable effects of electrotaxis. Other electrotactic effects may be reasonably included in the modelling process: for example, the EF may induce changes to the rate of polarisation and depolarisation (3). The electrotaxis model can straightforwardly be extended and recalibrated to account for any alternative hypothetical effects.

We have also considered EFs at a single reference strength, requiring a single parameter to quantify each hypothesised electrotactic effect. However, the characteristics of electrotaxis have been observed to vary nonlinearly with EF strength (5). The model is sufficiently flexible to account for this phenomenon through the replacement of the parameters γ_i with functions

$\Gamma_i(u)$ that vary with the EF strength, u mV mm⁻¹. The challenge will then be to use experimental data gathered from assays using EFs of different strengths to infer each of the functions Γ_i in place of each of the parameters γ_i .

The model we have presented predicts single-cell electrotactic behaviour. However, there is a wealth of data and analysis on electrotaxis in the context of cell populations (3, 4, 6–9, 13). The electrotaxis model in this paper is a starting point for a comprehensive agent-based model that also incorporates phenomena such as volume exclusion, adhesion, elastic collisions, contact inhibition, and so on (13, 30, 31). Furthermore, there is significant scope for linking the calibrated parameters of the single-cell model described in this work to the construction of lower-level models of the intracellular processes that give rise to electrotaxis. Multifidelity approaches (27, 32) that can link experiments and information at the intracellular, single-cell and multicellular level will be vital to identify and quantify the biasing effects of EFs on the collective motility of cell populations (12, 14, 33).

The model considered in this paper, and the Bayesian uncertainty quantification of its parameters, are important tools for enabling stochastic model predictive control designs of such policies based on output feedback and filtering (34). We have therefore provided a significant step towards the real-time model predictive control of populations of electrotactic cells.

AUTHOR CONTRIBUTIONS

The experiments and image analysis were performed by KZ and MZ. The modelling, simulation and inference were carried out by TPP and REB. All authors contributed to writing the manuscript.

ACKNOWLEDGMENTS

The experiments in the Zhao lab were supported by the AFOSR-MURI grant FA9550-16-1-0052 and NIH grant 1R01EY019101. TPP and REB are supported by BBSRC through grant BB/R000816/1, and REB is supported by a Royal Society Wolfson Research Merit Award.

REFERENCES

1. Stuelten, C. H., C. A. Parent, and D. J. Montell, 2018. Cell motility in cancer invasion and metastasis: insights from simple model organisms. *Nature Reviews Cancer* 18:296–312.
2. Kai, H., T. Yamauchi, Y. Ogawa, A. Tsubota, T. Magome, T. Miyake, K. Yamasaki, and M. Nishizawa, 2017. Accelerated wound healing on skin by electrical stimulation with a bioelectric plaster. *Advanced Healthcare Materials* 6:1700465.
3. Cortese, B., I. E. Palamà, S. D'Amone, and G. Gigli, 2014. Influence of electrotaxis on cell behaviour. *Integrative Biology* 6:817–830.

4. Cho, Y., M. Son, H. Jeong, and J. H. Shin, 2018. Electric field–induced migration and intercellular stress alignment in a collective epithelial monolayer. *Molecular Biology of the Cell* 29:2292–2302.
5. Saltukoglu, D., J. Grünewald, N. Strohmeyer, R. Bensch, M. H. Ulbrich, O. Ronneberger, and M. Simons, 2015. Spontaneous and electric field–controlled front–rear polarization of human keratinocytes. *Molecular Biology of the Cell* 26:4373–4386.
6. Zhao, M., A. Agius-Fernandez, J. V. Forrester, and C. D. McCaig, 1996. Directed migration of corneal epithelial sheets in physiological electric fields. *Investigative Ophthalmology & Visual Science* 37:2548–2558.
7. Zhao, M., B. Song, J. Pu, J. V. Forrester, and C. D. McCaig, 2003. Direct visualization of a stratified epithelium reveals that wounds heal by unified sliding of cell sheets. *The FASEB Journal* 17:397–406.
8. Cohen, D. J., W. J. Nelson, and M. M. Maharbiz, 2014. Galvanotactic control of collective cell migration in epithelial monolayers. *Nature Materials* 13:409–417.
9. Sun, Y.-H., Y. Sun, K. Zhu, B. Reid, X. Gao, B. W. Draper, M. Zhao, and A. Mogilner, 2017. Electric fields accelerate cell polarization and bypass myosin action in motility initiation. *Journal of Cellular Physiology* 233:2378–2385.
10. Hughes, R. M., and D. S. Lawrence, 2014. Optogenetic engineering: light-directed cell motility. *Angewandte Chemie International Edition* 53:10904–10907.
11. Wang, E., M. Zhao, J. V. Forrester, and C. D. McCaig, 2003. Bi-directional migration of lens epithelial cells in a physiological electrical field. *Experimental Eye Research* 76:29–37.
12. Lee, R. M., H. Yue, W.-J. Rappel, and W. Losert, 2017. Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. *Journal of The Royal Society Interface* 14:20170147.
13. Camley, B. A., and W.-J. Rappel, 2017. Physical models of collective cell motility: from cell to tissue. *Journal of Physics D: Applied Physics* 50:113002.
14. Matsiaka, O. M., C. J. Penington, R. E. Baker, and M. J. Simpson, 2018. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics. *Bulletin of Mathematical Biology* 80:738–757.
15. Nakajima, K., K. Zhu, Y.-H. Sun, B. Hegyi, Q. Zeng, C. J. Murphy, J. V. Small, Y. Chen-Izu, Y. Izumiya, J. M. Penninger, and M. Zhao, 2015. KCNJ15/Kir4.2 couples with polyamines to sense weak extracellular electric fields in galvanotaxis. *Nature Communications* 6.
16. Zhu, K., Y. Takada, K. Nakajima, Y. Sun, J. Jiang, Y. Zhang, Q. Zeng, Y. Takada, and M. Zhao, 2019. Expression of integrins to control migration direction of electrotaxis. *The FASEB Journal* 33:9131–9141.
17. Erban, R., and S. J. Chapman, 2019. Stochastic Modelling of Reaction–Diffusion Processes. Cambridge University Press.

- 458 18. Pavliotis, G., 2014. Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations.
459 Springer, New York, NY.
- 460 19. Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. Bayesian Data Analysis. Chapman
461 and Hall/CRC, 3rd edition.
- 462 20. Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah, 2017. Julia: A fresh approach to numerical computing. *SIAM*
463 *Review* 59:65–98.
- 464 21. Price, L. F., C. C. Drovandi, A. Lee, and D. J. Nott, 2017. Bayesian synthetic likelihood. *Journal of Computational and*
465 *Graphical Statistics* 27:1–11.
- 466 22. Priddle, J. W., S. A. Sisson, D. T. Frazier, and C. Drovandi, 2019. Efficient Bayesian synthetic likelihood with whitening
467 transformations. arXiv:1909.04857.
- 468 23. Picchini, U., U. Simola, and J. Corander, 2020. Adaptive MCMC for synthetic likelihoods and correlated synthetic
469 likelihoods. arXiv:2004.04558.
- 470 24. Hainy, M., C. C. Drovandi, and J. M. McGree, 2016. Likelihood-free extensions for Bayesian sequentially designed
471 experiments. In *mODa 11 - Advances in Model-Oriented Design and Analysis*, Springer International Publishing, 153–161.
- 472 25. Del Moral, P., A. Doucet, and A. Jasra, 2012. An adaptive sequential Monte Carlo method for approximate Bayesian
473 computation. *Statistics and Computing* 22:1009–1020.
- 474 26. Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, 2009. Approximate Bayesian computation scheme for
475 parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* 6:187–202.
- 476 27. Prescott, T. P., and R. E. Baker, 2020. Multifidelity approximate Bayesian computation with sequential Monte Carlo
477 parameter sampling. arXiv:2001.06256v1.
- 478 28. Akiyama, M., T. Sushida, S. Ishida, and H. Haga, 2017. Mathematical model of collective cell migrations based on cell
479 polarity. *Development, Growth & Differentiation* 59:471–490.
- 480 29. Vanegas-Acosta, J., D. Garzón-Alvarado, and A. Zwamborn, 2012. Mathematical model of electrotaxis in osteoblastic
481 cells. *Bioelectrochemistry* 88:134–143.
- 482 30. Binny, R. N., A. James, and M. J. Plank, 2016. Collective cell behaviour with neighbour-dependent proliferation, death and
483 directional bias. *Bulletin of Mathematical Biology* 78:2277–2301.
- 484 31. Warne, D. J., R. E. Baker, and M. J. Simpson, 2017. Optimal quantification of contact inhibition in cell populations.
485 *Biophysical Journal* 113:1920–1924.

- 486 32. Prescott, T. P., and R. E. Baker, 2020. Multifidelity approximate Bayesian computation. *SIAM/ASA Journal on Uncertainty*
487 *Quantification* 8:114–138.
- 488 33. Zhang, Y., G. Xu, R. M. Lee, Z. Zhu, J. Wu, S. Liao, G. Zhang, Y. Sun, A. Mogilner, W. Losert, T. Pan, F. Lin, Z. Xu, and
489 M. Zhao, 2017. Collective cell migration has distinct directionality and speed dynamics. *Cellular and Molecular Life*
490 *Sciences* 74:3841–3850.
- 491 34. Mesbah, A., 2016. Stochastic model predictive control: an overview and perspectives for future research. *IEEE Control*
492 *Systems* 36:30–44.

LIST OF FIGURES

493	1	Comparison of the stationary distributions for the random velocity, $\mathbf{v}$, under the autonomous and electrotactic models.	21
494			
495	2	Parameter inference and simulation of the autonomous model.	22
496			
497	3	Values for objective functions across $X \subseteq \{1, 2, 3, 4\}$	23
498	4	Model calibration results using all training data.	24
499	5	Comparing predictions of the calibrated electrotactic model against test data.	25

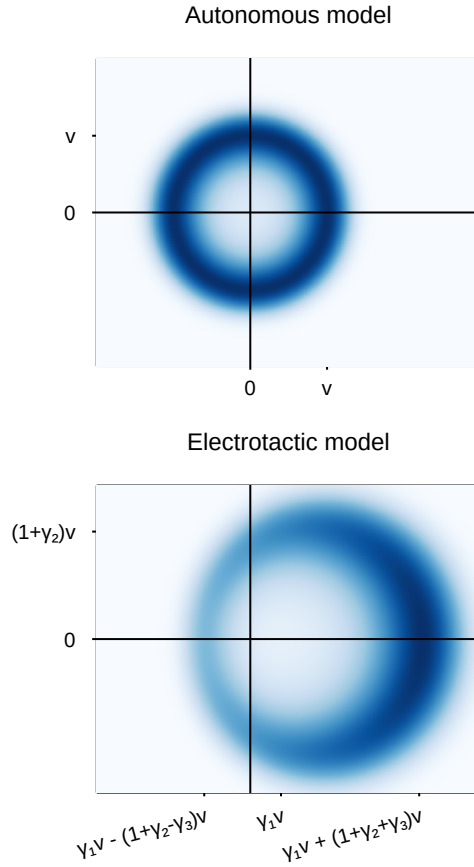


Figure 1: Comparison of the stationary distributions for the random velocity, $\mathbf{v}$, under the autonomous and electrotactic models, where darker regions correspond to greater probability. The bottom plot shows the hypothesised electrotactic effects of an EF, applied in the positive x direction, parametrised by $\gamma_1, \dots, \gamma_4$. The effects of γ_1, γ_2 and γ_3 are visible in the shape of the distribution. *Polarity bias* (γ_4) produces asymmetry in the distribution density, shown as a darker region to the right of the figure.

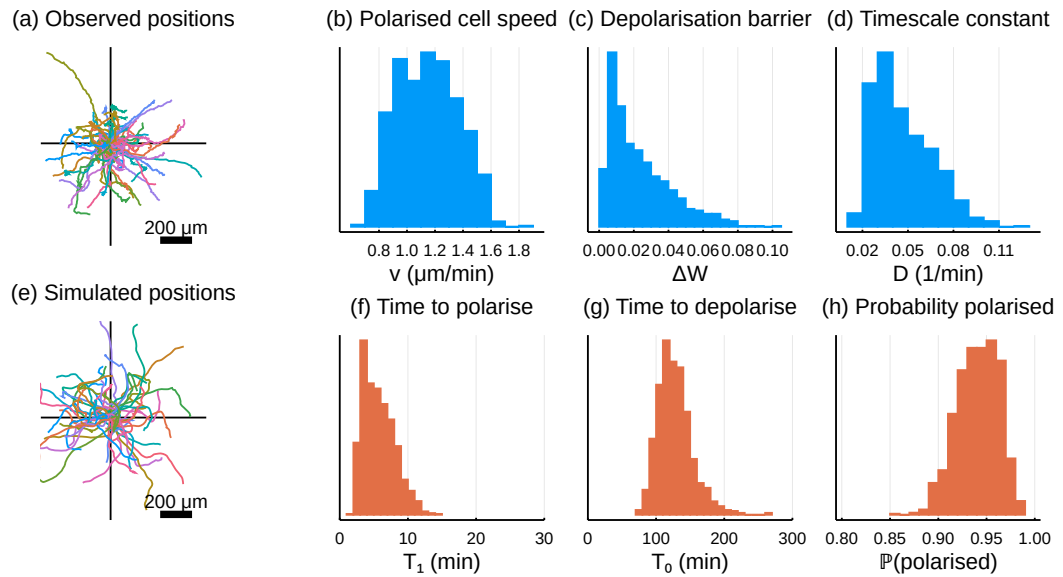


Figure 2: Parameter inference and simulation of the autonomous model. (a) The subset of the training data used to calibrate the autonomous model, corresponding to all observed trajectories under the autonomous experiment. (b–d) All one-dimensional projections of the posterior sample from $\pi(\theta_{\text{NoEF}} \mid \mathbf{x}_{\text{NoEF}})$. The covariance structure of the posterior is given in *Supplementary Material*, Figure S2. (e) Simulations of the calibrated model, using parameters randomly selected from the posterior depicted in (b–d). (f–h) Posterior predictive samples for T_1 (time to polarisation), T_0 (time to depolarisation), and $\mathbb{P}(\text{polarised})$ (probability of a cell being polarised at the final time) for simulations from the autonomous model with parameters taken from the posterior, $\pi(\theta_{\text{NoEF}} \mid \mathbf{x}_{\text{NoEF}})$, depicted in (b–d).

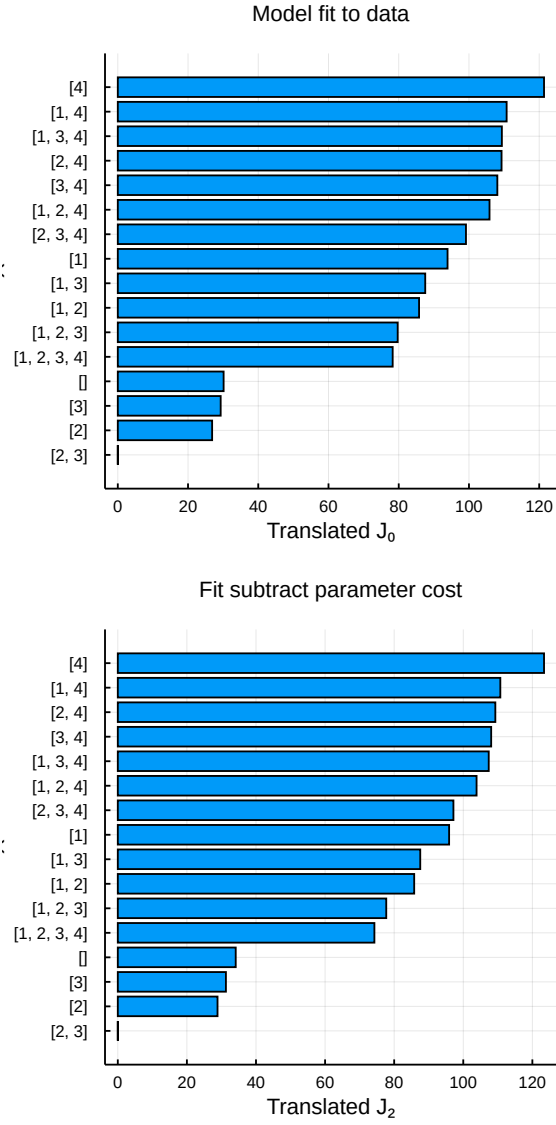


Figure 3: Objective functions $J_0(X)$ and $J_2(X)$ from Eq. (4), for combinations of electrostatic effects indexed by $X \subseteq \{1, 2, 3, 4\}$. Greater values are preferred. Each objective function is translated to have zero minimum value.

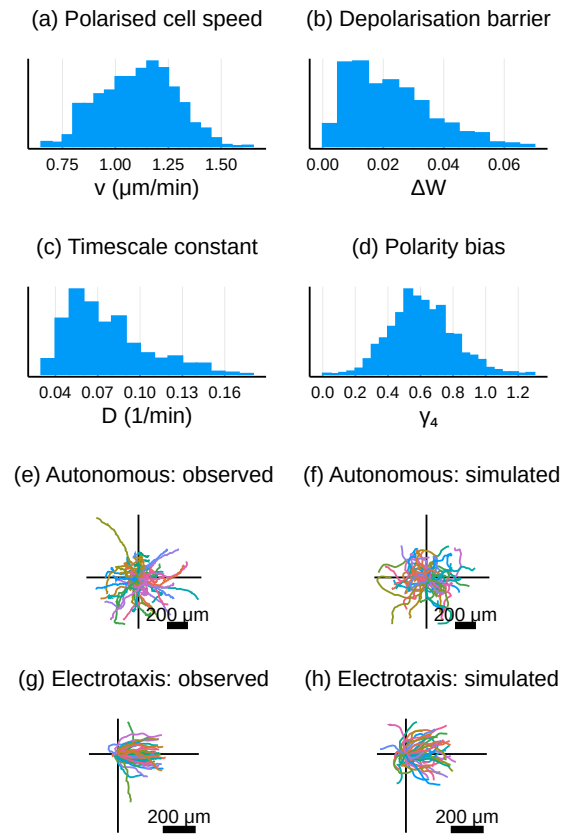


Figure 4: Empirical posterior samples inferred from training data: all trajectories from the autonomous experiment, and also from the electrotactic experiment for 0 min to 180 min. (a–d) One-dimensional projections of the empirical posterior distribution for all non-zero parameter values, based on the selected prior, $\pi_{[4]}$. In the *Supplementary Material*, the two-dimensional projections of this posterior are depicted in Figure S6, and a comparison of the posterior distributions of v , ΔW and D in Figure 2(b–d) and here is depicted in Figure S7. (e–h) Observed and simulated trajectories for (e, f) the autonomous experiment, and (g, h) the electrotactic experiment over 0 min to 180 min. Simulations in (f, h) were produced for randomly sampled parameter values from the posterior in (a–d).

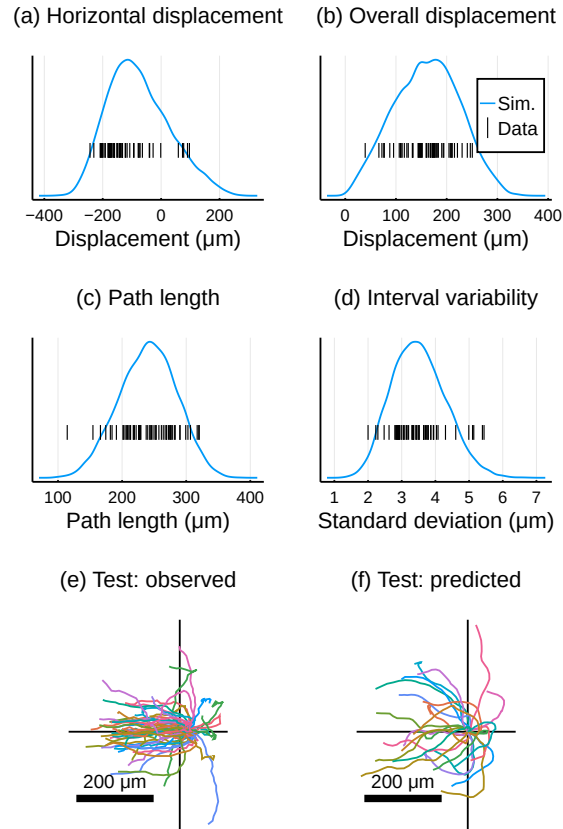


Figure 5: Comparing predictions of the calibrated electrotactic model against test data: observed trajectories from the electrotactic experiment for 180 min to 300 min. (a–d) One-dimensional projections of the summary statistics. Curves are an empirical distribution of simulated summary statistics for each parameter value from the posterior sample depicted in Figure 4(a–d). We overlay a barcode plot of each of the summary statistics of the observed test data. (e–f) Observed and simulated trajectories over 180 min to 300 min. Simulations in (f) were produced for randomly sampled parameter values from the posterior in Figure 4(a–d).