

Quantifying the impact of electric fields on single-cell motility

Thomas P. Prescott,^{1,3,*} Kan Zhu,² Min Zhao,² and Ruth E. Baker¹

¹Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford, United Kingdom; ²Department of Ophthalmology and Vision Science, Department of Dermatology, Institute for Regenerative Cures, University of California, Sacramento, California; and ³Alan Turing Institute, London, United Kingdom

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

SIGNIFICANCE Electrotaxis is attracting much interest and development as a technique to control cell migration because of 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 parameterized stochastic model that accurately replicates experimental single-cell data and enables the prediction of input-output behavior while quantifying uncertainty and stochasticity. The model allows us to elucidate and quantify how electric fields perturb the motile behavior of the cell. This model and the associated simulation-based calibration methodology will be central to future developments in the control of electrotaxis.

INTRODUCTION

Cell migration underpins key physiological processes central to developmental biology, as well as wound healing and tissue regeneration, and it plays a crucial role in invasive, metastatic cancers. There are ongoing efforts to intervene in and influence these phenomena to, for example, inhibit metastasis (1) or accelerate wound healing (2). However, the cellular processes driving collective migration are complex and multifaceted, deriving from diverse physical mechanisms and various external stimuli (3), making it challenging for

researchers to accurately and robustly direct cell motility. Because of the ease with which electric fields can be controlled and applied to cells, research into the control of cell motility has recently focused on exploiting electrotaxis (also known as galvanotaxis) (3–5). However, the precise effects of electric fields on intracellular processes and thus on cell motility are not fully understood, making quantitative predictions and control policy design impractical.

Electrotactic cells have been observed to change their motile behavior in response to the presence of a direct current (DC) electric field (EF) (3–7). Researchers seeking to control cell motility exploit this phenomenon by applying external electrical cues to cell populations (2,4–9). The key advantages of using electrical cues to guide cell migration include the ability to exploit endogenous, evolved

Submitted January 22, 2021, and accepted for publication June 28, 2021.

*Correspondence: tprescott@turing.ac.uk

Editor: Pablo Iglesias.

<https://doi.org/10.1016/j.bpj.2021.06.034>

© 2021 Biophysical Society.

biological functionality to respond to precisely controllable DC EFs. This compares favorably to using chemoattractants to guide motility, as chemical signals experienced by the cell cannot be so precisely or flexibly controlled, especially dynamically, and chemoattractants are usually highly cell specific. In contrast, light-directed motility allows for precise actuation signals. However, it requires sophisticated optogenetic manipulations of the cell population under control (10). As such, EFs provide a relatively precise and simply implemented actuation signal to achieve specified motile behaviors.

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

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

The agent-based modeling framework used in this work follows standard modeling 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 nonfrictional forces on the cell. We provide full details on the mathematical model in

[Materials and methods](#) and in the [Supporting materials and methods](#).

In the absence of any EF, the only nonfrictional 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 composed of a single component. A preliminary analysis of single-cell motility data, described more fully in the [Supporting materials and methods](#), suggests that the cell velocity arises from a cell being polarized in a particular direction and that the direction of polarization drifts stochastically over time. A polarized cell has a positive speed parameterized 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 characterizes the rate of decay in the autocorrelation of the polarization direction over time. [Equation 2](#) in the [Materials and methods](#) provides the mathematical formulation of this model.

We hypothesize 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 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 parameterize the magnitude of each hypothesized 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 hypothesized effect is not included in the model. [Equation 3](#) in the [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 as follows:

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}_{cell} , and an EF component, \mathbf{v}_{EF} . The EF velocity component acts in the direction of the field with magnitude $\gamma_1 v$.

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

Speed alignment (γ_3): Polarized cells travel more quickly when the direction of their polarization aligns with the EF but more slowly if opposed to the EF. The modal magnitude of \mathbf{v}_{cell} for polarized cells is increased by $\gamma_3 v \cos(\theta)$, where θ is the angle between \mathbf{v}_{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 polarize in the direction of the EF. The strength of this bias is parameterized by γ_4 .

Two models can be distinguished: the autonomous model, in which no EF is applied, and the electrotactic model, in which 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. Fig. 1 characterizes 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 ve-

locity distribution uniformly in the direction of the field, γ_2 rescales the domain of the distribution, γ_3 parameterizes asymmetry in the shape of the velocity distribution, and γ_4 parameterizes asymmetry in the density of the velocity distribution.

Outline

The primary goal of this work is to use single-cell experimental data to calibrate the parameterized mathematical model of spontaneous polarization and electrotaxis. The model calibration process enables the identification of which of the four hypothesized 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 are gathered from two assays in which the trajectories of motile human corneal epithelial cells are recorded for 5 h: 1) without any EF applied for the entire experiment and 2) 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 5 h of the autonomous experiment and the first 3 h 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 hypothesized 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 2 h of the electrotactic experiment, in which the EF input has changed direction. We compare these predictions with the cell trajectories observed over the final 2 h 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 toward the future goal of model-based policy design for the electrotactic control of single-cell and population-level motility.

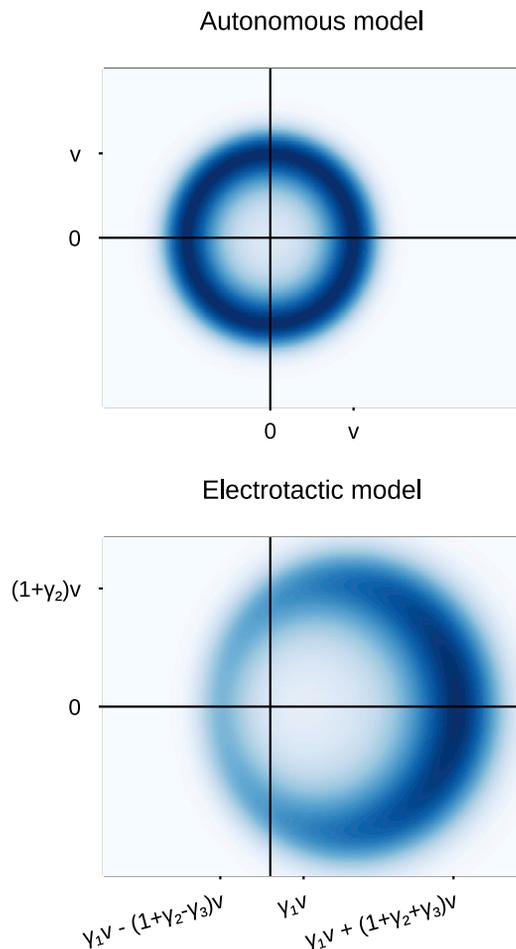


FIGURE 1 Comparison of the stationary distributions for the random velocity, \mathbf{v} , under the autonomous and electrotactic models, in which darker regions correspond to greater probability. The bottom plot shows the hypothesized electrotactic effects of an EF, applied in the positive x direction, parameterized 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. To see this figure in color, go online.

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

Prescott et al.

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 electrostatic experiment, the cells were subjected to a DC EF at a reference strength, 200 mV mm^{-1} , applied across the medium from $t = 60 \text{ min}$ to the end of the experiment. The EF was directed from left to right from 60 to 180 min, at which point the field direction was reversed from right to left for 180–300 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 electrostatic assay replicates, all over the entire time horizon. Visual confirmation from the raw experimental output confirms that cell collisions were rare because of the low density ($100 \text{ cells cm}^{-2}$) at which cells are initially seeded. We thus assume that cell-cell interactions can be neglected in this model. The data used in this work are shared online at <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 electrostatic experiments, respectively, in which 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, and $t_k = 5k \text{ min}$ 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 are used as training data, from which the model is calibrated. Thus, the training data consist of the trajectories from the autonomous experiment over the entire time horizon and the trajectories from the electrostatic experiment over 0–180 min, which are denoted by \mathbf{x}_{NoEF} and \mathbf{x}_{EF} , respectively. The test data, denoted \mathbf{x}_{test} , consist of all trajectories from the electrostatic experiment over 180–300 min, when the input EF has switched direction.

Materials

EpiLife culture medium with Ca^{2+} ($60 \mu\text{M}$), EpiLife defined growth supplement, and penicillin/streptomycin were purchased from Thermo Fisher Scientific (Waltham, MA). FNC Coating Mix was purchased from Athena Enzyme Systems (Baltimore, MD). Dow Corning high-vacuum grease was purchased from Thermo Fisher Scientific. Agar was purchased from MilliporeSigma (Burlington, MA). Silver wires with 99.999% purity were purchased from Advent Research Materials (Oxford, UK).

Cell culture

Telomerase-immortalized human corneal epithelial cells were routinely cultured in EpiLife medium supplemented with EpiLife defined growth supplement and 1% (v/v) penicillin/streptomycin. Cells were incubated at 37°C with 5% CO_2 until they reached $\sim 70\%$ confluence and were used between passages 55 and 65 for all cell migration assays.

Electrotaxis assay

Electrotaxis experiments were performed as previously described (15,16), with minor changes. Briefly, the electrotaxis chambers ($20 \times 10 \times 0.2 \text{ mm}$) were constructed in 100 mm petri dishes with glass strips and high-vacuum grease. The dimensions of the chambers were defined by the thickness and length of the glass slides, respectively. Chambers were coated with FNC Coating Mix, following the manufacturer's instructions to facilitate cell attachment. Cells were seeded at a low density ($100 \text{ cells cm}^{-2}$) and cultured overnight (12–18 h) in the chambers to allow sufficient attachment. Chambers were covered with glass coverslips and sealed with high-vacuum grease. Electric currents were applied to the chamber through agar-salt bridges connecting with silver-silver chloride electrodes in Steinberg's solution (58 mM NaCl and 0.67 mM KCl and 0.44 mM $\text{Ca}(\text{NO}_3)_2$, 1.3 mM MgSO_4 , and 4.6 mM Tris base (pH 7.4)). Fresh cell culture medium

(EpiLife) was added into reservoirs to ensure good salt bridge contact and to support cell viability during electric stimulation. An EF strength of 200 mV mm^{-1} was used unless otherwise noted. A pair of measuring electrodes was placed at the end of the electrotaxis chamber and connected to the multimeter for real-time monitoring of EF strength. The applied voltages were confirmed at the beginning of the experiment and every 30 min afterwards to ensure consistent EF application.

Time-lapse imaging and quantification of cell migration

Cell migration was monitored and recorded by phase-contrast microscopy using an inverted microscope (Carl Zeiss, Oberkochen, Germany) equipped with a motorized stage and a regular $10\times$ objective lens. Time-lapse images were acquired at 5 min intervals using Metamorph NX imaging software (Molecular Device, Sunnyvale, CA). To maintain standard cell culture conditions (37°C , 5% CO_2), a Carl Zeiss incubation system was used. Time-lapse images of cell migration were analyzed by using ImageJ software from the National Institutes of Health (<https://imagej.nih.gov/ij/>). Adherent cells in the images were manually tracked, and cells that divided, moved in and out of the field, or merged with other cells during the experiment were excluded from analysis. The position of a cell was defined by its centroid.

Model construction

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

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

and the stochastic dynamics of \mathbf{v} are modeled. The key to this modeling task is the nondimensional internal variable 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 motility, resulting in a velocity component $\mathbf{v}_{\text{cell}}(t)$.

Modeling spontaneous polarization and motility

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

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

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

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

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 walk approaches stationarity. The initial polarity, denoted $\mathbf{p}_0 = \mathbf{p}(0)$, also needs to be specified.

The potential function $W(\mathbf{p})$ in Eq. 2b is defined to capture the intended features of the autonomous model, namely that the magnitude of the cell velocity is randomly distributed around a modal value of v and that the direction of the polarity is uniformly distributed at stationarity. It can be shown (17,18) that the variability of the velocity around its modal value of v is determined by a nondimensional energy barrier, denoted ΔW , that is sufficient to define the potential function, $W(\mathbf{p})$. For further details on

the definition of W , see the [Supporting materials and methods](#). We will calibrate the autonomous model in [Eq. 2](#) by identifying the parameters v , D , and ΔW .

Modeling motility bias due to an EF

We use a vector-valued function, $\mathbf{u}(t)$, with nondimensional 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 60–180 min and reversed over 180–300 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, and arbitrary EF inputs are modeled using the notation $\mathbf{u}(t)$. The autonomous model in [Eq. 2](#) can be extended to include the four hypothesized effects of the EF. The velocity bias effect is included for by modeling 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 hypothesized electrotactic effects of speed increase and speed alignment are both modeled through adapting the velocity component induced by the cell polarity, originally defined in [Eq. 2a](#), into

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

where $\hat{\mathbf{p}}$ is the unit vector in the direction of the polarity, \mathbf{p} . Finally, the hypothetical polarity bias effect is modeled 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 parameterized 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, because \mathbf{u} is nondimensional, are all nondimensional.

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 [Fig. 1](#). The effect of each of the parameters γ_i , and hence each of the hypothesized electrotactic effects, can be identified by comparing the position, scale, and asymmetries of the two stationary distributions.

Summarizing 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 nonzero EF input $\mathbf{u}(t)$, the stochastic model in [Eq. 3](#) can be simulated. Note that if the EF is 0, 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 analyze the model outputs by mapping each simulated trajectory, ω , to a number (or small set of numbers) that summarize the trajectory. More details of the summary statistics can be found in the [Supporting materials and methods](#).

We define a set of summary statistics based on simulated cell positions at 5 min time points $t_j = 5j$ over any given time interval, $t_n < t_{n+m}$. We consider 1) the net horizontal cell displacement over the entire interval, $(\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)) \cdot \mathbf{i}$, denoted by $Y_1(\omega)$; 2) the net absolute cell displacement over the entire interval, $\|\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)\|$, denoted $Y_2(\omega)$; 3) 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 5 min sample points, denoted $Y_3(\omega)$; and 4) 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 the 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 polarize, T_1 , as the average time at which a simulated cell polarity, from initial polarity $\mathbf{p}_0 = 0$, first has polarity $\|\mathbf{p}(t)\| \geq \bar{p}$. Conversely, we define the time to depolarize, 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 $\mathbb{P}(\text{polarized})$ 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, as the polarity is not observed, and can only be used to summarize simulated trajectories.

Model calibration and selection

Given the experimental training data sets, \mathbf{x}_{NoEF} and \mathbf{x}_{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 behavior. 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,

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

where $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta)$ 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, $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, 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/tprescott/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, in which for each value of θ , the likelihood is approximated by the likelihood of summarized data under an empirical Gaussian

Prescott et al.

distribution. The empirical distribution is fit to the sample mean and covariance of a set of $n = 500$ summarized simulations (21–23). The summary statistics we use are Y_1 , Y_2 , Y_3 , and Y_4 , as described in [Summarizing simulations](#), above. We summarize the interval 0–300 min for the trajectories from the autonomous experiment and the separate intervals 0–60 and 60–180 min for the trajectories from the electrostatic 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 parallelization to speed inference (21,25–27). We provide full details of the SMC inference approach using summary statistics and synthetic likelihoods in the [Supporting materials and methods](#).

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. To identify which of the 16 possible combinations of the four hypothesized electrostatic effects are best supported by the experimental data, we will define 16 possible priors on Θ . For each of the 16 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 electrostatic model in [Eq. 3](#) to model only electrostatic effects included in the subset $X \subseteq \{1, 2, 3, 4\}$.

We define an optimization problem that aims to prevent overfitting by balancing the closeness of the model fit to data while prioritizing smaller parameter dimensions. The optimal subset, X , of electrostatic effects is defined as the maximizer 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 regularization parameter $\mu \geq 0$ controls the cost of overfitting by penalizing the total number of nonzero 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 electrostatic effects in X . This fit is defined for each $X \subseteq \{1, 2, 3, 4\}$ by the value of the partition function,

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

As the likelihoods $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta)$ 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, in which again the simulation-based synthetic likelihood is used in place of the true likelihood. More details of the specific SMC sampling methodology used for this estimate are given in the [Supporting materials and methods](#).

RESULTS

We initially calibrate the autonomous model, based on the portion of the training data set \mathbf{x}_{NoEF} from the autonomous

experiment alone, to confirm the principle of the modeling framework and its ability to replicate observed behaviors and to check that the parameters are identifiable from the data. Then, we calibrate the full electrostatic model using the full training data set, \mathbf{x}_{NoEF} and \mathbf{x}_{EF} , in two stages. We first assess which subset of the four hypothesized electrostatic effects is best supported by the data. After choosing the optimal combination of electrostatic effects, we then calibrate the parameters of the selected electrostatic model.

Parameters of the autonomous model are identifiable

We begin by confirming that the chosen modeling and inference approaches appropriately capture the autonomous experimental behavior, \mathbf{x}_{NoEF} in which no external EF is applied. The cell trajectories in this portion of the training data are depicted in [Fig. 2 a](#). This scenario is modeled 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 polarized cell, $v \mu\text{m min}^{-1}$; the timescale constant, $D \text{min}^{-1}$, which determines the characteristic timescale of the spontaneous polarization dynamics; and the dimensionless parameter, ΔW , which determines the variability of the cell polarity around its modal value.

[Fig. 2, b–d](#) depict 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 [Fig. 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 [Fig. 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 [Fig. 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}_{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 [Supporting materials and methods](#), 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.

[Fig. 2, b–d](#) quantifies the uncertainty in each parameter value resulting from the Bayesian approach to parameter

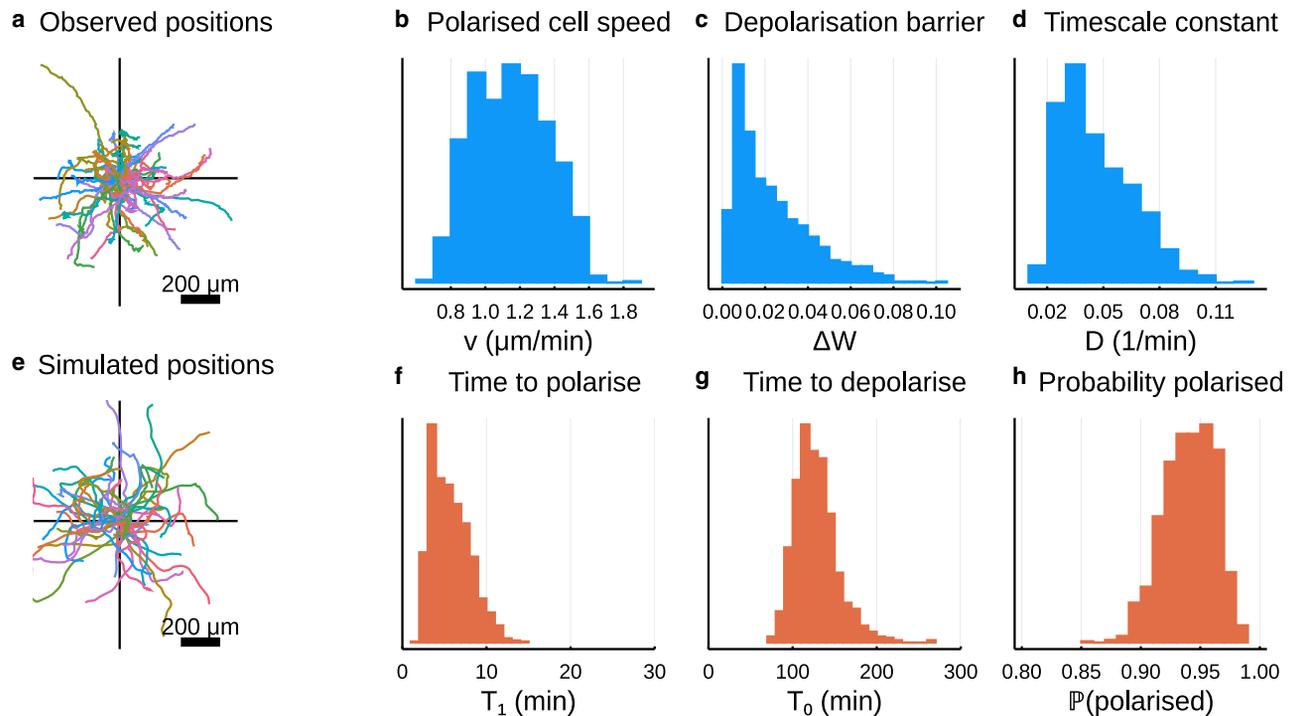


FIGURE 2 Parameter inference and simulation of the autonomous model. (a) The subset of the training data used to calibrate the autonomous model, corresponding to all observed trajectories under the autonomous experiment. (b–d) All one-dimensional projections of the posterior sample from $\pi(\theta_{\text{NoEF}} | \mathbf{x}_{\text{NoEF}})$. The covariance structure of the posterior is given in Fig. 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 polarization), T_0 (time to depolarization), and $\mathbb{P}(\text{polarized})$ (probability of a cell being polarized at the final time) for simulations from the autonomous model with parameters taken from the posterior, $\pi(\theta_{\text{NoEF}} | \mathbf{x}_{\text{NoEF}})$, depicted in (b)–(d). To see this figure in color, go online.

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

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

Given that the autonomous model can be calibrated to the data set from the autonomous experiment, we now seek to

calibrate the full electrotactic model to the entire training data set taken from both experiments. However, some or all of the hypothesized electrotactic effects used to define the model in Eq. 3 may not be supported by the experimental data. Thus, we first use the training data to select which of these proposed effects can be detected in the observed cell behaviors. Recall that the parameters γ_1 , γ_2 , γ_3 , and γ_4 correspond to four distinct hypothesized electrotactic effects: velocity bias, speed increase, speed alignment, 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 model. Conversely, setting any of these parameters to zero excludes the corresponding effect(s) from the model. There are a total of $2^4 = 16$ possible combinations of the four proposed electrotactic effects that the model in Eq. 3 can implement through combinations of positive and zero parameter values.

Each of the 16 possible combinations of the four electrotactic effects corresponds to a subset $X \subseteq \{1, 2, 3, 4\}$. We evaluate each combination of electrotactic effects, given by X , with respect to the objective function, $J_\mu(X)$, given in Eq. 4. This objective quantifies the tradeoff between the model fit and the number of nonzero parameters to select a suitably accurate model while avoiding overparameterization. Fig. 3 ranks each of the 16 possible combinations of

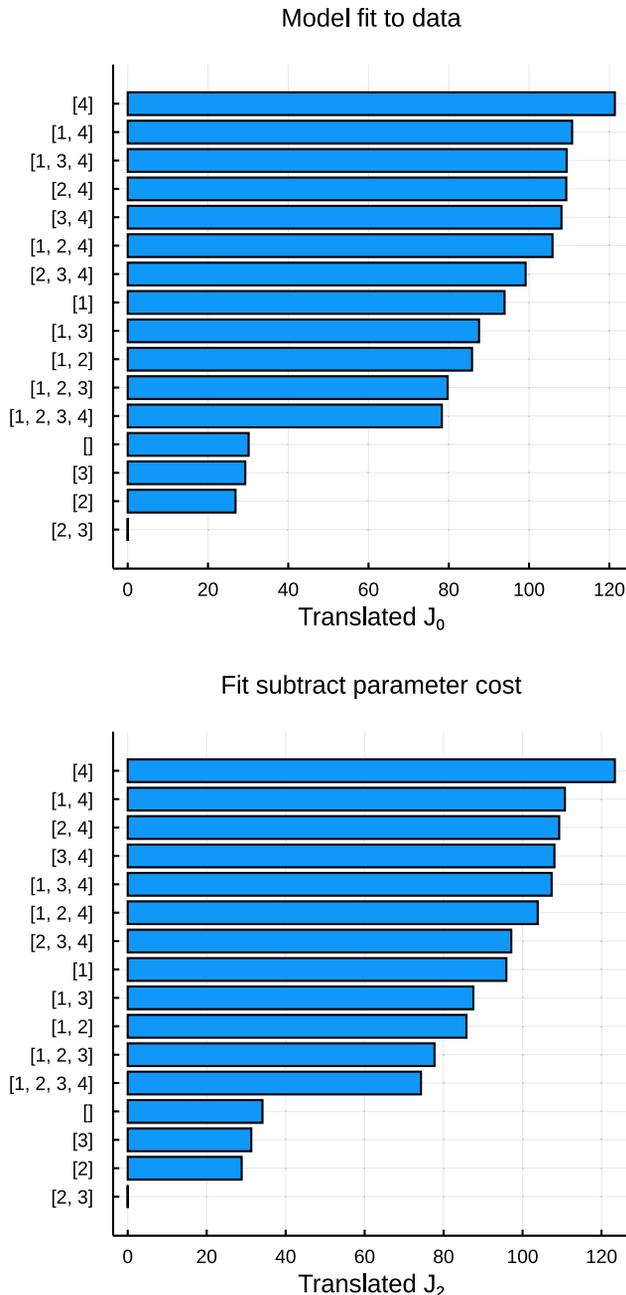


FIGURE 3 Objective functions $J_0(X)$ and $J_2(X)$ from Eq. 4, for combinations of electrostatic effects indexed by $X \subseteq \{1, 2, 3, 4\}$. Greater values are preferred. Each objective function is translated to have zero minimal value. To see this figure in color, go online.

electrotactic effects, $X \subseteq \{1, 2, 3, 4\}$, using two different objective functions. The top plot considers $\mu = 0$ such that the maximizer of J_0 is the combination that gives the best fit to data, with no consideration given to the dimension of parameter space. The bottom plot uses $\mu = 2$, which imposes a marginal cost on increasing the dimension of parameter space. Both objective functions are maximized by the singleton subset $X = \{4\}$, by a margin of over 10 from $X = \{1, 4\}$ in second place. This margin implies a Bayes

factor (19) greater than 10 in favor of a prior such that $\gamma_4 \in (0, 2]$ with $\gamma_1 = \gamma_2 = \gamma_3 = 0$, thus providing strong support for including only the polarity bias effect of the EF in our model and neglecting all of the other hypothesized effects. Indeed, we can also conversely conclude from Fig. 3 that any prior that sets $\gamma_4 = 0$ would induce a poor fit to the observed data.

The electrotactic effects of the EF on motility can be quantified

Recall that cell motility is modeled as the sum of an active force component deriving from cell polarization and a component composed of other external forces acting on the cell. In the selected model found in the preceding section, we found $\gamma_1 = 0$, meaning that the training data provide no evidence that the EF imparts an external force. Finding that $\gamma_2 = \gamma_3 = 0$ further implies that polarized cells do not travel any faster in the presence of a field, either uniformly or only if polarized in alignment with the field. Instead, the EF produces the observed bias in cell motility solely because of causing cells to preferentially polarize in the direction of the EF. In this section, we calibrate the electrotactic model by using the entire training data set, \mathbf{x}_{NoEF} and \mathbf{x}_{EF} , 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, $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, 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 $\gamma_1 = \gamma_2 = \gamma_3 = 0$.

Fig. 4, a–d show the empirical marginals from the posterior sample from $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, constructed using Bayesian synthetic likelihoods and SMC sampling. The marginals shown correspond to the four nonzero 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 Fig. 2, b–d, as depicted in Fig. S7. Similarly to Fig. 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 Fig. 4, a–d can be used as a point estimate for the parameter values: $v = 1.11 \mu\text{m min}^{-1}$, $\Delta W = 0.020$, $D = 0.069 \text{min}^{-1}$, and $\gamma_4 = 0.60$. Fig. 4, e–h compare the training data against 50 simulations from both models, Eq. 2 and Eq. 3, using parameter values

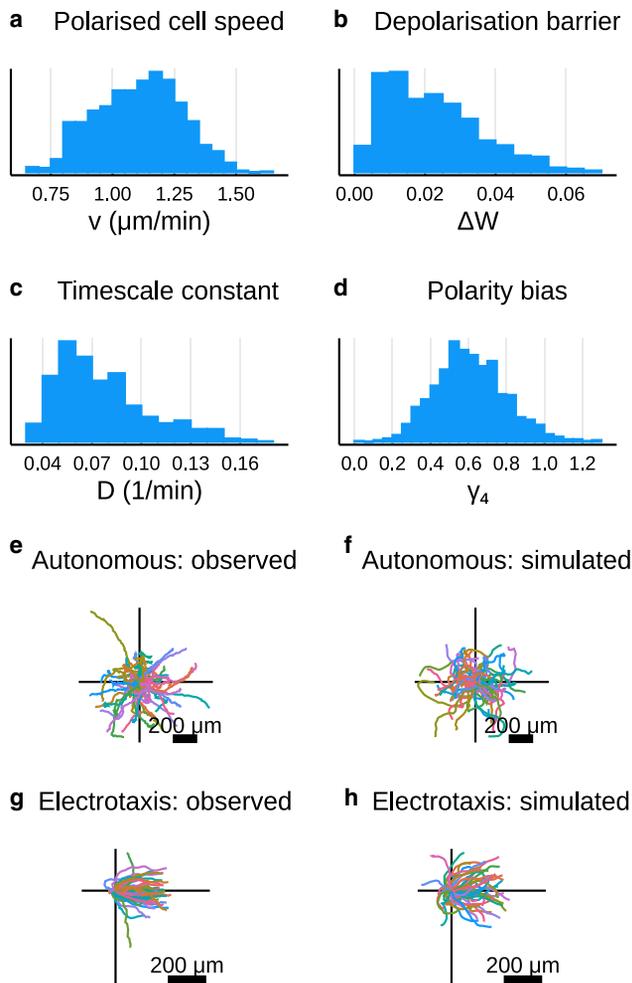


FIGURE 4 Empirical posterior samples inferred from training data: all trajectories from the autonomous experiment and also from the electrotactic experiment for 0–180 min. (a–d) One-dimensional projections of the empirical posterior distribution for all nonzero parameter values, based on the selected prior, $\pi_{[4]}$. In the [Supporting materials and methods](#), the two-dimensional projections of this posterior are depicted in [Fig. S6](#), and a comparison of the posterior distributions of v , ΔW , and D in [Fig. 2](#), b–d and here is depicted in [Fig. S7](#). (e–h) Observed and simulated trajectories for (e and f) the autonomous experiment and (g and h) the electrotactic experiment over 0–180 min. Simulations in (f) and (h) were produced for randomly sampled parameter values from the posterior in (a)–(d). To see this figure in color, go online.

randomly sampled from the empirical posterior depicted in [Fig. 4](#), a–d. The observed bias in motility toward 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 [Supporting materials and methods](#), 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 [Fig. S8](#).

Validation against test data

Recall that the portion of the data collected from the electrotactic experiment corresponding to the time interval 180–300 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.

In [Fig. 5](#), we compare the predictions to the test data in two ways. [Fig. 5](#) depicts the marginals of the empirical posterior predictive distribution for all four summary statistics over the interval 180–300 min, overlaid against the corresponding summarized 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 [Fig. 5](#), e and f, we also compare the observed trajectories over 180–300 min (translated to begin at the origin) against a set of simulated trajectories, using parameter values sampled randomly from the posterior in [Fig. 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

The primary goal of this work has been to use mathematical modeling to quantitatively identify the contributions of multiple hypothesized means by which EFs induce electrotaxis in single cells. We have presented an empirical, parameterized, 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 modeling 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, in which cells preferentially evolve the direction of their polarization to align with the direction of the EF. The data do not support the hypothesis that an EF contributes an external force on the cell or that polarized cells travel at different speeds in the presence of a field. By carefully calibrating the parameterized mathematical model to experimental data, we have quantified the effect of the polarity bias on the electrotactic phenotype of this cell line.

Prescott et al.

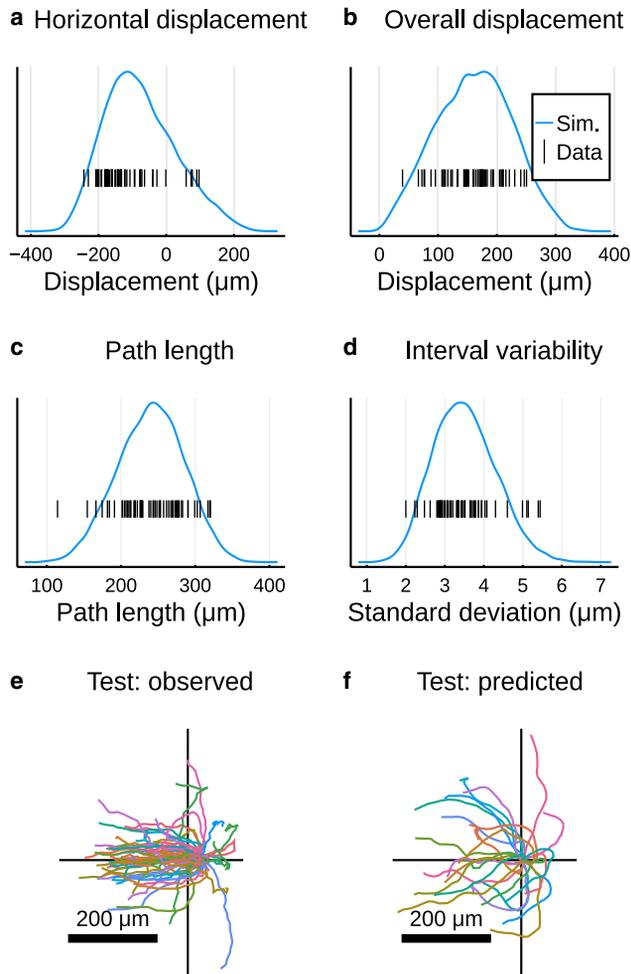


FIGURE 5 Comparing predictions of the calibrated electrostatic model against test data: observed trajectories from the electrostatic experiment for 180–300 min. (*a–d*) One-dimensional projections of the summary statistics. Curves are an empirical distribution of simulated summary statistics for each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

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 electrostatic assays using other cell types or with different experimental conditions. Thus, observed differences in spontaneous and electrostatic motility between different cells and experimental conditions (3,5) can be modeled and predicted within a common parametric framework. It is also important to acknowledge that we have chosen from only four hypothetical observable effects of electrostatic. Other electrostatic effects may be reasonably included in the modeling process; for example, the EF may induce changes to the rate of polarization and depolarization (3). The electrostatic 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 hypothesized electrostatic effect. However, the characteristics of electrostatic 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 electrostatic behavior. However, there is a wealth of data and analyses on electrostatic in the context of cell populations (3,4,6–9,13). The electrostatic model in this work 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 electrostatic. 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 work 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 toward the real-time model predictive control of populations of electrostatic cells.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2021.06.034>.

AUTHOR CONTRIBUTIONS

The experiments and image analysis were performed by K.Z. and M.Z.; the modeling, simulation, and inference were carried out by T.P.P. and R.E.B.; all authors contributed to writing the manuscript.

ACKNOWLEDGMENTS

The experiments in the Zhao lab were supported by the Air Force Office of Scientific Research (AFOSR) through the Multidisciplinary Research Program of the University Research Initiative (MURI) grant FA9550-16-1-0052 and National Institutes of Health grant 1R01EY019101. T.P.P. and R.E.B. are supported by Biotechnology and Biological Sciences Research

Council (BBSRC) through grant BB/R000816/1, and R.E.B. 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. *Nat. Rev. Cancer*. 18:296–312.
2. Kai, H., T. Yamauchi, ..., M. Nishizawa. 2017. Accelerated wound healing on skin by electrical stimulation with a bioelectric plaster. *Adv. Healthc. Mater*. 6:1700465.
3. Cortese, B., I. E. Palamà, ..., G. Gigli. 2014. Influence of electrotaxis on cell behaviour. *Integr. Biol.* 6:817–830.
4. Cho, Y., M. Son, ..., J. H. Shin. 2018. Electric field-induced migration and intercellular stress alignment in a collective epithelial monolayer. *Mol. Biol. Cell*. 29:2292–2302.
5. Saltukoglu, D., J. Grünewald, ..., M. Simons. 2015. Spontaneous and electric field-controlled front-rear polarization of human keratinocytes. *Mol. Biol. Cell*. 26:4373–4386.
6. Zhao, M., A. Agius-Fernandez, ..., C. D. McCaig. 1996. Directed migration of corneal epithelial sheets in physiological electric fields. *Invest. Ophthalmol. Vis. Sci.* 37:2548–2558.
7. Zhao, M., B. Song, ..., C. D. McCaig. 2003. Direct visualization of a stratified epithelium reveals that wounds heal by unified sliding of cell sheets. *FASEB J.* 17:397–406.
8. Cohen, D. J., W. J. Nelson, and M. M. Maharbiz. 2014. Galvanotactic control of collective cell migration in epithelial monolayers. *Nat. Mater*. 13:409–417.
9. Sun, Y.-H., Y. Sun, ..., A. Mogilner. 2018. Electric fields accelerate cell polarization and bypass myosin action in motility initiation. *J. Cell. Physiol.* 233:2378–2385.
10. Hughes, R. M., and D. S. Lawrence. 2014. Optogenetic engineering: light-directed cell motility. *Angew. Chem. Int. Ed. Engl.* 53:10904–10907.
11. Wang, E., M. Zhao, ..., C. D. McCaig. 2003. Bi-directional migration of lens epithelial cells in a physiological electrical field. *Exp. Eye Res.* 76:29–37.
12. Lee, R. M., H. Yue, ..., W. Losert. 2017. Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. *J. R. Soc. Interface*. 14:20170147.
13. Camley, B. A., and W.-J. Rappel. 2017. Physical models of collective cell motility: from cell to tissue. *J. Phys. D Appl. Phys.* 50:113002.
14. Matsiaka, O. M., C. J. Penington, ..., M. J. Simpson. 2018. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics. *Bull. Math. Biol.* 80:738–757.
15. Nakajima, K. I., K. Zhu, ..., M. Zhao. 2015. KCNJ15/Kir4.2 couples with polyamines to sense weak extracellular electric fields in galvanotaxis. *Nat. Commun.* 6:8532.
16. Zhu, K., Y. Takada, ..., M. Zhao. 2019. Expression of integrins to control migration direction of electrotaxis. *FASEB J.* 33:9131–9141.
17. Erban, R., and S. J. Chapman. 2019. Stochastic Modelling of Reaction–Diffusion Processes. Cambridge University Press, Cambridge, UK.
18. Pavliotis, G. 2014. Stochastic Processes and Applications: Diffusion Processes, The Fokker-Planck and Langevin Equations. Springer, New York.
19. Gelman, A., J. B. Carlin, ..., D. B. Rubin. 2013. Bayesian Data Analysis, Third Edition. Chapman and Hall/CRC, Boca Raton, FL.
20. Bezanson, J., A. Edelman, ..., V. B. Shah. 2017. Julia: a fresh approach to numerical computing. *SIAM Rev.* 59:65–98.
21. Price, L. F., C. C. Drovandi, ..., D. J. Nott. 2017. Bayesian synthetic likelihood. *J. Comput. Graph. Stat.* 27:1–11.
22. Priddle, J. W., S. A. Sisson, ..., C. Drovandi. 2019. Efficient Bayesian synthetic likelihood with whitening transformations. *arXiv*, arXiv:1909.04857 <https://arxiv.org/abs/1909.04857>.
23. Picchini, U., U. Simola, and J. Corander. 2020. Adaptive MCMC for synthetic likelihoods and correlated synthetic likelihoods. *arXiv*, arXiv:2004.04558 <https://arxiv.org/abs/2004.04558>.
24. Hainy, M., C. C. Drovandi, and J. M. McGree. 2016. Likelihood-free extensions for Bayesian sequentially designed experiments. In *mODA 11 - Advances in Model-Oriented Design and Analysis: Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis Held in Hamminkeln, Germany, June 12–17, 2016* J. Kunert, C. H. Müller, and A. C. Atkinson, eds.. Springer International Publishing, pp. 153–161.
25. Del Moral, P., A. Doucet, and A. Jasra. 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.* 22:1009–1020.
26. Toni, T., D. Welch, ..., M. P. H. Stumpf. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface*. 6:187–202.
27. Prescott, T. P., and R. E. Baker. 2020. Multifidelity approximate Bayesian computation with sequential Monte Carlo parameter sampling. *arXiv*, arXiv:2001.06256v1 <https://arxiv.org/abs/2001.06256v1>.
28. Akiyama, M., T. Sushida, ..., H. Haga. 2017. Mathematical model of collective cell migrations based on cell polarity. *Dev. Growth Differ.* 59:471–490.
29. Vanegas-Acosta, J. C., D. A. Garzón-Alvarado, and A. P. Zwamborn. 2012. Mathematical model of electrotaxis in osteoblastic cells. *Bioelectrochemistry*. 88:134–143.
30. Binny, R. N., A. James, and M. J. Plank. 2016. Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bull. Math. Biol.* 78:2277–2301.
31. Warne, D. J., R. E. Baker, and M. J. Simpson. 2017. Optimal quantification of contact inhibition in cell populations. *Biophys. J.* 113:1920–1924.
32. Prescott, T. P., and R. E. Baker. 2020. Multifidelity approximate Bayesian computation. *SIAM/ASA J. Uncertain. Quant.* 8:114–138.
33. Zhang, Y., G. Xu, ..., M. Zhao. 2017. Collective cell migration has distinct directionality and speed dynamics. *Cell. Mol. Life Sci.* 74:3841–3850.
34. Mesbah, A. 2016. Stochastic model predictive control: an overview and perspectives for future research. *IEEE Control Syst.* 36:30–44.

Biophysical Journal, Volume 120

Supplemental information

Quantifying the impact of electric fields on single-cell motility

Thomas P. Prescott, Kan Zhu, Min Zhao, and Ruth E. Baker

Supplementary Material

Quantifying the impact of electric fields on single-cell motility

TP Prescott* ^{1,3}, K Zhu², M Zhao², and RE Baker¹

¹*Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford
OX2 6GG, UK*

²*Department of Ophthalmology and Vision Science, Department of Dermatology, Institute for
Regenerative Cures, University of California, Sacramento, CA 95817, USA*

³*Alan Turing Insitute, 96 Euston Road, London NW1 2DB, UK*

1 Preliminary data analysis

To justify the assumption that the cell velocity evolves according to a random walk, we performed an initial analysis of both control data replicates. In Figure S1(a,b) we plot the distribution of displacement distances travelled in each five-minute sampling interval over the five-hour experiment, smoothed using a kernel density estimate, and where each curve corresponds to a different cell. These distributions in displacement distances are consistent with the modelling assumption that cells can spontaneously polarise, in which state they travel at a positive speed, and can also transiently depolarise.

Furthermore, we can also justify our assumption that the direction of motion of each cell drifts stochastically over time. Figure S1(c,d) plots the autocorrelations between the observed angle of cell displacements over each five-minute interval for each cell trajectory, measured for time lags from 5 min to 60 min. This figure demonstrates that the autocorrelation in the observed motility direction decays with increasing time lag. This observation is consistent with the modelling assumption that the velocity direction evolves according to a random walk.

2 Mathematical model of electrotaxis

The autonomous model of cellular velocity in Eq. (2) is given by

$$\mathbf{v}(t) = v\mathbf{p}(t), \tag{2a}$$

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

*tprescott@turing.ac.uk

and the electrotactic model in Eq. (3) is given by

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

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

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

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

Here, cell velocity is denoted by \mathbf{v} and cell polarity is denoted by \mathbf{p} . The vector \mathbf{u} represents the EF with magnitude $\|\mathbf{u}(t)\| = u(t)$, scaled such that $u(t) = \alpha$ represents a field of strength 200α mV mm⁻¹ applied at time t in the direction parallel with $\mathbf{u}(t)$. The two-dimensional standard Wiener process is denoted by \mathbf{B} , and $\hat{\mathbf{p}}$ is the unit vector in the direction of polarity. Both models depend on the parameters v , with units $\mu\text{m min}^{-1}$, and D , with units min^{-1} . Thus \mathbf{p} is a non-dimensional quantity. The additional parameters, $\gamma_1, \dots, \gamma_4$, in the electrotactic model parametrise the four hypothesised electrotactic effects, as described in the main text.

Also common to both models is the potential function $W(\mathbf{p})$. This function is defined to capture the intended features of the autonomous model, namely that cells stochastically and spontaneously polarise, and that the direction of the polarity is uniformly distributed at stationarity. Denoting $p = |\mathbf{p}|$, it follows from the latter requirement that the potential function $W(\mathbf{p}) = W(p)$ must be radially symmetric. The interpretation of the parameter v as the modal speed of a polarised cell also implies that the polarised state is characterised by \mathbf{p} stochastically evolving in the regime $p \approx 1$. We therefore require a potential function with local minimum at $p = 1$. Following [11], this function is implemented as

$$W(p) = \beta \left(\frac{1}{4} p^4 - \frac{1}{2} p^2 \right), \quad (S1)$$

where $\beta > 0$ defines the local minimum value of the well at $p = 1$. It can be shown [17, 18] that the rate at which the cell transiently depolarises is determined solely by the timescale parameter, D min⁻¹, and the non-dimensional value of the *energy barrier*, $\Delta W = \beta/4$. Hence, we calibrate the models in Eq. (2) and Eq. (3) by inferring the common parameters, v , D , and ΔW , together with the parameters $\gamma_1, \dots, \gamma_4$ specific to the electrotactic model.

Note on polarity definition Our description of the model interprets the variable \mathbf{p} as the cell polarity, and treats velocity as the combination of a polarity component and a component due to the EF. Another interpretation of \mathbf{p} is available if we specifically define single-cell polarity as the non-dimensionalisation of the velocity by v . In the electrotactic model, this alternative definition identifies cell polarity as the variable

$$\mathbf{v}/v = (1 + \gamma_2 u + \gamma_3 \mathbf{u} \cdot \hat{\mathbf{p}}) \mathbf{p} + \gamma_1 \mathbf{u}.$$

The variable \mathbf{p} , with dynamics (3d), is then interpreted as a slowly-responding component of the cell polarity (in the alternative definition) to the EF input, while $\gamma_1 \mathbf{u}$, identifiable with *velocity bias*, is an instantly-responding component of the cell polarity. However, these definitions are internal to the model, in the sense that they have no effect on the observable position or velocity of simulated cells. Thus, in the current work, we choose to identify ‘cell polarity’ as the modelled variable \mathbf{p} , while noting that alternative interpretations are possible.

3 Likelihood-free Bayesian inference

To infer the parameters of the model, we will use the data from both experimental assays. We use all recorded cell positions from the autonomous experiment over $t \in [0, 300]$, denoting this data set by \mathbf{x}_{NoEF} . In addition, we use the recorded cell positions from the electrotactic experiment, but only over $t \leq 180$ min, denoting this data set by \mathbf{x}_{EF} .

Note that the data from the electrostatic experiment gathered over $t > 180$ min will be held back as test data. The Bayesian inference framework uses the experimental training data, \mathbf{x}_{NoEF} and \mathbf{x}_{EF} , to update a prior distribution, $\pi(\theta)$, into a posterior distribution, $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, by multiplying by the *likelihood*, $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta)$, according to Bayes' rule,

$$\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) \propto \mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta)\pi(\theta).$$

To define the likelihood, we first consider simulations of the models in Eq. (2) and Eq. (3).

For a given parameter vector, θ , initial polarity, \mathbf{p}_0 , and non-zero EF input, $\mathbf{u}(t)$, the model in Eq. (3) is simulated and a trajectory, $\omega = (\mathbf{p}(t), \mathbf{x}(t))_{t \geq 0}$, is produced. This stochastic trajectory has conditional density $p(\omega | \theta, \mathbf{p}_0, \mathbf{u}(t))$. We assume that there is a known distribution, $\varphi(\mathbf{p}_0)$, for the initial polarity: for the inference procedure carried out in the main text, we assume that φ is a Gaussian distribution with zero mean and diagonal covariance matrix, with component-wise variances of 0.1. For the two specific experimental inputs, $\mathbf{u}_{\text{NoEF}}(t) \equiv 0$ and

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

we integrate the density p with respect to $\varphi(\mathbf{p}_0)$ and thus define two densities,

$$p_{\text{NoEF}}(\omega | \theta) = \int p(\omega | \theta, \mathbf{p}_0, \mathbf{u}_{\text{NoEF}})\varphi(\mathbf{p}_0) d\mathbf{p}_0, \quad (\text{S2})$$

$$p_{\text{EF}}(\omega | \theta) = \int p(\omega | \theta, \mathbf{p}_0, \mathbf{u}_{\text{EF}})\varphi(\mathbf{p}_0) d\mathbf{p}_0, \quad (\text{S3})$$

for trajectories simulated by the autonomous and electrostatic models, respectively. Each observed trajectory in the experimental training data set, $\mathbf{x}_{\text{NoEF},i}$ and $\mathbf{x}_{\text{EF},i}$, thus defines a set in the simulation space,

$$\begin{aligned} \Omega(\mathbf{x}_{\text{NoEF},i}) &= \left\{ \omega = (\mathbf{x}(t), \mathbf{p}(t))_{t \geq 0} : \mathbf{x}(t_j) = \mathbf{x}_{\text{NoEF},i}(t_j) \forall j = 0, \dots, 60 \right\}, \\ \Omega(\mathbf{x}_{\text{EF},i}) &= \left\{ \omega = (\mathbf{x}(t), \mathbf{p}(t))_{t \geq 0} : \mathbf{x}(t_j) = \mathbf{x}_{\text{EF},i}(t_j) \forall j = 0, \dots, 36 \right\}, \end{aligned}$$

of all simulated trajectories that are indistinguishable from the observed training data. We thus define the likelihoods of each simulation as

$$\begin{aligned} \mathcal{L}_{\text{NoEF}}(\mathbf{x}_{\text{NoEF},i} | \theta) &= \int_{\Omega(\mathbf{x}_{\text{NoEF},i})} p_{\text{NoEF}}(\omega | \theta) d\omega, \\ \mathcal{L}_{\text{EF}}(\mathbf{x}_{\text{EF},i} | \theta) &= \int_{\Omega(\mathbf{x}_{\text{EF},i})} p_{\text{EF}}(\omega | \theta) d\omega, \end{aligned}$$

for each cell index, i . The likelihoods of each trajectory thus combine to give the posterior,

$$\begin{aligned} \pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) &\propto \mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta)\pi(\theta) \\ &= \prod_i \mathcal{L}_{\text{NoEF}}(\mathbf{x}_{\text{NoEF},i} | \theta) \prod_j \mathcal{L}_{\text{EF}}(\mathbf{x}_{\text{EF},j} | \theta)\pi(\theta). \end{aligned} \quad (\text{S4})$$

However, it is clear that the likelihood of each of the experimentally observed trajectories cannot easily be calculated. We therefore identified the posterior parameter distribution using a likelihood-free (i.e. simulation-based) Bayesian inference approach, harnessing the concept of *synthetic likelihoods*.

3.1 Synthetic likelihoods

We focus on the autonomous case first; the electrotactic case follows similarly, with an obvious change of notation. The synthetic likelihood approach approximates the likelihoods, $\mathcal{L}_{\text{NoEF}}(\mathbf{x}_{\text{NoEF},i} \mid \theta)$ in two stages. The first stage is to reduce the dimension of the data space by defining a function of the simulated and observed trajectories that maps the data to a low-dimensional summary statistic. The second stage is to (a) use repeated simulation of the summarised model at the parameter value θ to fit an empirical multivariate Gaussian distribution for the summary statistic, and then (b) to approximate the likelihood with the *synthetic likelihood* of the experimental data, defined as the likelihood of the summarised data under the fitted empirical Gaussian distribution.

We define the function $Y : \omega \mapsto \mathbb{R}^4$ for the simulated trajectory $\omega = (\mathbf{x}(t), \mathbf{p}(t))$ on $t \in [t_n, t_{n+m}]$ as:

$$Y_1(\omega) = (\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)) \cdot \mathbf{i} \quad (\text{S5a})$$

$$Y_2(\omega) = \|\mathbf{x}(t_{n+m}) - \mathbf{x}(t_n)\|, \quad (\text{S5b})$$

$$Y_3(\omega) = \sum_{r=1}^m \|\mathbf{x}(t_{n+r}) - \mathbf{x}(t_{n+r-1})\|, \quad (\text{S5c})$$

$$Y_4(\omega) = \left(\frac{1}{m} \sum_{r=1}^m \|\mathbf{x}(t_{n+r}) - \mathbf{x}(t_{n+r-1}) - Y_3/m\|^2 \right)^{1/2}, \quad (\text{S5d})$$

for sample time points $t_j = 5j$ min. Thus, the entries of the vector $Y(\omega)$ denote the random values of

- the displacement over the interval $[t_n, t_{n+m}]$ in the positive x -direction,
- the net displacement,
- the path length,
- and the standard deviation of cell displacements over five-minute intervals,

for stochastic simulations ω of the electrotactic model in Eq. (3), given θ , \mathbf{p}_0 , and $\mathbf{u}(t)$. Note that we can also calculate the values of the function Y in Eq. (S5) for the experimentally observed data, $\mathbf{x}_{\text{NoEF},i}$, for each cell index, i . With a slight abuse of notation, we denote the summarised experimental data by $y_{\text{NoEF},i} = Y(\mathbf{x}_{\text{NoEF},i})$.

For a given value of θ , the synthetic likelihood approach [21–23] assumes that the random value of $Y(\omega)$ under the density $p_{\text{NoEF},i}(\omega \mid \theta)$ is a Gaussian random variable with parameter-dependent mean $\mu_{\text{NoEF}}(\theta)$ and covariance $\Sigma_{\text{NoEF}}(\theta)$. We estimate this mean and covariance with the sample mean and covariance of simulated summary statistics $Y(\omega_k)$, for $k = 1, \dots, n$, produced by simulating the autonomous model n times using the parameter value θ over the interval $t \in [0, 300]$. The resulting approximation of each trajectory’s likelihood, $\tilde{\mathcal{L}}_{\text{NoEF},n} \approx \mathcal{L}_{\text{NoEF}}$, is summarised as

$$\tilde{\mathcal{L}}_{\text{NoEF},n}(\mathbf{x}_{\text{NoEF},i} \mid \theta) = \mathcal{N}\left(y_{\text{NoEF},i} \mid \hat{\mu}_{\text{NoEF}}(\theta), \hat{\Sigma}_{\text{NoEF}}(\theta)\right) \quad i = 1, \dots, 50, \quad (\text{S6a})$$

$$\hat{\mu}_{\text{NoEF}}(\theta) = \frac{1}{n} \sum_{k=1}^n Y(\omega_k), \quad (\text{S6b})$$

$$\hat{\Sigma}_{\text{NoEF}}(\theta) = \frac{1}{n} \sum_{k=1}^n (Y(\omega_k) - \hat{\mu}_{\text{NoEF}}(\theta))(Y(\omega_k) - \hat{\mu}_{\text{NoEF}}(\theta))^T, \quad (\text{S6c})$$

$$\omega_k \sim p_{\text{NoEF}}(\cdot \mid \theta) \quad k = 1, \dots, n, \quad (\text{S6d})$$

where \mathcal{N} denotes the Gaussian density and where the chosen number of simulations, n , needs to be appropriately large [22]. In our case, we choose $n = 500$.

In the case of the electrostatic model with piecewise constant electric field input, $\mathbf{u}_{\text{EF}}(t)$, we adapt the procedure above to summarise the simulations and the training data with an eight-dimensional summary statistic. Treating the intervals $t \in [0, 60], [60, 180], [180, 300]$ separately, we summarise simulations, $\omega \sim p_{\text{NoEF}}(\cdot | \theta)$, and observations, $\mathbf{x}_{\text{EF},i}$, by calculating Y for each interval. We use only $t \leq 180$ to calibrate the model, while holding back the interval $t \in [180, 360]$ for the purpose of testing. Thus, we combine the values of Y for the intervals $t \in [0, 60]$ and $t \in [60, 180]$ into an eight-dimensional summary for calibrating the model. Using this eight-dimensional summary, we adapt the synthetic likelihood procedure summarised in Eq. (S6) to calculate $\tilde{\mathcal{L}}_{\text{EF},n}(\mathbf{x}_{\text{EF},i} | \theta)$.

Finally, we can multiply each of these trajectory synthetic likelihoods into an overall synthetic likelihood for the experimental data,

$$L_{\text{NoEF},n}(\theta) = \prod_i \tilde{\mathcal{L}}_{\text{NoEF},n}(\mathbf{x}_{\text{NoEF},i} | \theta), \quad (\text{S7a})$$

$$L_{\text{EF},n}(\theta) = \prod_i \tilde{\mathcal{L}}_{\text{EF},n}(\mathbf{x}_{\text{EF},i} | \theta), \quad (\text{S7b})$$

$$\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} | \theta) \approx L_n(\theta) = L_{\text{NoEF},n}(\theta)L_{\text{EF},n}(\theta), \quad (\text{S7c})$$

each calculation of which requires n simulations of the autonomous model and n of the electrostatic model.

3.2 SMC inference

In order to produce a sample from the posterior distribution, we use sequential Monte Carlo (SMC) with synthetic likelihoods [21–27], as outlined in Algorithm 1. This method is chosen in order to exploit parallelisation, mitigating the computational burden of MCMC-based approaches that is incurred due to the large numbers of model simulations required for accurate likelihood-free inference. SMC defines a sequence of intermediate importance distributions that evolve towards the target posterior. This approach is particularly useful in comparison to naive rejection sampling: since we will use non-informative priors, rejection sampling is too inefficient, as it proposes parameters in extremely low-likelihood regions of parameter space too frequently. Importantly, for each value of θ , the stochastic values of $L_{\text{NoEF},n}(\theta)$ and $L_{\text{EF},n}(\theta)$ can be called multiple times. At each call of these two likelihoods, we do *not* recycle previously computed values for the synthetic likelihoods but instead simulate the models again. Although this approach slows the inference procedure, it is necessary to ensure the correct stationary distribution of the Markov chain steps, (13–14).

In Algorithm 1 we produce a weighted sample from the Bayesian synthetic likelihood approximation to the posterior, $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$. The intermediate distributions at each iteration are proportional to the tempered distributions

$$\pi_T(\theta) \propto [L_{\text{NoEF},n}(\theta)L_{\text{EF},n}(\theta)]^T \pi(\theta),$$

where the sequence of temperatures T evolves from 0 to 1. In Algorithm 1, we define the initial perturbation kernel, $K(\cdot | \theta)$ to be a multivariate Gaussian density with mean θ and diagonal covariance matrix, with component-wise variances of 0.01, 0.0025 and 0.0001 for v , ΔW , and D , respectively, and 0.01 for each γ_i . The effective sample size, ESS , used to adaptively choose the increment in temperature, is defined as

$$ESS(\{W_i\}) = \left(\sum_i W_i \right)^2 / \sum_j W_j^2, \quad (\text{S8})$$

for any finite set of sample weights, W_i . We use $N = 1000$ particles, setting $N_{\text{min}} = 333$ as the effective sample size triggering resampling, and setting $\alpha = 0.8$ as the decay rate of the effective sample size.

The effective sample size of the sample produced by Algorithm 1, as defined in Equation (S8), can be an over-

Algorithm 1 Synthetic Likelihood SMC

Input: Observed summary statistics y_{NoEF} and y_{EF} ; prior π ; perturbation kernel $K(\cdot | \theta)$; .

Output: Weighted sample set of parameters θ_i with weights W_i , from the synthetic likelihood approximation to the posterior $\pi(\theta | \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$.

- 1: Sample N independent θ_i from π .
- 2: Set weights $W_i^0 = 1/N$ for $i = 1, \dots, N$.
- 3: Initialise $T = 0$ and $r = 0$.
- 4: **repeat**
- 5: Update $r \leftarrow r + 1$.
- 6: Find $\Delta T \in [\Delta T_{\min}, 1 - T]$ to solve $ESS(\{W_i^r\}) = \alpha ESS(\{W_i^{r-1}\})$, for weights W_i^r such that

$$\log W_i^r = \log W_i^{r-1} + \Delta T (\log L_{\text{NoEF},n}(\theta_i) + \log L_{\text{EF},n}(\theta_i)),$$

for the synthetic likelihoods, $L_{\text{NoEF},n}(\theta_i)$ and $L_{\text{EF},n}(\theta_i)$. Use $\Delta T = \Delta T_{\min}$ or $\Delta T = 1 - T$ if $ESS(\{W_i^r\})$ is, respectively, uniformly less than or uniformly greater than $\alpha ESS(\{W_i^{r-1}\})$ on the interval $[\Delta T_{\min}, 1 - T]$.

- 7: Update $T \leftarrow T + \Delta T$.
- 8: **if** $ESS(\{W_i^r\}) < N_{\min}$ **then**
- 9: Resample from $\{\theta_i\}$ according to weights W_i^r .
- 10: Reset weights $W_i^r = 1/N$.
- 11: Reset covariance of perturbation kernel, K , to equal the empirical covariance of the resampled parameter set, $\{\theta_i\}$.
- 12: **end if**
- 13: Propose perturbed parameter values $\theta_i^* \sim K(\cdot | \theta_i)$ for $i = 1, \dots, N$.
- 14: Accept $\theta_i \leftarrow \theta_i^*$ with Metropolis–Hastings probability

$$\left[\frac{L_{\text{NoEF},n}(\theta_i^*) L_{\text{EF},n}(\theta_i^*)}{L_{\text{NoEF},n}(\theta_i) L_{\text{EF},n}(\theta_i)} \right]^T \frac{\pi(\theta_i^*) K(\theta_i^* | \theta_i)}{\pi(\theta_i) K(\theta_i | \theta_i^*)}$$

for $i = 1, \dots, N$.

- 15: **until** $T = 1$.
-

estimate of the sample quality. The resampling step, in lines (9–11), tends to result in multiple particles with equal parameter values. These particles do not always separate through the single MCMC step in (13–14). The replicated parameter values in the sample thus degrade its quality, without being captured by the ESS calculation. To improve the quality of the final sample produced by Algorithm 1, we post-process the posterior. First, the final weighted Monte Carlo sample output from Algorithm 1 is resampled according to steps (9–11). Then, for each resampled particle, we calculate 100 MCMC steps (13–14) using $T = 1$. This procedure effectively produces 1000 short Markov chains of length 100, each beginning from a particle from the target distribution, and with stationary distribution equal to the target distribution. We use the end samples of each of these Markov chains as a higher-quality, 1000-particle sample from the Bayesian synthetic likelihood approximation to the posterior, each with equal weight. The resulting particles are checked to represent distinct parameter values, and thus every sample generated is of size 1000, where ‘size’ refers to both particle numbers and effective sample size, which are identical.

4 Model validation

4.1 Autonomous data cross-validation

We first evaluate the modelling approach by comparing the model outputs trained to each of the two replicates alone in turn, with the other replicate held back. Note that to produce a posterior based on the autonomous (control) data set alone, we can simply adapt Algorithm 1 to use only $L_{\text{NoEF},n}$ in calculating synthetic likelihoods (effectively setting $L_{\text{EF},n} = 1$). The one-dimensional marginals of the resulting posterior, trained on the entire autonomous data set, are shown in Figure 2 of the main text. The covariance structure of the posterior sample is depicted in Figure S2.

If training on each of the two control data sets alone, we can construct an additional two posteriors. Figure S3 demonstrates that these posteriors closely overlap with one another, and with the posterior trained on the combined data set. In addition to comparing the posteriors produced by each replicate, we can use repeated simulation to produce posterior predictive distributions for each of the four summary statistics used for inference. These can be compared to each of the observed data sets. In Figure S4, we plot posterior predictive distributions, based on 10 simulations for each of the 1000 sampled parameter values, for each of the three posteriors. In this figure, we demonstrate that there is a good agreement between these posterior predictive distributions and the empirically observed summary statistics for each replicate of the autonomous experiment.

We can quantify this agreement by estimating the log-likelihood of each data set, using the maximum likelihood normal approximations to each of the three empirical distributions depicted in Figure S4. The quantified cross-validation is shown in Table S1. For the posterior trained on replicate 1, the posterior predictive log-likelihood of the data from replicate 1 is, as expected, slightly greater than that of replicate 2. However, the drop in log-likelihood for the replicate 2 data is minimal, and thus the out-of-sample predictive performance of the posterior trained against replicate 1 is good. A similar cross-validation argument, swapping the roles of replicate 1 and replicate 2, holds. Thus, Figure S4 and Table S1 validate the suitability of the calibrated autonomous model.

Finally, the preliminary analysis of the observed displacements in each replicate, depicted in Figure S1, can be replicated for simulated trajectories using parameters sampled from the posteriors trained on each replicate alone. Figure S5 shows that model simulations produce qualitatively similar behaviour to the observed data. Importantly, these characteristics of the trajectories were not explicitly used to calibrate the model. However, they have been replicated by the calibrated model, which provides further evidence that the autonomous model is accurate.

4.2 Electrotactic model validation

After validating the autonomous model against the control data, we perform the synthetic likelihood inference procedure on the full training data set, comprising each observed trajectory from both control data sets, $\mathbf{x}_{\text{NoEF},i}$ on $t \in [0, 300]$, and the first section of each observed trajectory from the electrotactic data sets, $\mathbf{x}_{\text{EF},i}$ on $t \in [0, 180]$. Algorithm 1 is used to construct 16 separate posteriors, based on each of the 16 priors, π_X , indexed by $X \subset \{1, 2, 3, 4\}$, where $\gamma_i > 0$ if and only if $i \in X$. The greatest posterior likelihood is given by choosing $X = \{4\}$, and thus setting $\gamma_1 = \gamma_2 = \gamma_3 = 0$. For this choice of prior, we plot the one-dimensional marginals of the resulting posterior in Figure 4 of the main text. To illustrate the covariance structure of the resulting posterior, the two-dimensional marginals are given in Figure S6.

To validate the resulting posterior, in Figure S7 we compare the one-dimensional marginals of the parameters v , ΔW , and D for the posteriors trained against the autonomous data set only, versus the posterior trained against the entire training data set. The resulting posteriors are similar, showing that calibrating to the full training data set refines the predictions of model calibrated to the autonomous data set alone.

Moreover, the posterior predictive distributions of the four summary statistics over 0 min to 300 min (for the autonomous experiment) and each of 0 min to 60 min and 60 min to 180 min (for the electrotactic experiment), depicted in Figure S8, show that the model is a close fit to the observed summarised trajectories in the training data set. In particular, similarly to the autonomous model above, we have split the training data between replicates 1 and 2, and trained posteriors on each. We can perform cross-validation analysis, using the quantification of the log-likelihood of the observed summary statistics according to the posterior log-likelihoods in Table S2. As with the autonomous model, this table shows that the posteriors trained on each replicate provide a good prediction of the log-likelihood of the summary statistics of the other replicate. Thus, the calibrated model provides good out-of-sample predictive performance.

Note that, in addition to this cross-validation approach, a further test of the model validity is depicted in Figure 5 of the main manuscript. We use the electrotactic model, calibrated to the training data, \mathbf{x}_{NoEF} and \mathbf{x}_{EF} , to predict the behaviour of the held-back test data set, $\mathbf{x}_{\text{Switch}}$, comprising the observed cell trajectories over $t \in [180, 360]$ after the EF switches direction from the positive to negative x -direction. We demonstrate that the posterior predictive distributions of the summary statistics are a good match to the unseen test data, further confirming the ability of the calibrated model to predict cellular motility under dynamic EF inputs.

List of Figures

S1	Preliminary analysis of control data replicates.	10
S2	Posterior sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}})$	11
S3	Posterior distributions trained to each replicate, and the combined data set.	12
S4	Posterior predictive distributions of summary statistics used for inference.	13
S5	Analysis of simulated displacements over 5 min intervals.	14
S6	Weighted sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$	15
S7	Comparison of one-dimensional projections of empirical posterior samples from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$ and $\pi(\theta \mid \mathbf{x}_{\text{NoEF}})$	16
S8	Posterior predictive distributions of summary statistics used for inference.	17

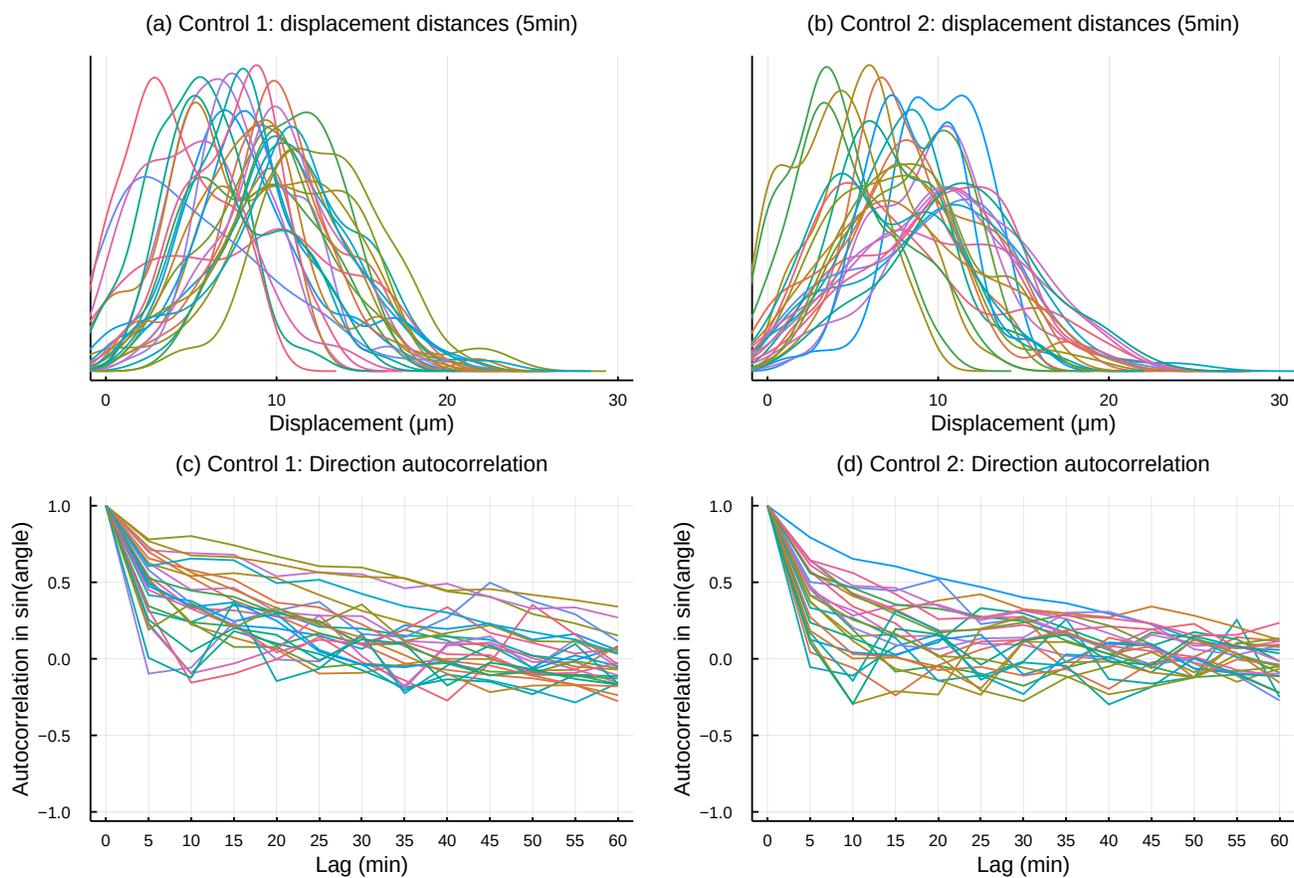


Figure S1: Preliminary analysis of autonomous data sets. (a, b) Distributions of observed displacement distances for each tracked cell in each autonomous data set. (c, d) Autocorrelations of observed displacement angles for all tracked cells in each autonomous data set, for intervals at lags of 5 min to 60 min.

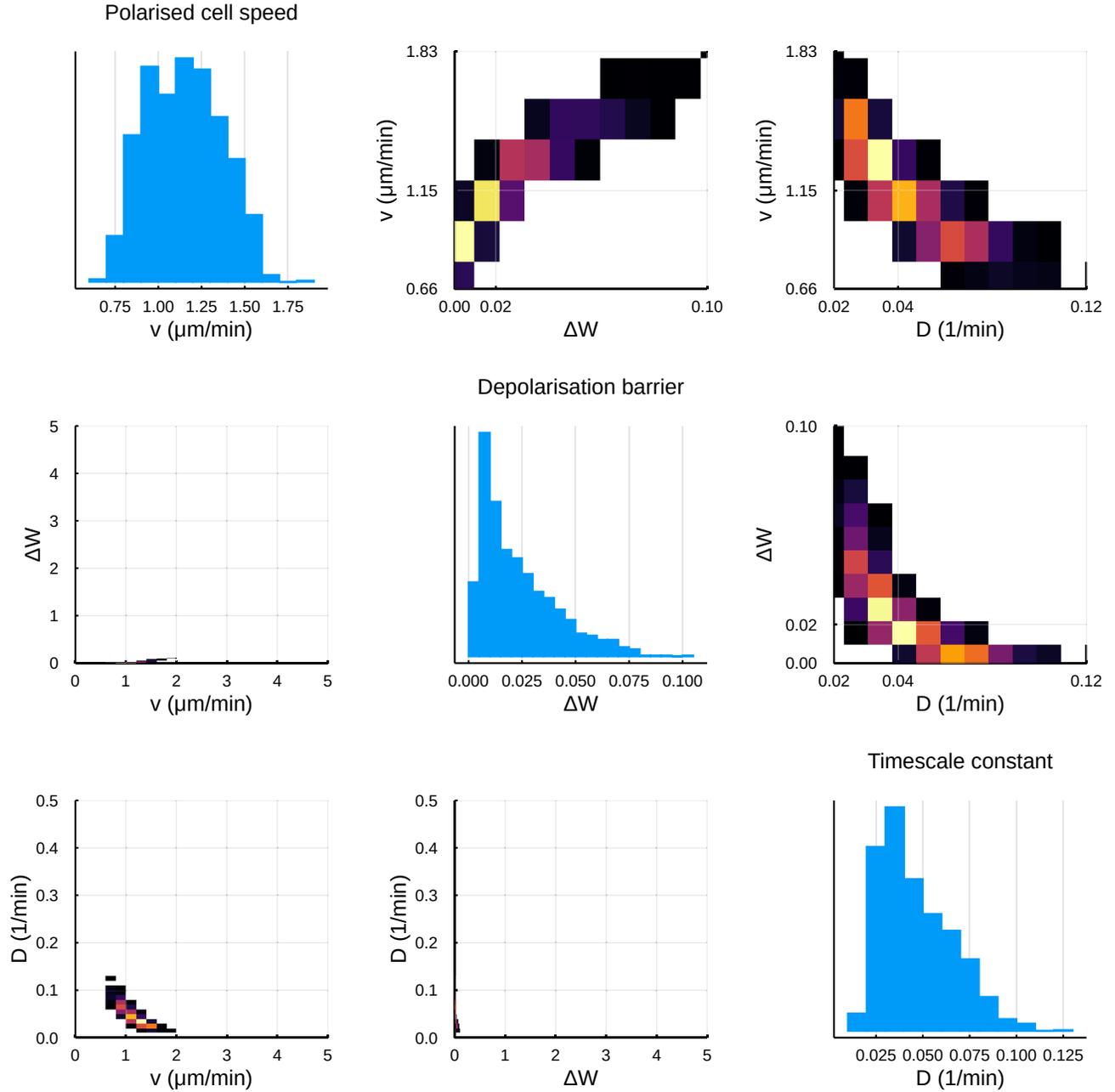


Figure S2: Posterior sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}})$, generated by Algorithm 1 neglecting $L_{\text{EF},n}$, for identified parameters, v , ΔW , and D . Diagonal plots are empirical histograms (as in Figure 2 in the main text). Off-diagonal heatmaps represent empirical pairwise distributions, where brighter colours correspond to greater density. Axes in the bottom-left are scaled to the prior support; the top-right scales are zoomed in to the region of greatest positive posterior likelihood.

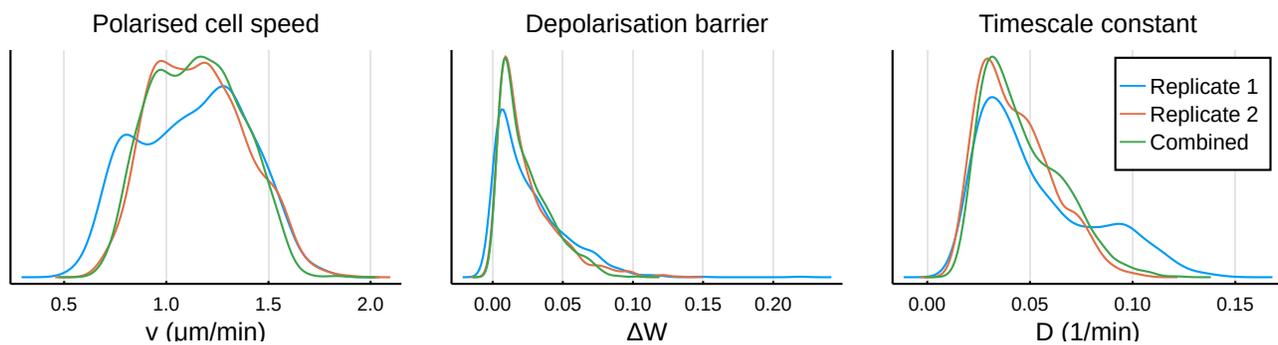


Figure S3: Posterior distributions for the parameters of the autonomous model, trained to each replicate, and also to the combined data set. Posterior samples have been generated using Algorithm 1, with MCMC post-processing, and depicted as densities using a default KDE.

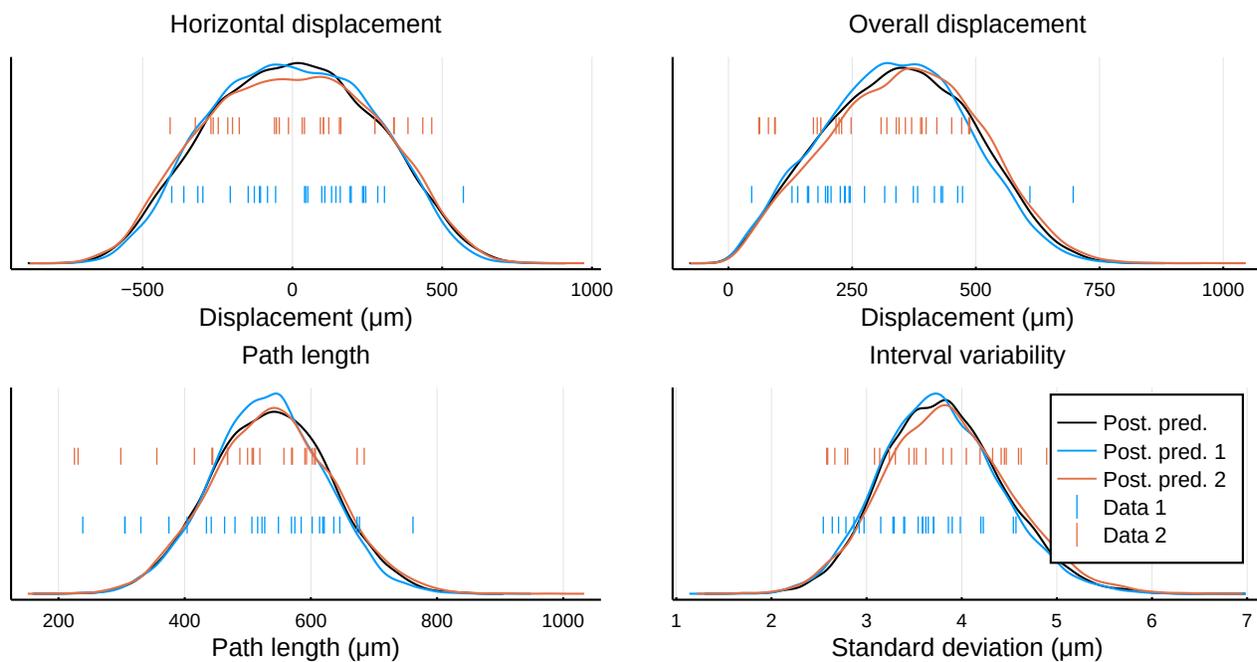


Figure S4: Posterior predictive distributions of the summary statistics used for inference, for autonomous data \mathbf{x}_{NoEF} only. For each posterior sample trained against replicate 1, replicate 2, and both replicates together, we simulated 10,000 summary statistics to produce three posterior predictive distributions (kernel density estimates, represented as solid curves). These are plotted with the observed summary statistics from each of the data sets. The posterior predictive distributions for each replicate can thus be cross-validated against the data in the other replicate: the log-likelihoods of the observations in each replicate, under each posterior predictive distribution, are recorded in Table S1.

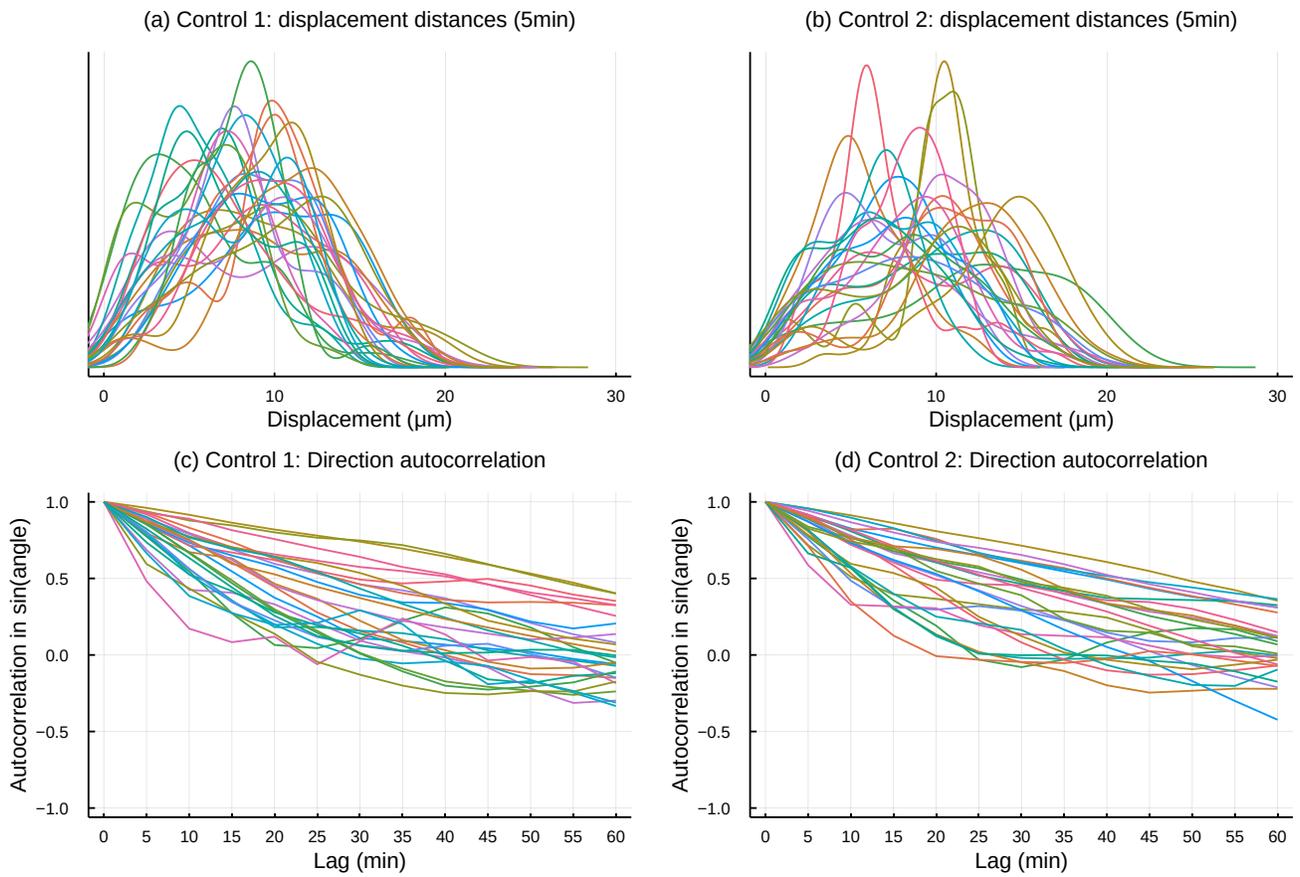


Figure S5: Analysis of simulated displacements over 5 min intervals. (a, b) Distributions of simulated displacement distances for cells simulated using parameters sampled from posteriors trained on each autonomous data set. (c, d) Autocorrelations of simulated displacement angles for cells simulated using parameters sampled from posteriors trained on each autonomous data set, for intervals at lags of 5 min to 60 min.

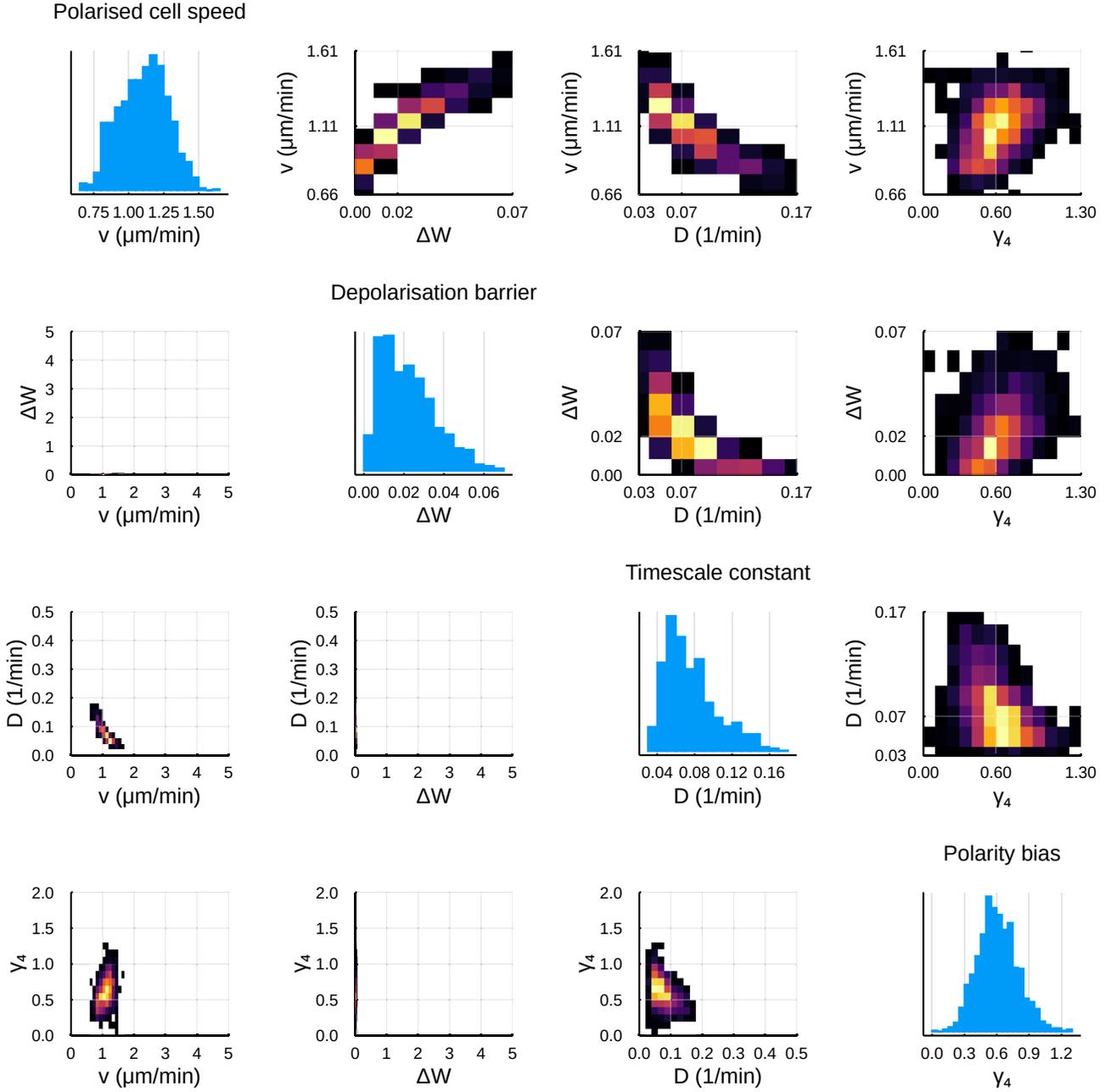


Figure S6: Weighted sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, generated by the completion of Algorithm 1 for identified parameters, v , ΔW , D , and γ_4 . Plots are as described in Figure S2.

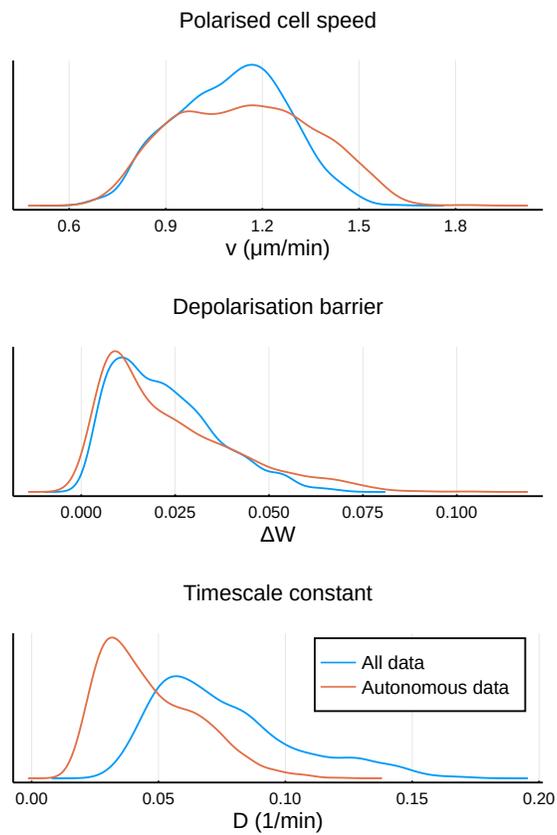


Figure S7: One-dimensional projections for v , ΔW , and D of the empirical posteriors $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$ and $\pi(\theta \mid \mathbf{x}_{\text{NoEF}})$, generated by Algorithm 1, trained on all data and the autonomous data, respectively.

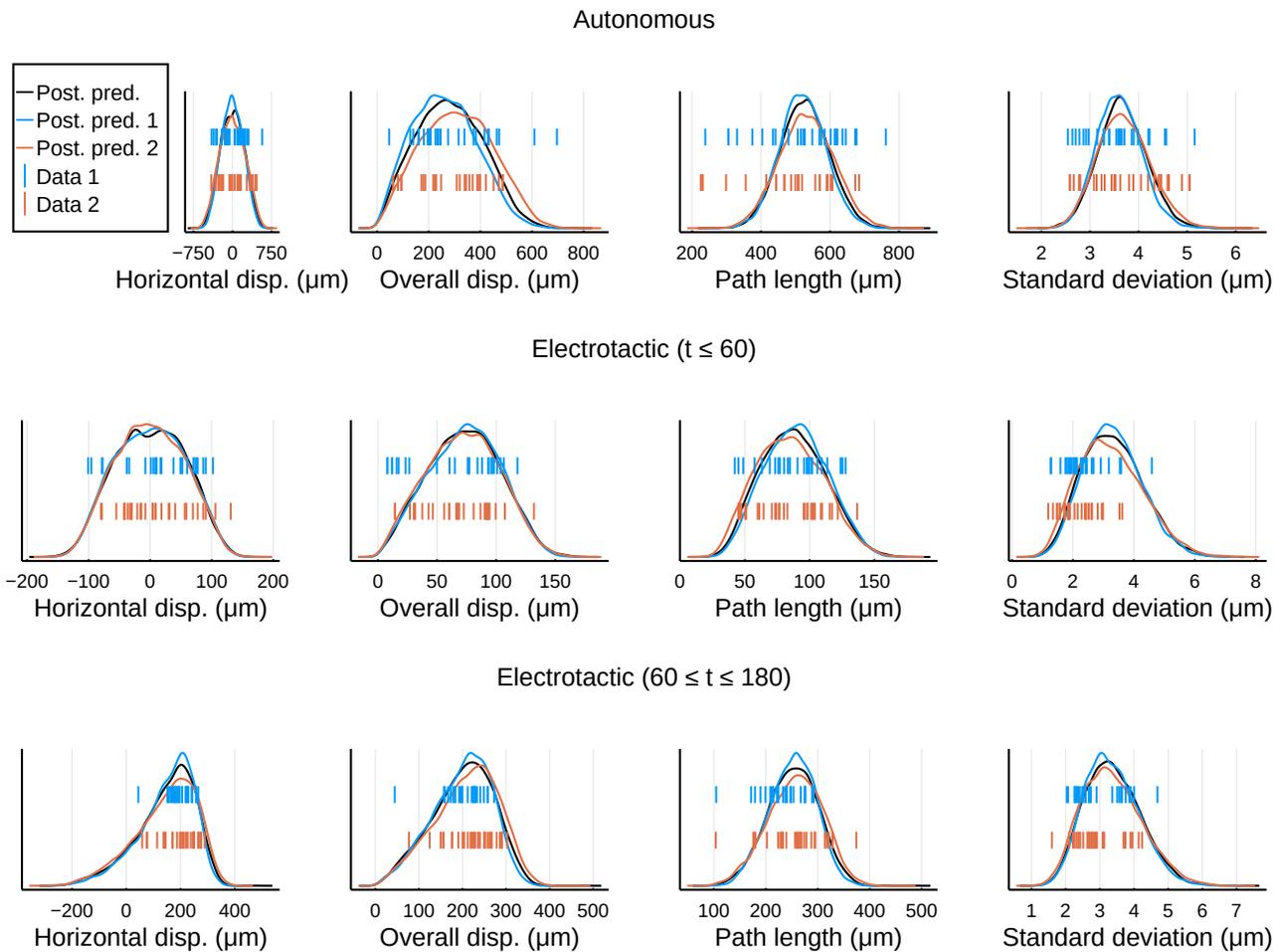


Figure S8: Posterior predictive distributions of all summary statistics used for inference with combined training data \mathbf{x}_{NoEF} and \mathbf{x}_{EF} . Details are as for Figure S4, with log-likelihoods of the observations under each posterior predictive distribution recorded in Table S2.

List of Tables

S1	Cross-validated log-likelihoods of training data under empirical posterior predictive distributions: autonomous data only.	19
S2	Cross-validated log-likelihoods of training data under empirical posterior predictive distributions: all training data.	20

Log likelihoods	Replicate 1 training data: 27 autonomous trajectories	Replicate 2 training data: 26 autonomous trajectories
Replicate 1 posterior	-558.6	-537.4
Replicate 2 posterior	-559.9	-537.2
Combined posterior	-559.3	-537.5

Table S1: Cross-validated log-likelihoods of training data under empirical posterior predictive distributions, trained on data from the autonomous experiment only. Entries (i, j) correspond to the log-likelihood of observed data from Replicate j under the posterior predictive distribution for posteriors trained on data from Replicate i , for $i, j \in \{1, 2\}$. The bottom row corresponds to log-likelihoods under the posterior predictive distribution for the posterior trained on the combined data set. This table quantifies the log-likelihood of Figure S4.

Log likelihoods	Replicate 1 training data: 27 autonomous trajectories 26 electrotactic trajectories	Replicate 2 training data: 26 autonomous trajectories 30 electrotactic trajectories
Replicate 1 posterior	-1468	-1550
Replicate 2 posterior	-1497	-1543
Combined posterior	-1479	-1547

Table S2: Cross-validated log-likelihoods of training data under empirical posterior predictive distributions, trained on all identified training data. Details are as for Table S1. This table quantifies the log-likelihood of Figure S8.