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: 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 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) that dictates the individual cancer cell migration, should account also for the divergence of the diffusion of the PDE. Furthermore, 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. focussing 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/emerges from the migration of individual cells so should the two mathematical approaches be connected. Concerning cancer invasion, a 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].

The remainder of the paper is organised as follows. 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 the more typical continuum cancer invasion models (e.g. [7], [8], [9]) and consider the following general Advection-Diffusion partial differential equation (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 will 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 $\mathbf{x} \in \mathbb{R}^d$.

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 quite typically described via stochastic differential equations (SDEs) that track the position of the cells. These SDEs generally have 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 $t \geq 0$, $\mathbf{X}_t \in \mathbb{R}^d$, 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 we address 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 will denote by $\mathbf{x}_t \in \mathbb{R}^d$ the position of its centroid at time t, and following [11–14] we will exhibit in

Section 3 that the less intuitive numerical scheme:

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

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

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

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.

Clearly, though, 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), albeit $\mathbf{D}$ is included in a very particular fashion; via it's divergence. In the more intuitive scheme (5) the drift term accounts only for the advection term $\mathbf{A}$.

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

We will, hence, conclude that there is not a 1-1 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 assumption that a cell can be viewed as a sufficiently small cell-cluster that satisfies (1). We can assume, without loss of generality, that 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 (scalar) density and V_∞ is the volume of the cell-cluster.

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

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

and

$$\nu_2(t) = \int_{V_\infty} \mathbf{x}\mathbf{x}^T u(t,\mathbf{x})\, d\mathbf{x}. \tag{7b}$$

We also consider the second moment about the mean

$$N(t) = \int_{V_\infty} (\mathbf{x} - \boldsymbol{\nu}_1(t))\,(\mathbf{x} - \boldsymbol{\nu}_1(t))^T\, u(t,\mathbf{x})\, d\mathbf{x} = \nu_2(t) - \boldsymbol{\nu}_1(t)\boldsymbol{\nu}_1^T(t). \tag{8}$$

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 (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 $\dfrac{dN(t)}{dt}$ respectively.

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

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

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 V_∞ and $\mathbf{z}$ an n-dimensional vector.

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

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

We work on each one of the terms on the left hand side separately; for the first one we obtain

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

Similarly, for the second

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

and the third term

$$\begin{aligned} I_{13} &= \int_{V_\infty} x_i \nabla \cdot (\mathbf{D}\nabla u)\, d\mathbf{x} = \int_{V_\infty} \nabla \cdot (x_i \mathbf{D}\nabla u)\, d\mathbf{x} - \int_{V_\infty} (\nabla u)^T \mathbf{D}_i\, d\mathbf{x} \\ &= \int_{S_\infty} x_i (\mathbf{D}\nabla u)^T \mathbf{n}\, dS - \int_{V_\infty} \nabla \cdot (u \mathbf{D}_i)\, d\mathbf{x} + \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x} \\ &\overset{(10b)}{=} - \int_{S_\infty} u \mathbf{D}_i^T \mathbf{n}\, dS + \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x} \\ &\overset{(10c)}{=} \int_{V_\infty} u \nabla \cdot \mathbf{D}_i\, d\mathbf{x} \end{aligned} \tag{11d}$$

So, by combining (11a)-(11d) we obtain the following representation for the rate of change of the first moment:

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

In a similar way as above, we identify the relation satisfied by the ij^{th} element of the second moment ν_2:

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

Yet again, we calculate these three integrals one-by-one. The first one reads

$$I_{21} = \int_{V_\infty} x_i x_j \frac{\partial u}{\partial t}\, d\mathbf{x} = \frac{\partial}{\partial t}\int_{V_\infty} x_i x_j u\, d\mathbf{x} = \frac{d(\nu_2)_{ij}}{dt}, \tag{13b}$$

the second integral recasts into

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

and for the third term, it holds

$$\begin{aligned} I_{23} &= \int_{V_\infty} x_i x_j \nabla\cdot(\mathbf{D}\nabla u)\, d\mathbf{x} \\ &= \int_{V_\infty} \nabla\cdot(x_i x_j \mathbf{D}\nabla u)\, d\mathbf{x} - \int_{V_\infty} \left[x_j(\mathbf{D}\nabla u)_i + x_i(\mathbf{D}\nabla u)_j\right] d\mathbf{x} \\ &= \int_{S_\infty} x_i x_j (\mathbf{D}\nabla u)^T \mathbf{n}\, dS - \int_{V_\infty} \nabla\cdot(u x_j \mathbf{D}_i + u x_i \mathbf{D}_j)\, d\mathbf{x} \\ &\quad + \int_{V_\infty} \left[x_j \nabla\cdot\mathbf{D}_i + x_i \nabla\cdot\mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x} \\ &= -\int_{S_\infty} (u x_j \mathbf{D}_i + u x_i \mathbf{D}_j)^T \mathbf{n}\, d\mathbf{x} \\ &\quad + \int_{V_\infty} \left[x_j \nabla\cdot\mathbf{D}_i + x_i \nabla\cdot\mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x} \\ &= \int_{V_\infty} \left[x_j \nabla\cdot\mathbf{D}_i + x_i \nabla\cdot\mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x}. \end{aligned} \tag{13d}$$

By combining (13a)-(13d), we obtain the following relation for the second moment

$$\frac{d(\nu_2)_{ij}}{dt} = \int_{V_\infty} (x_j A_i + x_i A_j) u\, d\mathbf{x} + \int_{V_\infty} \left[x_j \nabla\cdot\mathbf{D}_i + x_i \nabla\cdot\mathbf{D}_j + 2D_{ij}\right] u\, d\mathbf{x}. \tag{14}$$

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

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

$$\frac{d\nu_2}{dt} = \int_{V_\infty} \left(\mathbf{x}(\mathbf{A} + \nabla\cdot\mathbf{D})^T + (\mathbf{A} + \nabla\cdot\mathbf{D})\mathbf{x}^T + 2\mathbf{D}\right) u(t,\mathbf{x})\, d\mathbf{x} \tag{16}$$

$$\frac{dN}{dt} = \int_{V_\infty} \big(\mathbf{x}(\mathbf{A} + \nabla \cdot \mathbf{D})^T + (\mathbf{A} + \nabla \cdot \mathbf{D})\mathbf{x}^T + 2\mathbf{D}\big) u(t, \mathbf{x})\, d\mathbf{x} - \boldsymbol{\nu}_1(t) \left(\int_{V_\infty} (\mathbf{A} + \nabla \cdot \mathbf{D}) u(t, \mathbf{x})\, d\mathbf{x} \right)^T - \left(\int_{V_\infty} (\mathbf{A} + \nabla \cdot \mathbf{D}) u(t, \mathbf{x})\, d\mathbf{x} \right) \boldsymbol{\nu}_1^T(t). \tag{17}$$

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

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

where δ is the Dirac function and $\mathbf{x}_t$. Then the velocity of the cell is given as the sum of the advection $\mathbf{A}$ and the divergence of the diffusion matrix $\mathbf{D}$:

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

Accordingly, we also see that equation (17) recasts into

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

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 and note that by substituting (18) and (19) into (9) we 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 clearly different from the one without the correction of the drift term, namely from

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

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

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

In the next section we will 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 corrected SDE scheme (20) incorporates a diffusion-based contribution that is not found in the SDE scheme (21). This is a significant difference and is central to 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 will consider 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 will investigate the behaviour of the two SDE schemes, (21) and (20), on a particular application where the corresponding noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ are the same. This will allow to compare the two schemes as a result of their difference on the drift term alone.

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

In **Experiment** 3 we will 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 implicitly through a function that describes the environment. For both SDE schemes (21) and (20), we will perform a large number of realisations and compare them with the solution of the underlying PDE. We will accordingly conclude that the corrected SDE scheme (20) provides a much better approximation to the PDE (1) than the SDE scheme (21).

For the numerical solution of the PDE (1) we use a numerical method that was previously developed in [15, 16] and which we briefly discuss in the Appendix A. All algorithm implementations and visualisations were conducted in MATLAB [17].

Experiment 1. In the first experiment, we consider an experimental setting where the advection and diffusion terms, $\mathbf{A}$ and $\mathbf{D}$ of the PDE (1) have an

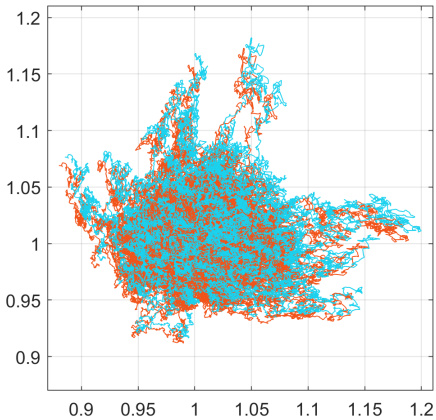

Fig. 1 – Exp. 1. Cell migration tracks of 100 realisations of the SDE schemes (23), in orange, and (24), in light blue, with the same random noise. The clear shift between the two schemes is due to the adaptation of the drift term of the scheme (24).

explicit dependence on the space variable $\mathbf{x}$, namely:

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

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

$$\frac{\partial u}{\partial t} + \nabla \cdot \left[\begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} u \right] - \nabla \cdot \left[\begin{pmatrix} bx_1^7 & 0 \\ 0 & bx_2^7 \end{pmatrix} \nabla u \right] = 0, \tag{22}$$

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

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

as follows:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} \Delta t + \sqrt{2\Delta t} \begin{pmatrix} \sqrt{bx_1^7} & 0 \\ 0 & \sqrt{bx_2^7} \end{pmatrix} \boldsymbol{\xi}_1, \tag{23}$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \left[\begin{pmatrix} ax_1^7 \\ ax_2^7 \end{pmatrix} + \nabla \cdot \begin{pmatrix} bx_1^7 & 0 \\ 0 & bx_2^7 \end{pmatrix} \right] \Delta t + \sqrt{2\Delta t} \begin{pmatrix} \sqrt{bx_1^7} & 0 \\ 0 & \sqrt{bx_2^7} \end{pmatrix} \boldsymbol{\xi}_2. \tag{24}$$

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 the realisations of the schemes.

The parameters used in these numerical simulations can be found in Table 1 and the simulation results are presented in Figure 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,

the corrected scheme (24) introduces an additional displacement of the cell migration track, as a whole, directed towards larger values of x_1 and x_2. This is justified by the nature and structure of the drift term, the diffusion-based correction introduced in (24) and the positivity of the parameter b, cf. Table 1.

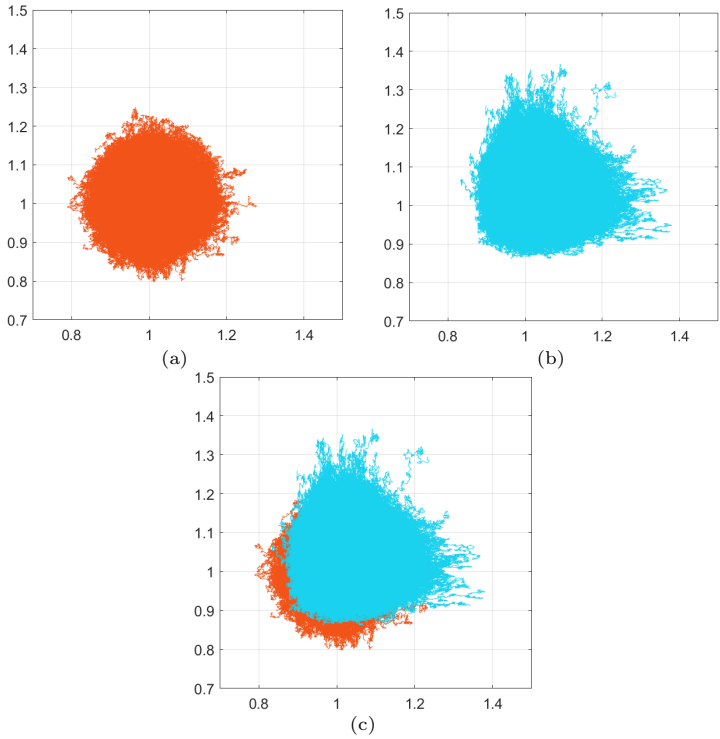

Fig. 2 – **Exp. 2.** Comparison of the migration tracks traversed the cells following the SDE schemes (23) and (24) respectively. For both schemes, we considered initial position at $(1, 1)$ and a total travelling time $T = 1$. (a) The full migration tracks of 10^5 cells following the SDE scheme (23). (b) The full migration tracks of 10^5 cells following the corrected scheme (24). (c) Superimposing the migration patterns of (a) and (b) and clearly see that the corrected scheme (24) introduces additional migratory bias.

Experiment 2. The intuition we have gained from Experiment 1, namely the way the adaptation of the drift term in the corrected scheme (24) 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 current experiment, we choose independent noise terms $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ for the two SDE schemes (23) and (24) and perform 10^5 new realisations with each. We otherwise consider exactly the same setting and parameters as in **Experiment** 1.

The simulation results are presented in Figures 2 and 3. In Figure 2, in particular, we see that cells migrate further and in a more biased fashion

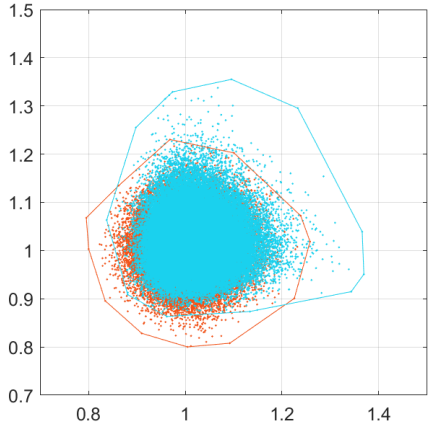

Fig. 3 – **Exp. 2.** Showing here the final positions of the cells migration after 10^5 realisations of the SDE schemes (23) and (24)—in orange and light blue respectively—as shown in Figure 2, along with the convex hull of these points. The impact of the adaptation in the drift term of the corrected scheme (24) induces additional migration and spread than the scheme.

when the follow the corrected scheme (24) rather than the scheme (23). This is clearly the result of the additional bias introduced in the drift term of the corrected scheme (24). The difference between the two schemes can be further seen in Figure 3 where we present the convex hull of the positions of the cells in the final time T.

We can furthermore quantify the difference between the two schemes by measuring the average distance traversed by the cells from their initial position. In more detail, we perform K realisations of the schemes and calculate for each one of them the distance between the initial and the final position. We then calculate the average distance by the formula

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

where $\mathbf{x}^i_{t_0} = \left(x^i_{1,t_0}, x^i_{2,t_0}\right)$, and $\mathbf{x}^i_T = \left(x^i_{1,T}, x^i_{2,T}\right)$ represent the initial and final positions of the cells in the realisation $i = 1, ..., K$. We apply the above formula for the two schemes (23) and (24), i.e. we perform $K = 10^6$ realisations of each scheme, calculate their respective average distances d_1 and d_2, and evaluate the relative difference between the two to obtain

$$E = \frac{|d_1 - d_2|}{d_1} \approx 0.076.$$

Experiment 3. In this experiment we consider a PDE that is typical in the field of cancer invasion more typically, than the PDE (22). 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

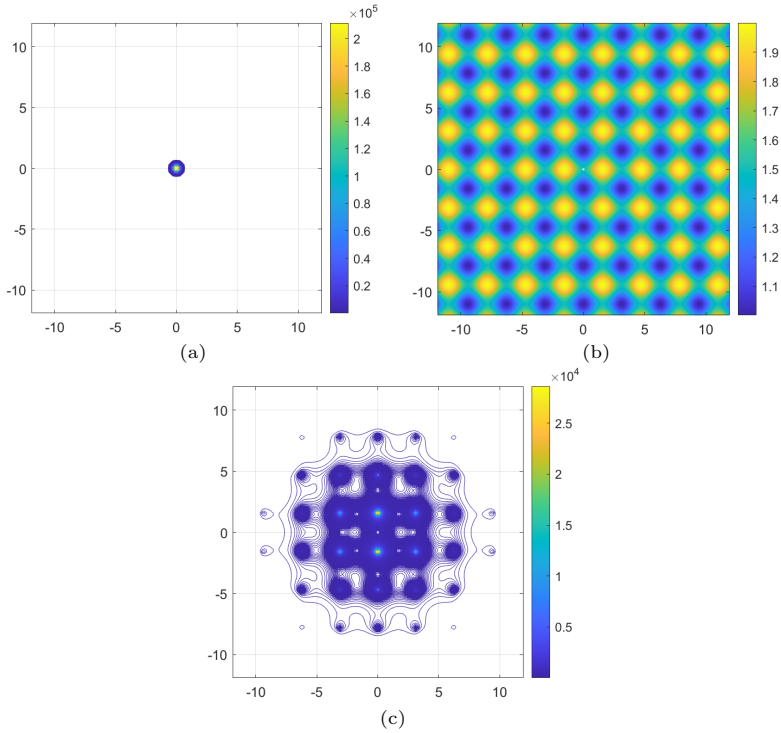

Fig. 4 – **Exp. 3.** (a) Showing here are isolines of the concentration u at the time $t = 0$; they serve as initial conditions for the PDE (25). (b) The structure of the (fixed) environment v that participates in the advection and diffusion terms of (25); the formula of v is given in (27). (c) Isolines of the solution u of the PDE (25) at the final time $t = 10$; they reveal a higher concentration of the at the "valleys" of the (fixed) environment function v seen in (b).

of the extracellular matrix, or a completely different extracellular queue. Still, for the sake of simplicity of presentation, we will not make any particular biological assumptions on the nature of the extracellular environment and rather refer to it simply as "environment".

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

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) - \nabla \cdot \Big(\nabla v(\mathbf{x}) u(t,\mathbf{x}) \Big) - \nabla \cdot \Big(\left(v(\mathbf{x}) - 1\right) \nabla u(t,\mathbf{x}) \Big) = 0, \qquad (25)$$

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 cancer cells. As previously mentioned, we do not investigate the biomedical realism of this model, nor do we interpret its findings under this light.

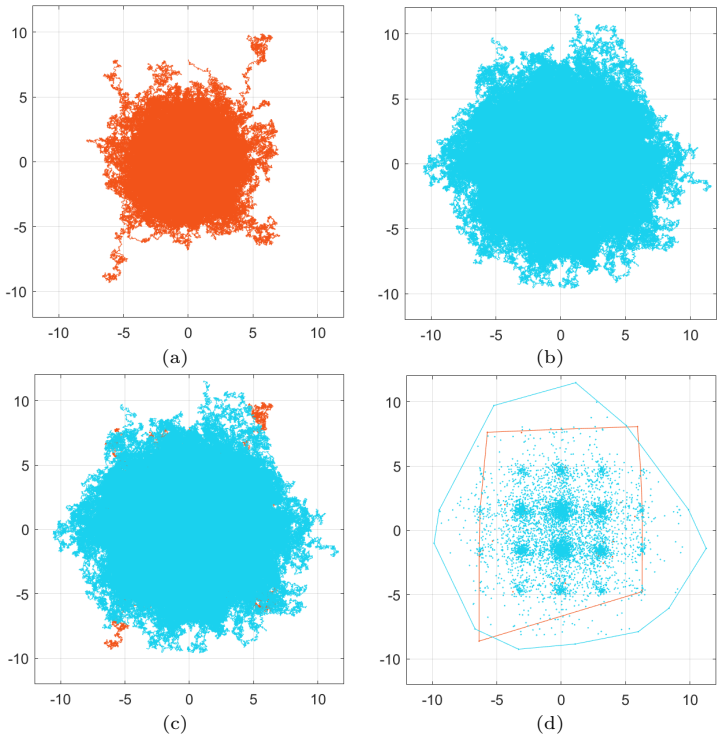

Fig. 5 – **Exp. 3.** Simulations and comparison of 10^4 realisations of SDE schemes (28a) and (28b). (a), (b) & (c): Full set of tracks for the schemes (28a) and (28b) respectively. It is clearly seen that the spread in (b) is much wider than in (a); this is justified by the adaptation introduced in the drift term of (24). This remark is confirmed by superimposing the cell tracks of (a) in (b) in (c). (d): The final position of the 10^4 realisations of (a) and (b) along with convex hulls of the corresponding points. It is clearly seen that the cells concentrate at the "valleys" of the environment v, cf. Figure 4, and that the final positions of (28b) (shown in light blue) spread more than those of (28a).

The PDE (25) is augmented with the initial condition

$$u(0,\mathbf{x}) = u(0,x_1,x_2) = 3\times 10^5\, e^{-20(x_1^2+x_2^2)} \tag{26}$$

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

$$v(\mathbf{x}) = v(x_1,x_2) = 1 + 0.5\left(\sin^2(x_1)+\sin^2(x_2)\right). \tag{27}$$

Note that, for v given in (27), the diffusivity $v(\mathbf{x})-1$ of (25) is non-negative for all $\mathbf{x}\in\Omega$.

Based on the advection and diffusion terms of the PDE (25), we rewrite the SDE schemes (21) and (20) in the form:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t - \nabla v(\mathbf{x}_t)\Delta t + \sqrt{2\Delta t}\sqrt{v(\mathbf{x}_t)-1}\,\boldsymbol{\xi}_1, \tag{28a}$$

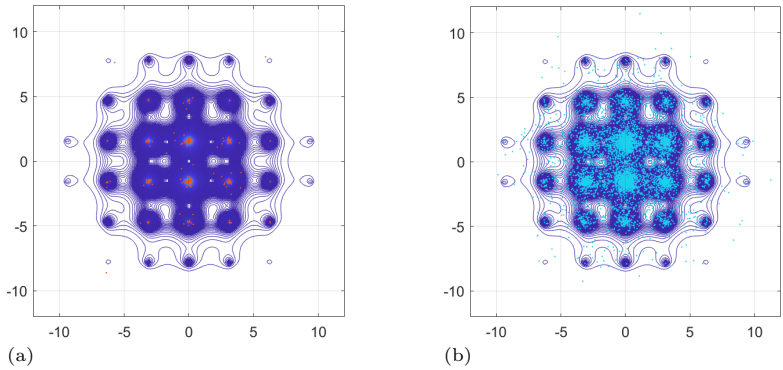

Fig. 6 – **Exp. 3.** Comparison of the numerical solution of the PDE (25) against the predictions of the SDE schemes (28a) and (28b). (a) Isolines of the final time solution u of the PDE (25), cf. Figure 4, superimposed with 10^4 (final time) realisations of the SDE scheme (28a). (b) Similarly as in (a), we superimpose with 10^4 (final time) realisations of the corrected SDE scheme (28a) with the isolines of the final time solution u of the PDE (25). We note that the cells described by the SDE scheme (28a) are concentrated, at the final time, almost exclusively at the "valleys" of the (fixed) environment v, cf. Figure 4, much more than cells described by the SDE scheme (28b), and much more than the final time solution u of the PDE (25).

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \sqrt{2\Delta t}\sqrt{v(\mathbf{x}_t) - 1}\boldsymbol{\xi}_2. \tag{28b}$$

Note that the corrected SDE scheme (28b) lacks a drift term; this is a result of the particular structure of the advection and diffusion terms of the PDE (25) and the way they are combined in the drift term of the corrected scheme (20).

In Figure 6 we present a direct comparison between the numerical solution of the PDE (25) and the SDE schemes (28a) and (28a). What we exhibit there is that, at the final time $t = 10$, the positions of the cells of 10^4 realisations of the SDE scheme (28a) are much more concentrated than the corresponding solution of the PDE (25). On the other hand, the distribution at the same final time $t = 10$, of 10^4 realisations of the corrected scheme (28b) is much closer to the solution of the PDE (25).

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)?

As we have exhibited in this paper, 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 regime would have been solely dependent on the advection term $\mathbf{A}$, as e.g. is the case in (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 involve 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—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. 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. [18], [19], [20].

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

$$\mathbf{w}_t = A(\mathbf{w}) + R(\mathbf{w}) + D(\mathbf{w}), \tag{29}$$

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

$$\partial_t \mathbf{w}_h = \mathcal{A}(\mathbf{w}_h) + \mathcal{R}(\mathbf{w}_h) + \mathcal{D}(\mathbf{w}_h), \tag{30}$$

where the numerical operators $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{D}$ are *discrete approximations* of the operators A, R, and D in (29) respectively.

We use a second order *Implicit-Explicit Runge-Kutta* (IMEX-RK) *Finite Volume* (FV) numerical method that was previously developed in [15, 16] where we refer for more details, see also [21]. This method is based on a *splitting* in explicit and implicit terms in the form

$$\partial_t \mathbf{w}_h = \mathcal{I}(\mathbf{w}_h) + \mathcal{E}(\mathbf{w}_h). \tag{31}$$

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

$$\begin{cases} \mathbf{W}_i^* = \mathbf{w}_h^n + \tau_n \sum_{j=1}^{i-2} \bar{a}_{i,j}\mathbf{E}_j + \tau_n \bar{a}_{i,i-1}\mathbf{E}_{i-1}, & i = 1\dots s \\ \mathbf{W}_i = \mathbf{W}_i^* + \tau_n \sum_{j=1}^{i-1} a_{i,j}\mathbf{I}_j + \tau_n a_{i,i}\mathbf{I}_i, & i = 1\dots s \ , \\ \mathbf{w}_h^{n+1} = \mathbf{w}_h^n + \tau_n \sum_{i=1}^{s} \bar{b}_i \mathbf{E}_i + \tau_n \sum_{i=1}^{s} b_i \mathbf{I}_i \end{cases} \tag{32}$$

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. [22].

We solve the linear systems in (32) using the *iterative biconjugate gradient stabilised Krylov subspace* method, for which we refer to [23, 24].

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)

Table 2 Butcher tableaux for the explicit (upper) and the implicit (lower) parts of the third order IMEX scheme (32), see also [22].

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	$\frac{1471266399579}{7840856788654}$	$-\frac{4482444167858}{7529755066697}$	$\frac{11266239266428}{11593286722821}$	$\frac{1767732205903}{4055673282236}$
	$\frac{1471266399579}{7840856788654}$	$-\frac{4482444167858}{7529755066697}$	$\frac{11266239266428}{11593286722821}$	$\frac{1767732205903}{4055673282236}$

[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] 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)

[16] 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)

[17] MATLAB: MATLAB Version 9.10.0.1613233 (R2021a). The Mathworks, Inc., Natick, Massachusetts (2021)

[18] 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)

[19] 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)

[20] 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)

[21] 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)

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

[23] 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)

[24] 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)