Stochastic differential equation modelling of cancer cell migration and tissue invasion

Dimitrios Katsaounis[1], Mark A.J. Chaplain[1] and Nikolaos Sfakianakis[1*]

[1]School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, KY16 9SS, Scotland, UK.

*Corresponding author(s). E-mail(s): n.sfakianakis@st-andrews.ac.uk;
Contributing authors: dk204@st-andrews.ac.uk; majc@st-andrews.ac.uk;

Abstract

Invasion of the surrounding tissue is a key aspect of cancer growth and spread involving a coordinated effort between cell migration and matrix degradation, and has been the subject of mathematical modelling for almost 30 years. In this current paper we address a long-standing question in the field of cancer cell migration modelling. Namely, identify the migratory pattern and spread of individual cancer cells, or small clusters of cancer cells, when the macroscopic evolution of the cancer cell colony is dictated by a specific partial differential equation (PDE).

We show that the usual heuristic understanding of the diffusion and advection terms of the PDE being one-to-one responsible for the random and biased motion of the solitary cancer cells, respectively, is not precise. On the contrary, we show that the drift term of the correct stochastic differential equation (SDE) scheme that dictates the individual cancer cell migration, should account also for the divergence of the diffusion of the PDE. We support our claims with a number of numerical experiments and computational simulations.

Keywords: Cancer invasion, multiscale modelling, hybrid continuum-discrete, coupled partial and stochastic partial differential equations

1 Introduction

Cancer invasion is a complex process involving numerous interactions between the cancer cells and the extracellular matrix (ECM) (cf. the tumour microenvironment) facilitated by matrix degrading enzymes. Along with active cell migration (both individual and collective) and increased/excessive proliferation, these processes enable the local spread of cancer cells into the surrounding tissue. Any encounter with blood or lymphatic vessels (cf. tumour-induced angiogenesis, lymph-angiogenesis) in the tumour microenvironment initiates and facilitates the spread of the cancer to secondary locations in the host, i.e., metastasis or metastatic spread. A comprehensive historical overview of the biology of metastastic spread can be found in the article by Talmadge and Fidler [1], while an overview of the core aspects of invasion can be found in the articles of Hanahan and Weinberg [2], [3] and the review article of Friedl and Wolf [4]. From a mathematical modelling perspective, cancer invasion has been a topic of interest for almost 30 years with a range of approaches and techniques being used, and an overview can be found in the recent review paper by Sfakianakis and Chaplain [5]. Broadly speaking, two different approaches have been used to model cancer invasion-continuum approaches (i.e. using differential equations with cancer cell density as one of the dependent variables) and individual-based or agent-based approaches (i.e. focusing on the movement of individual cells). Some have also adopted a so-called hybrid approach e.g. [6], deriving a discrete model governing the migration of individual cancer cells from the discretization of an associated system of PDEs.

It is not our intention here to discuss in detail the previous modelling work in the area. Rather the aim of this paper is to investigate mathematically the connection between the stochastic differential and the partial differential equations (SDEs and PDEs respectively) that are typically used to describe the migration of living cells. The precise interplay between the SDE and PDE approach is not yet clear and has been in the research focus the last years. The difference between these two approaches is significant and it lies primarily in the immediate focus of the mathematical model. Namely, whereas the SDE approach focuses primarily on the migratory behaviour of the individual cells, the PDE approach describes the macroscopic behavioural pattern of a large collective of cells. As the behavioural pattern, in real life biology, of large cell collectives is related to the migration of individual cells so should the two mathematical approaches be connected.

By identifying the interplay between the two mathematical approaches, we shed light in the complexity of multiscale modelling and simulations. This has direct implications in the biological understanding of solitary cancer cell migration and the development/growth of tumours. Such detailed knowledge of how far individual cancer cells can penetrate into the local tissue is very important from a surgical point of view and can help to minimise the amount of resection required, a point initially raised and investigated in the work of Anderson et al. [6].

Our aim, hence, in this work is to investigate the interplay between the SDE and PDE modeling approaches of cancer invasion and growth. Namely, we exhibit that the terms of the numerical scheme solving the underlying SDE are not in a one-to-one correspondence to the terms of the PDE. In more detail, we show that a particular "correction" in the drift terms of the numerical scheme of the SDE that improves the approximation qualities of the schemes when compared with the numerical solution of the PDE.

These ideas are studied in the remainder paper in the following way. In Section 2 we provide some background for the motivation to this paper stating the general forms of PDE and SDE to be considered, while in Section 3 we derive the SDE schemes in some detail. In Section 4 we undertake numerical experiments and compare results from computational simulations of the underlying cell migration PDE model with simulations of two different SDE schemes. Finally in Section 5 concluding remarks are made.

2 Motivation

We are motivated in this work by typical continuum cancer invasion models (e.g. [7], [8], [9]) and consider the following general Advection-Diffusion PDE

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{A}u) - \nabla \cdot (\mathbf{D}\nabla u) = 0, \tag{1}$$

where, according to the usual practice, $u : [0, T] \times \Omega \to \mathbb{R}$ represents the space-time dependent density of cancer cells, and where $\Omega \subset \mathbb{R}^d$, with $d = 2, 3$, is a Lipschitz domain. We assume throughout this work that both the advection and diffusion coefficients $\mathbf{A}$ and $\mathbf{D}$ are non-constant in the sense that they depend on $\mathbf{x} \in \mathbb{R}^d$.

It is biologically understood that the macroscopic patterns of a large collective of cells is related to the migration of individual cells. Hence, our aim is to understand the connection between the continuous model (1), and models that capture the migration of solitary cancer cells. Following the ideas developed in the seminal works by Einstein [10] on the investigation of Brownian motion, as well as by Stratonovich, Ito and Kitanidis [11–13], the motion of the solitary cancer cells is usually described via SDEs that track the position of the cells. These SDEs typically take the form

$$d\mathbf{X}_t = \mathbf{a}(\mathbf{X}_t, t)\, dt + \mathbf{d}(\mathbf{X}_t, t)\, d\mathbf{W}_t\,, \tag{2}$$

with $t \geq 0$. Here $\mathbf{X}_t \in \mathbb{R}^d$ is a stochastic process that represents the position of the solitary cancer cells, and where $\mathbf{a}(\mathbf{X}_t, t)$ and $\mathbf{d}(\mathbf{X}_t, t)$ are the drift and the diffusion coefficients respectively. Here $\mathbf{W}_t$ represents a d-dimensional *Wiener process*.

The critical question that arises in this work is the following: how do the advection and diffusion terms $\mathbf{A}$, $\mathbf{D}$ of the deterministic PDE (1) relate to the drift and diffusion terms $\mathbf{a}$, $\mathbf{d}$ of the SDE (2)?

By identifying these relations we shed light in the complexity of multiscale modelling and simulations in general, and in the migration of solitary cancer cells and macroscopic tumour growth within a living organism in particular. Succeeding, hence, in identifying such relations will allow well established—and phenomenologically verifiable—macroscopic models to be used in order to extract, at the smaller scale, the migratory pattern of solitary cells or small clusters of cancer cells. Vice-versa, these relations will provide an additional validation to the use of deterministic models or even encourage ideas for a hybrid method using both deterministic and stochastic models.

We proceed in this effort under the assumption that u represents a single solitary cancer cell or a small cluster of cancer cells. We denote by $\mathbf{x}_t \in \mathbb{R}^d$ the numerical approximation of the solution stochastic process $\mathbf{X}_t \in \mathbb{R}^d$ that represents the position of the cell's centroid at time t. Following [11–15] we exhibit in Section 3 that the *less* intuitive numerical scheme:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \big(\mathbf{A}(\mathbf{x}_t) + \nabla \cdot \mathbf{D}(\mathbf{x}_t)\big)\,\Delta t + \sqrt{2\Delta t}\mathbf{V}(\mathbf{x}_t)\,\xi\ , \tag{3}$$

where $\boldsymbol{\xi}$ is a vector of d independent and normally distributed random variables of mean 0 and variance 1, and where $\mathbf{V}$ given by

$$\mathbf{V}(\mathbf{x}_t)\mathbf{V}^T(\mathbf{x}_t) = \mathbf{D}(\mathbf{x}_t)\ , \tag{4}$$

provides a better approximation to the PDE (1) than the *more* intuitive numerical scheme:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \mathbf{A}(\mathbf{x}_t)\,\Delta t + \sqrt{2\Delta t}\mathbf{V}(\mathbf{x}_t)\,\xi\ . \tag{5}$$

Both schemes account, in the same way, for the "square root" of the diffusion $\mathbf{D}$ in their corresponding diffusion/noise terms. As we will see in Section 3 this is derived in a very natural way.

The SDE (2) is of an Ito-type and, accordingly, the schemes (3) and (5) are derived using the Euler-Maruyama approximation. The less intuitive scheme (3) should not be mistaken for the equivalent scheme of a Stratonovich-type SDE emerging from (2), cf. [15]; it should rather be understood as an Ito-type scheme that is different from (5).

Structurally, the difference of the two schemes (3) and (5) lies in their corresponding drift terms. In particular, in the "corrected" scheme (3) the drift term accounts for both the advection, $\mathbf{A}$, and the diffusion, $\mathbf{D}$, coefficients of the PDE (1). In the scheme (5) the drift term accounts only for the advection term $\mathbf{A}$. This final remark, the one-to-one correspondence between the terms of the scheme (5) and the equation (2), justifies the characterisation of the scheme (5) as "more intuitive", cf. [15, 16].

In Section 3 we derive the scheme (3) from the PDE (1) and in Section 4 we compare it numerically with the scheme (5) and the PDE (1).

We, hence, conclude that there is not a one-to-one relation between the advection term of the PDE and the drift term of the SDE—unless the divergence of the diffusion $\mathbf{D}$ vanishes; this is clearly the case when the diffusion $\mathbf{D}$ does not depend on the spatial variable $\mathbf{x}$.

3 Derivation of the SDE Schemes

We make the fundamental modeling assumption that a cell can be viewed as a sufficiently small cell-cluster that satisfies the PDE (1). We furthermore assume that, without loss of generality, such a cell-cluster has unit mass i.e.

$$\int_{V_\infty} u(t,\mathbf{x})\, d\mathbf{x} = 1 \ , \tag{6}$$

where $u(t,\mathbf{x})$ is the density and V_∞ the volume of the cell-cluster.

To capture the behaviour of the cell-cluster, and in particular, the position of its centroid and spread, we use the method of moments. Accordingly, the first and second moments of u read respectively

$$\boldsymbol{\nu}_1(t) = \int_{V_\infty} \mathbf{x} u(t,\mathbf{x})\, d\mathbf{x} \ , \tag{7a}$$

$$\boldsymbol{\nu}_2(t) = \int_{V_\infty} \mathbf{x}\mathbf{x}^T u(t,\mathbf{x})\, d\mathbf{x} \ . \tag{7b}$$

Note that $\boldsymbol{\nu}_1(t)$ is a d-dimensional vector and $\boldsymbol{\nu}_2(t)$ a $d\times d$-dimensional matrix, and they are both understood via their physical rather than probabilistic interpretation. We also consider the second moment about the mean

$$\boldsymbol{N}(t) = \int_{V_\infty} (\mathbf{x} - \boldsymbol{\nu}_1(t))\,(\mathbf{x} - \boldsymbol{\nu}_1(t))^T\, u(t,\mathbf{x})\, d\mathbf{x} = \boldsymbol{\nu}_2(t) - \boldsymbol{\nu}_1(t)\boldsymbol{\nu}_1^T(t) \ , \tag{8}$$

where $\boldsymbol{N}(t)$ is $d \times d$-dimensional matrix. The first moment $\boldsymbol{\nu}_1$ in (7a) can be interpreted as the position of the cell-cluster centroid, and, accordingly, it's (time) rate of change as the velocity of the cell-cluster. On the other hand, the rate of change of the second moment $\boldsymbol{\nu}_2$ in (7b) represents the spreading of the cell-cluster. These two remarks together allow to (heuristically) identify the drift and diffusion coefficients, $\mathbf{a}$ and $\mathbf{d}$, of the SDE (2) as $\dfrac{d\boldsymbol{\nu}_1(t)}{dt}$ and $\dfrac{d\boldsymbol{N}(t)}{dt}$ respectively.

As a direct consequence of that, we will construct a numerical scheme in line with the classical Euler-Maruyama approximation, cf. [14], as follows

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \frac{d\boldsymbol{\nu}_1(t)}{dt}\Delta t + \sqrt{\frac{d\boldsymbol{N}(t)}{dt}\Delta t}\,\boldsymbol{\xi}, \tag{9}$$

where the square root in the above, should be understood in the usual matrix notation, see e.g. (4).

To this end, we impose the following boundary conditions that indicate an exponential decay of $u(t,\mathbf{x})$ as $|\mathbf{x}| \to \infty$.

$$\int_{S_\infty} e^{\mathbf{z}^T\mathbf{x}} u(t,\mathbf{x})\mathbf{A}^T\mathbf{n}\, dS = 0 \ , \tag{10a}$$

$$\int_{S_\infty} e^{\mathbf{z}^T\mathbf{x}} (\mathbf{D}\nabla u(t,\mathbf{x}))^T\mathbf{n}\, dS = 0 \ , \tag{10b}$$

$$\int_{S_\infty} e^{\mathbf{z}^T\mathbf{x}} u(t,\mathbf{x})\mathbf{D}\mathbf{n}\, dS = 0 \ , \tag{10c}$$

where S_∞ represents the surface of the cell-cluster V_∞ and $\mathbf{z}$ a d-dimensional vector. Through the i-th derivative of (10a)-(10c) with respect to $\mathbf{z}$ and setting $\mathbf{z} = 0$ one can retrieve the boundary conditions for the i-th moment. In our case we only need to differentiate twice since we only need the first two moments.

With all these in mind, we proceed by calculating through (7a) and (8) the time derivatives of $\boldsymbol{\nu}_1(t)$ and $\boldsymbol{N}(t)$. For $\boldsymbol{\nu}_1$ we multiply (1) by x_i, for $i = 1, ..., n$, integrate over the cell-cluster volume V_∞, and obtain

$$\int_{V_\infty} x_i \left[\frac{\partial u}{\partial t} + \nabla\cdot(\mathbf{A}u) - \nabla\cdot(\mathbf{D}\nabla u)\right] d\mathbf{x} = 0 \iff$$
$$\underbrace{\int_{V_\infty} x_i \frac{\partial u}{\partial t}\, d\mathbf{x}}_{=I_{11}} + \underbrace{\int_{V_\infty} x_i \nabla\cdot(\mathbf{A}u)\, d\mathbf{x}}_{=I_{12}} - \underbrace{\int_{V_\infty} x_i \nabla\cdot(\mathbf{D}\nabla u)\, d\mathbf{x}}_{=I_{13}} = 0 \ . \tag{11a}$$

We work on each of the terms I_{11}, I_{12}, I_{13} on the left hand side separately; and obtain for I_{11}

$$I_{11} = \int_{V_\infty} x_i \frac{\partial u}{\partial t}\, d\mathbf{x} = \frac{\partial}{\partial t}\int_{V_\infty} x_i u\, d\mathbf{x} = \frac{d(\boldsymbol{\nu}_1)_i}{dt}. \tag{11b}$$

For the second term I_{12}, after invoking the Divergence Theorem and the boundary conditions (10a)-(10c), we obtain

$$\begin{aligned} I_{12} = \int_{V_\infty} x_i \nabla\cdot(\mathbf{A}u)\, d\mathbf{x} &= \int_{V_\infty} \nabla\cdot(\mathbf{A}u x_i)\, d\mathbf{x} - \int_{V_\infty} A_i u\, d\mathbf{x} \\ &= \int_{S_\infty} u x_i \mathbf{A}^T\mathbf{n}\, dS - \int_{V_\infty} A_i u\, d\mathbf{x} \\ &\overset{(10a)}{=} -\int_{V_\infty} A_i u\, d\mathbf{x} \ , \end{aligned} \tag{11c}$$

where A_i is the i-th element of the vector $\mathbf{A}$. Similarly, the third term I_{13} recasts into

$$I_{13} = \int_{V_\infty} x_i \nabla\cdot(\mathbf{D}\nabla u)\, d\mathbf{x} = \int_{V_\infty} \nabla\cdot(x_i\mathbf{D}\nabla u)\, d\mathbf{x} - \int_{V_\infty} (\nabla u)^T\, \mathbf{D}_i\, d\mathbf{x}$$

$$
\begin{aligned}
&= \int_{S_\infty} x_i (\mathbf{D}\nabla u)^T \mathbf{n}\, dS - \int_{V_\infty} \nabla \cdot (u\mathbf{D}_i)\, d\mathbf{x} + \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x} \\
&\overset{(10b)}{=} - \int_{S_\infty} u \mathbf{D}_i^T \mathbf{n}\, dS + \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x} \\
&\overset{(10c)}{=} \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x}\ , \qquad (11d)
\end{aligned}
$$

where $\mathbf{D}_i$ is the i-th column of the diffusion matrix $\mathbf{D}$. So, by combining (11a) with (11b)-(11d) we obtain the following representation for the rate of change of the i-th (vector) component of the first moment:

$$
\frac{d(\boldsymbol{\nu}_1)_i}{dt} = \int_{V_\infty} (A_i + \nabla \cdot \mathbf{D}_i) u\, d\mathbf{x}. \qquad (12)
$$

In a similar way, we identify the relation satisfied by the ij^{th} element of the second moment $\boldsymbol{\nu}_2$, which allows us to compute the rate of change of $\boldsymbol{N}(t)$ in (8):

$$
\begin{aligned}
&\int_{V_\infty} x_i x_j \left[\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{A}u) - \nabla \cdot (\mathbf{D}\nabla u) \right] d\mathbf{x} = 0 \iff \\
&\underbrace{\int_{V_\infty} x_i x_j \frac{\partial u}{\partial t}\, d\mathbf{x}}_{=I_{21}} + \underbrace{\int_{V_\infty} x_i x_j \nabla \cdot (\mathbf{A}u)\, d\mathbf{x}}_{=I_{22}} - \underbrace{\int_{V_\infty} x_i x_j \nabla \cdot (\mathbf{D}\nabla u)\, d\mathbf{x}}_{=I_{23}} = 0. \qquad (13a)
\end{aligned}
$$

As with (11a), we employ the Divergence Theorem and the boundary condition (10a)-(10c) and calculate these three integrals one-by-one. The first one, I_{21}, reads

$$
I_{21} = \int_{V_\infty} x_i x_j \frac{\partial u}{\partial t}\, d\mathbf{x} = \frac{\partial}{\partial t} \int_{V_\infty} x_i x_j u\, d\mathbf{x} = \frac{d(\boldsymbol{\nu}_2)_{ij}}{dt}, \qquad (13b)
$$

the second integral, I_{21}, recasts into

$$
\begin{aligned}
I_{22} = \int_{V_\infty} x_i x_j \nabla \cdot (\mathbf{A}u)\, d\mathbf{x} &= \int_{V_\infty} \nabla \cdot (x_i x_j \mathbf{A}u)\, d\mathbf{x} - \int_{V_\infty} (x_j A_i + x_i A_j) u\, d\mathbf{x} \\
&= \int_{S_\infty} x_i x_j u \mathbf{A}^T \mathbf{n} dS - \int_{V_\infty} (x_j A_i + x_i A_j) u\, d\mathbf{x} \\
&\overset{(10a)}{=} - \int_{V_\infty} (x_j A_i + x_i A_j) u\, d\mathbf{x}, \qquad (13c)
\end{aligned}
$$

and for the third term, I_{21}, it holds

$$
\begin{aligned}
I_{23} &= \int_{V_\infty} x_i x_j \nabla \cdot (\mathbf{D}\nabla u)\, d\mathbf{x} \\
&= \int_{V_\infty} \nabla \cdot (x_i x_j \mathbf{D}\nabla u)\, d\mathbf{x} - \int_{V_\infty} \left[x_j (\mathbf{D}\nabla u)_i + x_i (\mathbf{D}\nabla u)_j \right] d\mathbf{x}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{S_\infty} x_i x_j (\mathbf{D}\nabla u)^T \mathbf{n}\, dS - \int_{V_\infty} \nabla \cdot (u x_j \mathbf{D}_i + u x_i \mathbf{D}_j)\, d\mathbf{x} \\
&\quad + \int_{V_\infty} \left[x_j \nabla \cdot \mathbf{D}_i + x_i \nabla \cdot \mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x} \\
&\overset{(10b)}{=} - \int_{S_\infty} (u x_j \mathbf{D}_i + u x_i \mathbf{D}_j)^T \mathbf{n}\, d\mathbf{x} \\
&\quad + \int_{V_\infty} \left[x_j \nabla \cdot \mathbf{D}_i + x_i \nabla \cdot \mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x} \\
&\overset{(10c)}{=} \int_{V_\infty} \left[x_j \nabla \cdot \mathbf{D}_i + x_i \nabla \cdot \mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x}, \qquad (13d)
\end{aligned}
$$

where D_{ij} is the element in the i-th row and j-th column of the matrix $\mathbf{D}$. By combining (13a) with (13b)-(13d), we obtain the following relation for ij-th (matrix) element of the second moment

$$
\frac{d(\boldsymbol{\nu}_2)_{ij}}{dt} = \int_{V_\infty} (x_j A_i + x_i A_j) u\, d\mathbf{x} + \int_{V_\infty} \left[x_j \nabla \cdot \mathbf{D}_i + x_i \nabla \cdot \mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x}. \quad (14)
$$

Summing up (12) and (14), we deduce the following set of equations in a vector form for any given advection and diffusion terms $\mathbf{A}$ and $\mathbf{D}$ in (1)

$$
\frac{d\boldsymbol{\nu}_1}{dt} = \int_{V_\infty} (\mathbf{A} + \nabla \cdot \mathbf{D})\, u(t, \mathbf{x})\, d\mathbf{x} \qquad (15)
$$

$$
\frac{d\boldsymbol{\nu}_2}{dt} = \int_{V_\infty} \left(\mathbf{x}(\mathbf{A} + \nabla \cdot \mathbf{D})^T + (\mathbf{A} + \nabla \cdot \mathbf{D})\mathbf{x}^T + 2\mathbf{D}\right) u(t, \mathbf{x})\, d\mathbf{x} \qquad (16)
$$

$$
\begin{aligned}
\frac{d\boldsymbol{N}}{dt} &= \int_{V_\infty} \left(\mathbf{x}(\mathbf{A} + \nabla \cdot \mathbf{D})^T + (\mathbf{A} + \nabla \cdot \mathbf{D})\mathbf{x}^T + 2\mathbf{D}\right) u(t, \mathbf{x})\, d\mathbf{x} \\
&\quad - \boldsymbol{\nu}_1(t) \left(\int_{V_\infty} (\mathbf{A} + \nabla \cdot \mathbf{D}) u(t, \mathbf{x})\, d\mathbf{x} \right)^T \\
&\quad - \left(\int_{V_\infty} (\mathbf{A} + \nabla \cdot \mathbf{D}) u(t, \mathbf{x})\, d\mathbf{x} \right) \boldsymbol{\nu}_1^T(t). \qquad (17)
\end{aligned}
$$

We retrieve (17) by taking the time derivative of (8) and substituting equations (15) and (16).

Assuming that the mass of the cell-cluster is concentrated, at time t, in a single point $\mathbf{x}_t$, we can represent the density of the cell-cluster as

$$
u(t, \mathbf{x}) = \delta\left(\mathbf{x} - \mathbf{x}_t\right),
$$

where δ is the Dirac function centered at $\mathbf{x}_t$. Then the velocity of the cell-cluster in (15), is given as the sum of the advection $\mathbf{A}$ and the divergence of

Table 1 Parameters used in **Experiments** 1 and 2.

-	t_0	T	N	a	b	$\mathbf{x_0}$	realisations
Experiment 1	0	1	1000	0.01	0.001	$(1,1)^T$	10^2
Experiment 2	0	1	1000	0.01	0.001	$(1,1)^T$	10^5

the diffusion matrix $\mathbf{D}$ at $\mathbf{x}_t$:

$$\frac{d\boldsymbol{\nu}_1}{dt} = \mathbf{A}(x_t) + \nabla \cdot \mathbf{D}|_{\mathbf{x}_t}. \tag{18}$$

Accordingly, the rate of change of second moment about the mean in (17) recasts into

$$\begin{aligned}\frac{d\boldsymbol{N}}{dt} &= \mathbf{x}_t(\mathbf{A} + \nabla \cdot \mathbf{D})^T\Big|_{\mathbf{x}_t} + (\mathbf{A} + \nabla \cdot \mathbf{D})\Big|_{\mathbf{x}_t} \mathbf{x}_t^T + 2\mathbf{D}\Big|_{\mathbf{x}_t} \\ &\quad - \mathbf{x}_t(\mathbf{A} + \nabla \cdot \mathbf{D})^T\Big|_{\mathbf{x}_t} - (\mathbf{A} + \nabla \cdot \mathbf{D})\Big|_{\mathbf{x}_t} \mathbf{x}_t^T \\ &= 2\mathbf{D}\Big|_{\mathbf{x}_t},\end{aligned} \tag{19}$$

which clearly indicates that the rate of spreading of the cell-cluster is given by the diffusion matrix at the point $\mathbf{x}_t$.

We close this section by substituting (18) and (19) into (9) to obtain the corrected numerical scheme

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \big(\mathbf{A}(\mathbf{x}_t) + \nabla \cdot \mathbf{D}(\mathbf{x}_t)\big)\Delta t + \sqrt{2\Delta t}\mathbf{V}(x_t)\boldsymbol{\xi}_2, \tag{20}$$

where $\mathbf{V}(\mathbf{x}_t)\mathbf{V}^T(\mathbf{x}_t) = \mathbf{D}(\mathbf{x}_t)$ and where $\boldsymbol{\xi}_2$ is a vector of d independent and normally distributed random variables with 0 mean and variance 1.

The scheme (20) that we have just derived, is different from the one without the correction of the drift term, i.e. (5), which we repeat here for completion:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \mathbf{A}(\mathbf{x}_t)\Delta t + \sqrt{2\Delta t}\mathbf{V}(x_t)\boldsymbol{\xi}_1, \tag{21}$$

where, as before, $\boldsymbol{\xi}_1$ is a vector of independent and normally distributed random variables with mean 0 and variance 1.

In the next section we numerically investigate the two schemes (21) and (20) and provide evidence of their differences and their fitting with the corresponding/underlying PDE (1).

4 Numerical Experiments

We have seen in the previous section that the drift term of the SDE scheme (20) incorporates a diffusion-based contribution/correction that is not found

in the more intuitive scheme (21). This is a significant difference between the two schemes and is central in our numerical investigations. Namely our aim in this section is to numerically investigate the impact that this correction has on the simulations of these two schemes. To this end, we consider three specific numerical settings that highlight the difference of the two schemes, and compare their predictions with each other and with the corresponding/underlying PDE (1).

In more detail, in Experiment 1, we investigate the behaviour of the two schemes, (21) and (20), on a particular application where the corresponding noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ are the same. This allows to csompare the two schemes as a result of their differences on the drift terms alone.

In Experiment 2, we consider the same computational setting as in Experiment 1, with the only difference being that the two SDE schemes (21), (20) are augmented with different noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$. With a large number of scheme realisations we extract information on the full spectrum of differences between the two schemes.

In Experiment 3 we consider a more generic, and common in the cancer invasion modelling literature, experimental setting where the advection and diffusion terms $\mathbf{A}$, $\mathbf{D}$ of (1) depend on the spatial variable $\mathbf{x} \in \mathbb{R}^d$ through their dependence on the non-uniform tumour microenvironment $v : \mathbb{R}^d \to \mathbb{R}$. For both SDE schemes (21) and (20), we perform a large number of realisations and compare them with the solution of the underlying PDE (1). We accordingly conclude that the corrected SDE scheme (20) provides a much better approximation to the PDE (1) than the more intuitive SDE scheme (21).

For the numerical solution of the PDE (1) we use a numerical method that was previously developed in [17, 18] and which we briefly discuss in the Appendix A. All algorithm implementations, simulations, and visualisations were conducted in MATLAB [19].

Experiment 1. In the first experiment, we consider a modelling setting where the advection and diffusion terms, $\mathbf{A}$ and $\mathbf{D}$, of the PDE (1) have an explicit dependence on the space variable $\mathbf{x}$, namely:

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix}, \qquad \mathbf{D}(\mathbf{x}) = \begin{pmatrix} bx_1^7 & 0 \\ 0 & bx_2^7 \end{pmatrix}, \tag{22}$$

where $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and $a, b \in \mathbb{R}$ are constants. Accordingly, the PDE (1) reads as

$$\frac{\partial u}{\partial t} + \nabla \cdot \left[\begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} u \right] - \nabla \cdot \left[\begin{pmatrix} bx_1^7 & 0 \\ 0 & bx_2^7 \end{pmatrix} \nabla u \right] = 0, \tag{23}$$

where $t \in [0, T]$. Similarly, the SDE schemes (21) and (20) are re-formulated, for

$$\mathbf{V}(\mathbf{x_t}) = \begin{pmatrix} \sqrt{bx_1^7} & 0 \\ 0 & \sqrt{bx_2^7} \end{pmatrix}, \tag{24}$$

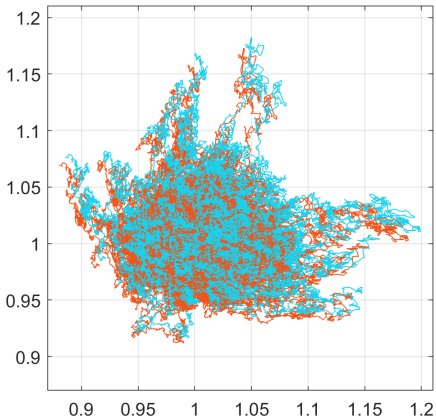


Fig. 1 – Exp. 1. Showing here the cell migration tracks/sample paths of 100 realisations of the SDE schemes (25), in orange, and (26), in light blue, with the same random noise. The initial position of the cells is set at (1,1) for all realisations and the total travelling time is $T = 1$. Note that the sample paths of the two schemes are parallel shifts of one another. The shift between the two schemes is due to the adaptation of the drift term of the scheme (26).

as follows:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} \Delta t + \sqrt{2\Delta t} \begin{pmatrix} \sqrt{bx_1^7} & 0 \\ 0 & \sqrt{bx_2^7} \end{pmatrix} \boldsymbol{\xi}_1, \tag{25}$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \left[\begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} + \nabla \cdot \begin{pmatrix} bx_1^7 & 0 \\ 0 & bx_2^7 \end{pmatrix} \right] \Delta t + \sqrt{2\Delta t} \begin{pmatrix} \sqrt{bx_1^7} & 0 \\ 0 & \sqrt{bx_2^7} \end{pmatrix} \boldsymbol{\xi}_2. \tag{26}$$

These two SDE schemes are augmented with the same noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ to allow for a direct, one-to-one, comparison of their realisations.

The parameters used in this numerical experiment can be found in Table 1 and the simulation results are presented in Figure 1. In more detail, we perform 100 realisations of the two schemes (25), (26)—each pair of experiments with the same noise—starting from the same initial position $(x_1, x_2) = (1, 1)$ and running over the time $t \in [0, 1]$. These results show that the realisations of the two schemes appear to be—in a one-to-one conformation—slightly shifted and parallel to each other. More specifically, as can be seen, the corrected scheme (26) introduces an additional displacement of the cell migration track/sample paths directed towards larger values of x_1 and x_2. This is due to the nature and structure of the drift term as well as the diffusion-based correction introduced in (26) and the positivity of the parameter b, cf. Table 1.

Experiment 2. The intuition we have gained from Experiment 1, namely the way the adaptation of the drift term in the corrected scheme (26) affects the distance and direction of the cell migration, can be investigated further by considering independent noise terms for the two schemes. Hence, in the

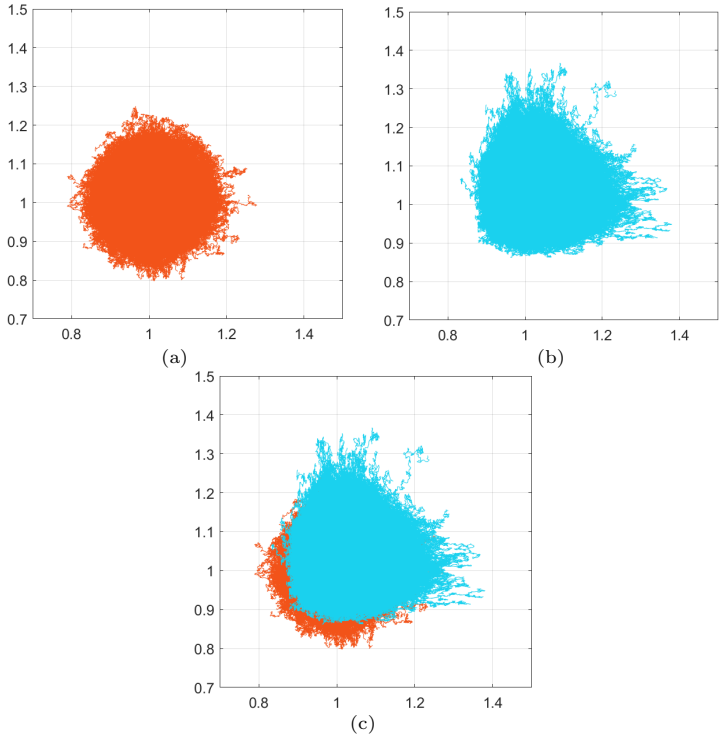


Fig. 2 – **Exp. 2.** Comparison of the migration tracks traversed the cells following the SDE schemes (25) and (26). For both schemes, we considered initial position at $(1, 1)$ and a total travelling time $T = 1$. (a) & (b) The full migration tracks of 10^5 cells following the SDE scheme (25) and (26) respectively. (c) Superimposing the migration patterns of (a) and (b) clearly reveals that the corrected scheme (26) introduces additional migratory bias.

current experiment, we choose independent noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ for the two SDE schemes (25) and (26) and perform 10^5 new realisations with each, all start from the initial point $(x_1, x_2) = (1, 1)$ and running over time $t \in [0, 1]$. We otherwise consider the same setting and parameters as in Experiment 1; the parameters for this experiment can be found in in Table 1.

The simulation results are presented in Figures 2 and 3. In Figure 2, in particular, we see that cells migrate further away from the origin and in a more biased fashion when they follow the corrected scheme (26) rather than the scheme (25). This is clearly the result of the additional bias introduced in the drift term of the corrected scheme (26). The qualitative difference between the two schemes can be further seen in Figure 3 where we present the final positions of the cells at time T of both schemes, along with their corresponding convex hulls.

To measure the quantitative difference of the two schemes (25) and (26), we first measure the average distance traversed by the cells from their initial position until the final time T. In more detail, we perform K realisations of the schemes and calculate, for each one, the distance between the initial and the

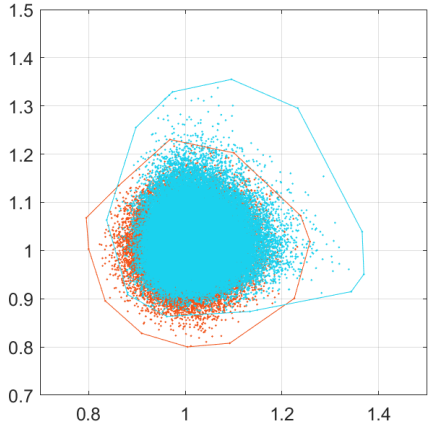


Fig. 3 – **Exp. 2.** Showing here the final positions of the cells migration after 10^5 realisations of the SDE schemes (25) and (26)—in orange and light blue respectively—as shown in Figure 2. All realisations start from $(1, 1)$ and run for time $t \in [0, 1]$. The corresponding convex hulls of these final positions are also shown. Note that the adaptation in the drift term of the corrected scheme (26), as opposed to the scheme (25), induces additional migration of the cells and spread of their final positions.

final position. We then calculate the average distance the cells have traversed by the formula

$$d = \frac{1}{K} \sum_{i=1}^{K} \sqrt{(x^i_{1,T} - x^i_{1,t_0})^2 + (x^i_{2,T} - x^i_{2,t_0})^2}, \tag{27}$$

where $\left(x^i_{1,t_0}, x^i_{2,t_0}\right)$ and $\left(x^i_{1,T}, x^i_{2,T}\right)$ represent the initial and final positions of the cells in the realisation $i = 1, ..., K$. We apply the above formula for the two schemes (25) and (26), after performing $K = 10^6$ realisations of each, calculate their respective average distances d_1 and d_2 respectively, and evaluate the signed relative difference between the two to obtain

$$E = \frac{d_2 - d_1}{d_1} \approx 0.076\,. \tag{28}$$

To further quantify the difference between the two schemes, we perform a uniform binning approach of the final positions of the cancer cells as described in [20]. Namely, we consider a partition of the x_1-axis into non-overlapping bins of fixed size r, and assign to each bin the quantity:

$$U_i = \sum_{j=1}^{N_i} I_{i,j}\,, \tag{29}$$

where N_i is the total number of positions (x_1, x_2) in the i-th bin and $I_{i,j}$ is the j-th element of the set $I_i = \{x_2 : \text{for positions } (x_1, x_2) \text{ in the } i\text{-th bin}\}$.

The quantification was made by computing the J_2 criterion which is defined as follows

$$J_2 = \frac{|S_w + S_b|}{S_w}, \tag{30}$$

where

$$\begin{aligned} S_w &= \sum_{i=1}^{M} P_i S_i, \qquad S_b = \sum_{i=1}^{M} P_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^T, \\ P_i &= \frac{n_i}{K}, \qquad S_i = \mathbb{E}\big[(\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)^T\big], \qquad \boldsymbol{\mu}_0 = \sum_{i=1}^{M} P_i \boldsymbol{\mu}_i \end{aligned} \tag{31}$$

where n_i is the number of positions in the i-th bin, K is the number of realisations, and $\boldsymbol{\mu}_i$ is the mean value of i-th bin. After computing the J_2 values of both schemes (25) and (26), we compute the signed relative error for different values r of the size of the bins

$$\mathcal{E}^r = \frac{J_2^{2,r} - J_2^{1,r}}{J_2^{1,r}}, \tag{32}$$

where $J_2^{1,r}, J_2^{2,r}$ are the values of the J_2-criterion for schemes (25) and (26) respectively. Larger values of J_2 correspond to better separated data. For the choice of $r = 0.01 + kh$, where $h = 0.001$, for $k = 1, ..., 60$, we observe that for smaller sizes of the bins we get larger values of J_2 and that $J_2^{1,r} < J_2^{2,r}$ for all the values of r. The average value of the $k-$different values of $\mathcal{E}^r$ is

$$\mathcal{E}^r \approx 0.0082. \tag{33}$$

This result provides a second confirmation that (26) introduces additional migratory bias to the one side of the plane.

Experiment 3. In this experiment we consider a PDE that is more common in the field of cancer invasion than the PDE in equation (23). Namely, we consider here a model where the growth of the tumour depends on the extracellular environment. This could represent, e.g. an extracellular chemical signal, the density of the extracellular matrix, or a completely different extracellular biochemical queue. Still, for the sake of simplicity of presentation, we do not make any particular biological assumptions on the nature of the extracellular environment and rather refer to it simply as "environment".

Furthermore, we assume that this environment is non-uniform in space, does not change in time, and influences the growth of tumour in a very specific fashion. These assumptions are incorporated in the following model:

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) - \nabla \cdot \Big(\nabla v(\mathbf{x}) u(t, \mathbf{x})\Big) - \nabla \cdot \Big(\big(v(\mathbf{x}) - 1\big) \nabla u(t, \mathbf{x})\Big) = 0, \tag{34}$$

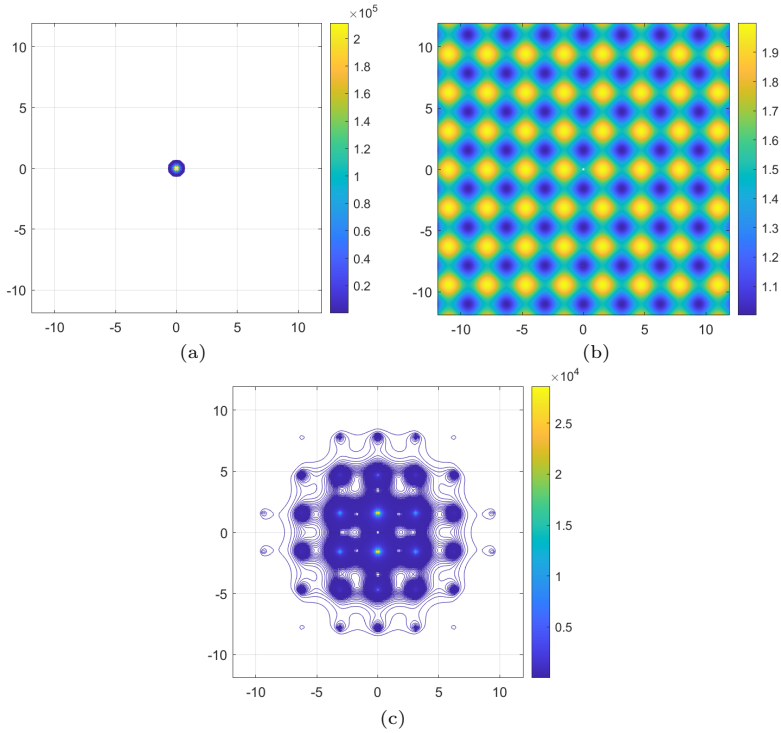


Fig. 4 – **Exp. 3.** Numerical solution of the PDE (34). (a) Showing here the isolines of the concentration u at time $t = 0$; they serve as initial conditions for the PDE (34). (b) The structure of the (fixed) extracellular environment v that participates in the advection and diffusion terms of (34); the formula of v is given in (36). (c) Isolines of the solution u of the PDE (34) at the final time $t = 10$; they reveal a higher concentration of the cancer cells in the "valleys" of the (fixed) extracellular environment v shown in (b).

where $t \geq 0$, $\mathbf{x} = (x_1, x_2) \in \Omega = [-12, 12]^2$, and where $u : [0, \infty) \times \Omega \to \mathbb{R}$ represents the density of the cancer cells. As previously mentioned, we do not investigate the biomedical realism of this model, nor do we interpret its findings under this light.

The PDE (34) is augmented with the initial condition

$$u(0, \mathbf{x}) = u(0, x_1, x_2) = 3 \times 10^5 \, e^{-20(x_1^2 + x_2^2)}, \tag{35}$$

for $\mathbf{x} \in \Omega$ as shown in Figure 4, and zero-Neumann boundary conditions over $\partial\Omega$. Furthermore, the (fixed) extracellular environment $v : \Omega \to \mathbb{R}$, shown in Figure 4, is given by

$$v(\mathbf{x}) = v(x_1, x_2) = 1 + 0.5 \left(\sin^2(x_1) + \cos^2(x_2) \right). \tag{36}$$

Note that, for v given in (36), the diffusivity $v(\mathbf{x}) - 1$ of (34) is non-negative for all $\mathbf{x} \in \Omega$.

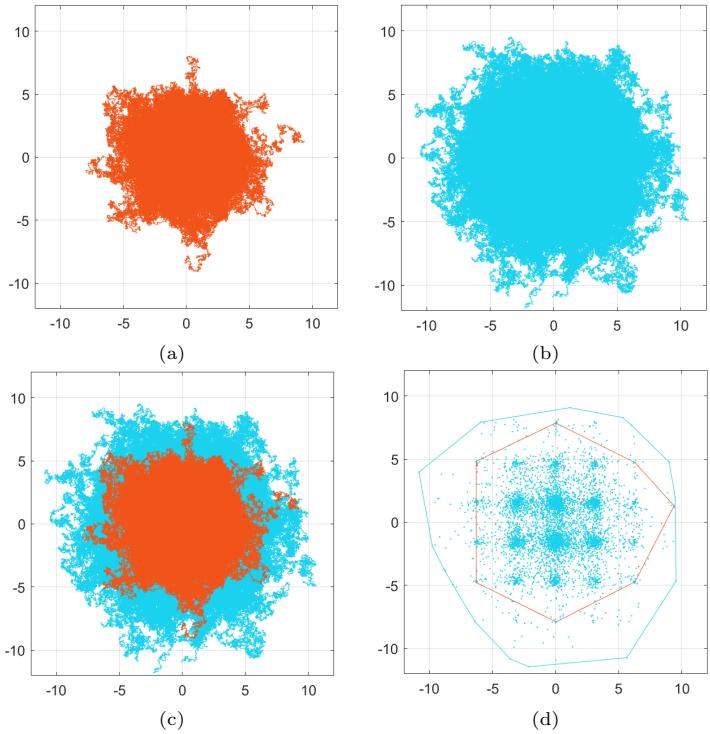


Fig. 5 – Exp. 3. Simulations and comparison of 10^4 realisations of SDE schemes (37a) and (37b). (a) & (b) Full set of tracks/sample paths for the more intuitive scheme (37a) and the corrected scheme (37b) respectively. It can be clearly seen that the spread in (37b) is much wider than in (37a); this is justified by the adaptation introduced in the drift term of the corrected scheme (26). This remark is confirmed by superimposing the cell tracks of (a) in (b) in (c). (d) This shows the final positions of the 10^4 realisations of (37a) and (37b), in orange and light blue respectively, along with the convex hulls of the corresponding points. It is clearly seen that the cells concentrate in the "valleys" of the environment v, cf. with the solution of the PDE (34) in Figure 4, and that the final positions of (37b) (shown in light blue) spread more than those of (37a). The final positions of (37a) are not visible as they are overlapped by the final positions of (37b).

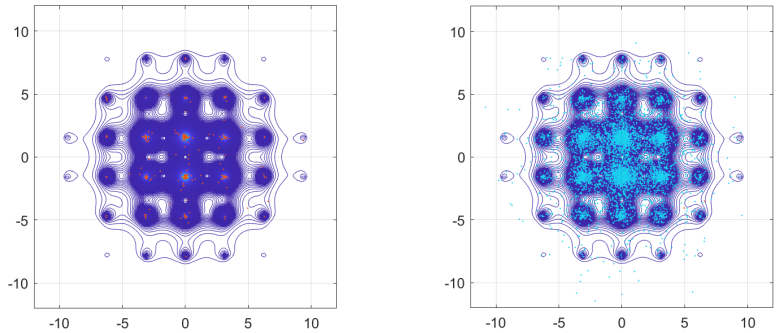
We plot in Figure 4 the isolines of the solution of the PDE (34) at the final time $t = 10$. These illustrate clearly a significantly higher concentration of the cancer cells in the "valleys" of the extracellular environment v.

Based on the advection and diffusion terms of the PDE (34), we rewrite the SDE schemes (21) and (20) as follows:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t - \nabla v(\mathbf{x}_t)\Delta t + \sqrt{2\Delta t}\sqrt{v(\mathbf{x}_t) - 1}\boldsymbol{\xi}_1, \tag{37a}$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \sqrt{2\Delta t}\sqrt{v(\mathbf{x}_t) - 1}\boldsymbol{\xi}_2. \tag{37b}$$

Note that the corrected SDE scheme (37b), which corresponds to the corrected scheme (20), lacks a drift term; this is a result of the particular structure of the



(a) (b)

Fig. 6 – **Exp. 3.** Qualitative comparison of the numerical solution of the PDE (34) against the predictions of the SDE schemes (37a) and (37b). (a) Isolines of the final time solution u of the PDE (34), cf. Figure 4, superimposed with 10^4 (final time) realisations of the SDE scheme (37a). (b) Isolines of the final time solution u of the PDE (34), cf. Figure 4, superimposed with 10^4 (final time) realisations of the SDE scheme (37b). We note that the cells described by the SDE scheme (37a) are concentrated, at the final time, almost exclusively in the "valleys" of the (fixed) environment v, cf. Figure 4, much more than cells described by the corrected SDE scheme (37b), and much more than the final time solution u of the PDE (34). The results confirm that the corrected scheme (37b) offers a much better approximation of the PDE (34) than the more more intuitive scheme (37a).

advection and diffusion terms of the PDE (34) and the way they are combined in the drift term of the corrected scheme (20).

The simulations that we perform for Experiment 3 are shown in Figure 5 and are similar to the ones for Experiment 2 (cf. Figures 2 and 3). We note that the cell tracks of the more intuitive scheme (37a) spread out from the origin in a lesser extend than the corrected scheme (37b). This is due to the presence of the drift term in the scheme (37a) which leads almost all cancer cells to high densities of the extracellular environment v. On the other hand, the final positions of the sample paths given by (37b) concentrate less in the "valleys" of the extracellular environment.

In Figure 6 we present a direct qualitative comparison between the numerical solution of the PDE (34) and the SDE schemes (37a) and (37b). What this figure shows is that, at the final time $t = 10$, the positions of the cells of 10^4 realisations of the SDE scheme (37a) are much more concentrated than the corresponding solution of the PDE (34). On the other hand, the distribution at the same final time $t = 10$, of 10^4 realisations of the corrected scheme (37b) is more spread out and much closer to the solution of the PDE (34). In effect, these simulation results indicate that the corrected scheme (37b) is a better approximation to the numerical solution of the PDE (34) than the scheme (37a).

Similarly to Experiment 2, we quantify the difference between the two schemes by measuring the corresponding average distances traversed by the cells from their initial position. In more detail, we perform $K = 10^4$ realisations for each scheme, calculate the average distances d_1 and d_2 using the formula

(27), and their relative difference via:

$$E = \frac{d_2 - d_1}{d_1} \approx 0.516\,. \tag{38}$$

To further quantify the difference of the two schemes, we perform a uniform binning and calculate the relative error of the J_2-criterion for (37a) and (37b) for $k-$different values of the size r of the bins as in (32). We noticed that, for smaller values r we have that again $J_2^{1,r} < J_2^{2,r}$. We choose $r = 0.01 + kh$, where $h = 0.001$, for $k = 1, ..., 140$, and calculate the average value of the $k-$different values of the signed relative error $\mathcal{E}^r$ which is

$$\mathcal{E}^r \approx 3.0692\,. \tag{39}$$

This confirms, yet again, that (37b) brings an additional migratory bias to the movement of solitary cancer cells to the entire plain.

5 Discussion

In this paper we have investigated the long-standing question of bridging the scales in problems of multiscale modelling and simulations. As we are motivated by the study of cancer growth and invasion models, we can rephrase this question as follows: how does one identify the correct SDEs (2) that dictate the migratory pattern of solitary cancer cells or small clusters of cancer cell, when the macroscopic evolution of the cancer cell colony follows the advection-diffusion PDE (1)?

We have exhibited in this paper, that the answer to this question is not trivial. The usual heuristic understanding that the advection and diffusion terms of the PDE (1) are responsible for the biased and random motion of the cancer cells, respectively, is not precise. Were this correct, the drift term in the SDE schemes that describe the migration of the cells in the solitary cell regime would have been solely dependent on the advection term $\mathbf{A}$ of the PDE (1), as e.g. is the case in the numerical scheme (5). We have seen with the derivations of the SDE schemes in Section 3 and with the numerical investigations in Section 4 that this is not the case. On the contrary, we have shown with (3) that the drift term of the correct SDE schemes should account for the advection $\mathbf{A}$ as well as the divergence of the diffusion $\mathbf{D}$ of the PDE (1) in a very specific way.

When comparing the corrected scheme (3) with the more heuristically expected one (5), we see that their difference depends solely on the divergence of the diffusion $\mathbf{D}$. This clearly indicates that the two schemes would be/are identical in the case of space independent diffusion $\mathbf{D}$. If though, both the advection and diffusion terms depend on the spatial variable—as is typically the case in cancer invasion models—then, identifying the inconsistency of the SDE schemes with the underlying PDE is not trivial and quite often is not readily apparent due to the inherent stochasticity.

From a biological/oncological perspective this is important in determining accurately how far individual cancer cells can penetrate the local tissue, and so from a clinical point of view the model results have much predictive potential. It is known that resection margins (i.e. how much tissue is removed surgically) and the pattern of cancer invasion are predictors both of local recurrence and of survival in patients who undergo surgery [21], [22], [23], [24], [25], [26]. With further refinement, accurate parameterisation and testing, using the results from the model would enable quantitative estimates to be made of the likely extent of local spread by an invasive cancer. This would then enable a surgical oncologist to tailor the radicality of surgical excision for a given individual situation. Further, more accurate estimation of metastatic spread (with the associated implications for adjuvant systemic therapy) will also be possible.

Future work in further developing the insights gained from the current modelling will focus on parameterising the model more accurately in order to make quantitative comparison with any available data. A promising avenue of development in this regard, given the difficulty in obtaining *in vivo* data, is to use *in vitro* data from organotypic invasion assays cf. [27], [28], [29].

Statements and Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Supplementary Material/Appendix

A Numerical solution of Advection-Reaction-Diffusion PDEs

We present here the main components of the numerical method we use to solve the generic *Advection-Reaction-Diffusion* (ARD) system of the form

$$\mathbf{w}_t = A(\mathbf{w}) + R(\mathbf{w}) + D(\mathbf{w}), \tag{40}$$

where $\mathbf{w} = \mathbf{w}(t, \mathbf{x})$, $t \geq 0$, $\mathbf{x} \in \Omega$ (domain), represents the solution vector, and A, R, and D the advection, reaction, and diffusion operators respectively.

We denote by $\mathbf{w}_h(t)$ the corresponding (semi-)discrete numerical approximation—indexed here by the maximal spatial grid diameter h—that satisfies the system of ODEs

$$\partial_t \mathbf{w}_h = \mathcal{A}(\mathbf{w}_h) + \mathcal{R}(\mathbf{w}_h) + \mathcal{D}(\mathbf{w}_h), \tag{41}$$

where the numerical operators $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{D}$ are *discrete approximations* of the operators A, R, and D in (40) respectively.

We use a second order *Implicit-Explicit Runge-Kutta* (IMEX-RK) *Finite Volume* (FV) numerical method that was previously developed in [17, 18]

where we refer for more details, see also [30]. This method is based on a *splitting* in explicit and implicit terms in the form

$$\partial_t \mathbf{w}_h = \mathcal{I}(\mathbf{w}_h) + \mathcal{E}(\mathbf{w}_h). \tag{42}$$

The actual splitting depends on the particular problem in hand but in a typical case, the advection terms $\mathcal{A}$ are treated explicitly in time, the diffusion terms $\mathcal{D}$ implicitly, and the reaction terms $\mathcal{R}$ partly explicit and partly implicit.

After splitting, we employ a diagonally implicit RK method for the implicit part, and an explicit RK for the explicit part. They are combined in the following scheme

$$\begin{cases} \mathbf{W}_i^* = \mathbf{w}_h^n + \tau_n \sum_{j=1}^{i-2} \bar{a}_{i,j}\mathbf{E}_j + \tau_n \bar{a}_{i,i-1}\mathbf{E}_{i-1}, & i = 1\dots s \\ \mathbf{W}_i = \mathbf{W}_i^* + \tau_n \sum_{j=1}^{i-1} a_{i,j}\mathbf{I}_j + \tau_n a_{i,i}\mathbf{I}_i, & i = 1\dots s \ , \\ \mathbf{w}_h^{n+1} = \mathbf{w}_h^n + \tau_n \sum_{i=1}^{s} \bar{b}_i\mathbf{E}_i + \tau_n \sum_{i=1}^{s} b_i\mathbf{I}_i \end{cases} \tag{43}$$

where $s = 4$ are the stages of the IMEX method, $\mathbf{E}_i = \mathcal{E}(\mathbf{W}_i)$, $I_i = \mathcal{I}(\mathbf{W}_i)$, $i = 1\dots s$, $\{\bar{b}, \bar{A}\}$, $\{b, A\}$ are respectively the coefficients for the explicit and the implicit part of the scheme, given by the Butcher Tableau in Table 2, cf. [31].

We solve the linear systems in (43) using the *iterative biconjugate gradient stabilised Krylov subspace* method, for which we refer to [32, 33].

Table 2 Butcher tableaux for the explicit (upper) and the implicit (lower) parts of the third order IMEX scheme (43), see also [31].

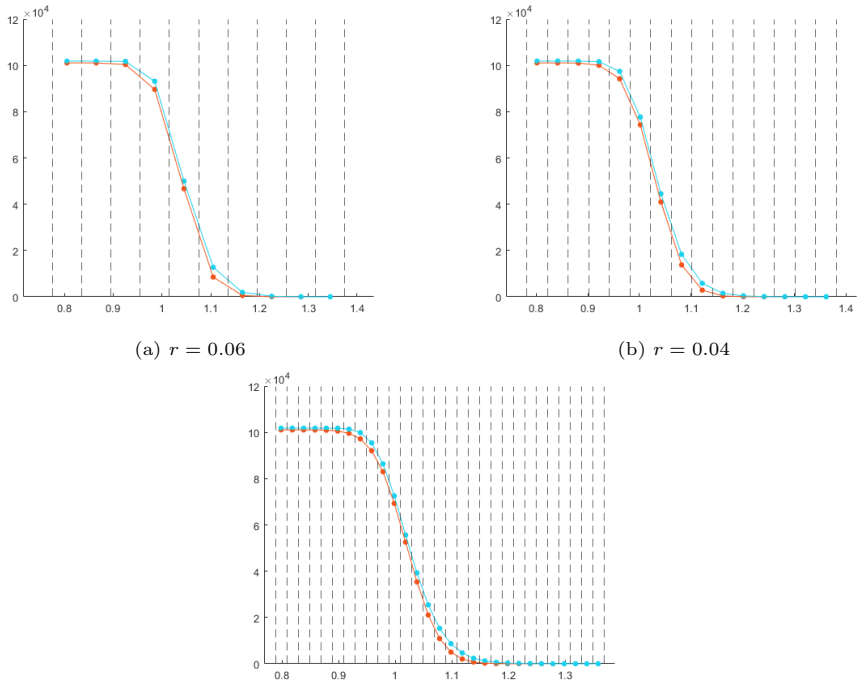
0				
$\frac{1767732205903}{2027836641118}$	$\frac{1767732205903}{2027836641118}$			
$\frac{3}{5}$	$\frac{5535828885825}{10492691773637}$	$\frac{788022342437}{10882634858940}$		
1	$\frac{6485989280629}{16251701735622}$	$-\frac{4246266847089}{9704473918619}$	$\frac{10755448449292}{10357097424841}$	
	$\frac{1471266399579}{7840856788654}$	$-\frac{4482444167858}{7529755066697}$	$\frac{11266239266428}{11593286722821}$	$\frac{1767732205903}{4055673282236}$
0	0			
$\frac{1767732205903}{2027836641118}$	$\frac{1767732205903}{4055673282236}$	$\frac{1767732205903}{4055673282236}$		
$\frac{3}{5}$	$\frac{2746238789719}{10658868560708}$	$-\frac{640167445237}{6845629431997}$	$\frac{1767732205903}{4055673282236}$	
1	$\frac{1471266399579}{7840856788654}$	$-\frac{4482444167858}{7529755066697}$	$\frac{11266239266428}{11593286722821}$	$\frac{1767732205903}{4055673282236}$
	$\frac{1471266399579}{7840856788654}$	$-\frac{4482444167858}{7529755066697}$	$\frac{11266239266428}{11593286722821}$	$\frac{1767732205903}{4055673282236}$

Table 3 Bin size r and J_2 values in **Exp.** 2.

r	0.01	0.02	0.03	0.04	0.05	0.06	0.07
J_2 of (25) $\approx$	6.90562	1.67652	1.18584	1.07234	1.03471	1.01896	1.01138
J_2 of (26) $\approx$	7.24178	1.70981	1.19329	1.07589	1.03633	1.01929	1.01199

B Uniform Binning

In this section we present some graphic results of the uniform binning performed for Experiment 2 in Section 4. We illustrate the binning of the final positions of 10^5 realisations of both schemes (25) and (26). At the centre of each bin we place the value U_i calculated in (29) for both schemes. We perform this sequence with different sizes r of the bins. In addition we present in Table 3 the J_2 values of the both schemes (25) and (26) for $r = \{0.02, 0.04, 0.06\}$.



(a) $r = 0.06$

(b) $r = 0.04$

(c) $r = 0.02$

Fig. 7 – **Exp. 2.** Graphical comparison of uniform binning of the final positions of the migration tracks traversed the cells following the SDE schemes (25) and (26) respectively. For both schemes, we considered initial position at $(1, 1)$ and a total travelling time $T = 1$. (a) Uniform binning of the final positions of migration tracks of 10^5 cells for the SDE schemes (25) and (25) with bin size $r = 0.06$. (b) Uniform binning of the final positions of migration tracks of 10^5 cells for the SDE schemes (25) and (25) with bin size $r = 0.04$. (c) Uniform binning of the final positions of migration tracks of 10^5 cells for the SDE schemes (25) and (25) with bin size $r = 0.02$.

We also worked in a similar fashion for the schemes (37a) and (37b) in the Experiment 3.

References

[1] Talmadge, J.E., Fidler, I.J.: AACR centennial series: The biology of cancer metastasis: historical perspective. Cancer Res. **70**, 5649–5669 (2010)

[2] Hanahan, D., Weinberg, R.A.: The hallmarks of cancer. Cell **100**, 57–70 (2000)

[3] Hanahan, D., Weinberg, R.A.: The hallmarks of cancer: The next generation. Cell **144**, 646–671 (2011)

[4] Friedl, P., Wolf, K.: Tumour-cell invasion and migration: diversity and escape mechanisms. Nat. Rev. Cancer **5**, 362–374 (2003)

[5] Sfakianakis, N., Chaplain, M.A.J.: Mathematical modelling of cancer invasion: A review. In: Suzuki, T., Poignard, C., Chaplain, M., Quaranta, V. (eds.) Methods of Mathematical Oncology. MMDS 2020. Springer Proceedings in Mathematics & Statistics, Vol 370, pp. 153–172. Springer, Singapore (2021)

[6] Anderson, A.R.A., Chaplain, M.A.J., Newman, E.L., Steele, R.J.C., Thompson, A.M.: Mathematical modelling of tumour invasion and metastasis. J. Theor. Medicine **2**, 129–154 (2000)

[7] Chaplain, M.A.J., Lolas, G.: Mathematical modelling of cancer cell invasion of tissue: the role of the urokinase plasminogen activation system. Math. Modell. Methods. Appl. Sci. **15**, 1685–1734 (2005)

[8] Chaplain, M.A.J., Lolas, G.: Mathematical modelling of cancer invasion of tissue: Dynamic heterogeneity. Net. Hetero. Med. **1**, 399–439 (2006)

[9] Andasari, V., Gerisch, A., Lolas, G., South, A.P., Chaplain, M.A.J.: Mathematical modelling of cancer cell invasion of tissue: biological insight from mathematical analysis and computational simulation. J. Math. Biol. **63(1)**, 141–71 (2011)

[10] Einstein, A.: Investigations on the Theory of the Brownian Movement. Dover Publications, New York (1998)

[11] Stratonovich, R.I.: A new representation for stochastic integrals and equations. SIAM J. Control **4(2)**, 362–371 (1966)

[12] Ito, E., Ozawa, S., Kijima, H., Kazuno, A., Nishi, T., Chino, O., Shimada, H., Tanaka, M., Inoue, S., Inokuchi, S., Makuuchi, H.: New

invasive patterns as a prognostic factor for superficial esophageal cancer. J. Gastroenterol. **47**(12), 1279–1289 (2012)

[13] Kitanidis, P.K.: Particle-tracking equations for the solution of the advection-dispersion equation with variable coefficients. Water Resour. Res. **30**, 3225–3227 (1994)

[14] Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)

[15] Evans, L.C.: An Introduction to Stochastic Differential Equations. American Mathematical Society, Providence (2013)

[16] Arnold, L.: Stochastic Differential Equations: Theory and Applications, 1st edn. John Wiley, New York (1974)

[17] Kolbe, N., Katuchova, J., Sfakianakis, N., Hellmann, N., Lukacova-Medvidova, M.: A study on time discretization and adaptive mesh refinement methods for the simulation of cancer invasion: The urokinase model. Appl. Math. Comput. **273**, 353–376 (2016)

[18] Sfakianakis, N., Kolbe, N., Hellmann, N., Lukacova-Medvidova, M.: A multiscale approach to the migration of cancer stem cells: Mathematical modelling and simulations. Bull. Math. Biol. **79**, 209–235 (2017)

[19] MATLAB: MATLAB Version 9.13.0.2105380 (R2022b). The Mathworks, Inc., Natick, Massachusetts (2022)

[20] Anderson, P.E., Reo, N.V., DelRaso, N.J., Doom, T.E., Raymer, M.L.: Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. Metabolomics **4**(3), 261–272 (2008)

[21] Tam, K., Alhiyari, Y., Huang, S., Han, A., Stafsudd, O., Shori, R., John, M.S.: Label-free, real-time detection of perineural invasion and cancer margins in a murine model of head and neck cancer surgery. Sci. Rep. **12** (2022)

[22] Lee, J.H., Ahn, S.H., Park, D.J., Kim, H.H., Lee, H.J., Yang, H.K.: Clinical impact of tumor infiltration at the transected surgical margin during gastric cancer surgery. J. Surg. Oncol. **106(6)**, 772–6 (2012)

[23] Spiro, R.H., Guillamondegui, O.J., Paulino, A.F., Huvos, A.G.: Pattern of invasion and margin assessment in patients with oral tongue cancer. Head & Neck **21(5)**, 408–13 (1999)

[24] Yokota, T., Sawai, K., Yamaguchi, T., Taniguchi, H., Shimada, S., Yoneyama, C., Takahashi, T.: Resection margin in patients with gastric

cancer associated with esophageal invasion: clinicopathological study. J. Surg. Oncol. **53(1)**, 60–3 (1993)

[25] Kim, B.S., Oh, S.T., Yook, J.H., Kim, H.S., Lee, I.S., Kim, B.S.: Appropriate gastrectomy resection margins for early gastric carcinoma. J. Surg. Oncol. **109(3)**, 198–201 (2014)

[26] Tsujitani, S., Okuyama, T., Orita, H., Kakeji, Y., Maehara, Y., Sugimachi, K., Kaibara, N.: Margins of resection of the esophagus for gastric cancer with esophageal invasion. Hepatogastroenterology **42(6)**, 873–7 (1995)

[27] Franssen, L.C., Sfakianakis, N., Chaplain, M.A.J.: A novel 3d atomistic-continuum cancer invasion model: In silico simulations of an in vitro organotypic invasion assay. J. Theor. Biol. **522**, 110677 (2021)

[28] Nurmenniemi, S., Sinikumpu, T., Alahuhta, I., Salo, S., Sutinen, M., Santala, M., Risteli, J., Nyberg, P., Salo, T.: A novel organotypic model mimics the tumor microenvironment. Am. J. Pathol. **175(3)**, 1281–1291 (2009)

[29] Nyström, M.L., Thomas, G.J., Stone, M., Mackenzie, I.C., Hart, I.R., Marshall, J.F.: Development of a quantitative method to analyse tumour cell invasion in organotypic culture. J. Pathol. **205(4)**, 468–475 (2005)

[30] Lakkis, O., Madzvamuse, A., Venkataraman, C.: Implicit–explicit timestepping with finite element approximation of reaction–diffusion systems on evolving domains. SIAM J. Numer. Anal. **51**, 2309–2330 (2012)

[31] Kennedy, C.A., Carpenter, M.H.: Additive Runge-Kutta schemes for convection-diffusion-reaction equations. Appl. Numer. Math. **1**(44), 139–181 (2003)

[32] Krylov, A.N.: On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined. Otdel. mat. i estest. nauk. **VII**(4), 491–539 (1931)

[33] van der Vorst, H.A.: Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. SIAM J. Sci. Comput. **13**(2), 631–644 (1992)