Data-driven spatio-temporal modelling of glioblastoma

Andreas Christ Sølvsten Jørgensen, Ciaran Scott Hill, Marc Sturrock, Wenhao Tang, Saketh R. Karamched, Dunja Gorup, Mark F. Lythgoe, Simona Parrinello, Samuel Marguerat and Vahid Shahrezaei

Mathematical oncology provides unique and invaluable insights into tumour growth on both the microscopic and macroscopic levels. This review presents state-of-the-art modelling techniques and focuses on their role in understanding glioblastoma, a malignant form of brain cancer. For each approach, we summarize the scope, drawbacks and assets. We highlight the potential clinical applications of each modelling technique and discuss the connections between the mathematical models and the molecular and imaging data used to inform them. By doing so, we aim to prime cancer researchers with current and emerging computational tools for understanding tumour progression. By providing an in-depth picture of the different modelling techniques, we also aim to assist researchers who seek to build and develop their own models and the associated inference frameworks. Our article thus strikes a unique balance. On the one hand, we provide a comprehensive overview of the available modelling techniques and their applications, including key mathematical expressions. On the other hand, the content is accessible to mathematicians and biomedical scientists alike to accommodate the interdisciplinary nature of cancer research.

Received: 8 November 2022
Accepted: 23 February 2023
1. Introduction

Glioblastoma (GBM) is a malignant hierarchically organized brain cancer. It is both the most common and most aggressive type of primary brain cancer in adults [1]. Not only does the diffusive invasion of glioma cancer cells into healthy tissue impede complete resection, but also GBM harbours a sub-population of highly therapy-resistant stem-like cells [2]. Tumour recurrence is inevitable, resulting in a median survival time of 15 months for patients despite maximal treatment [3]. The current gold standard of treatment is the Stupp protocol, which consists of maximal safe surgical resection, followed by radiotherapy and chemotherapy with temozolomide, an alkylating agent [4,5].

Mathematical cancer models have provided a deeper understanding of this immensely complex disease by unveiling the underlying mechanisms and offering quantitative insights [6–9]. Such models have entered all areas of GBM research, ranging from the classification and detection of brain tumours to therapy [10,11].

This review provides the reader with an overview of existing mathematical and computational models that aim to simulate spatially resolved tumour growth. We discuss three main paradigms that have emerged for such in silico experiments. In §2, we introduce the so-called continuum models that treat variables, such as the tumour cell density, as continuous macroscopic quantities based on conservation laws. Alternatively, one might represent each cell as an individual agent. Such discrete models are discussed in §3. Section 4 deals with hybrid multi-scale and multi-resolution models that merge and bridge different approaches.

Each of the three methods has its advantages and shortcomings. One must choose between them based on computational limitations and the level of detail required to answer the research questions of interest. With this review, we aim to assist researchers in choosing between the different methods by highlighting the drawbacks and assets of each approach and by showing how the different methods can complement each other. Moreover, we summarize the main concepts and the key mathematical expressions that lie at the core of each approach. We hereby aim to strike a balance between providing a brief overview and showing the mathematics involved. In contrast, many other recent reviews of cancer modelling limit the mathematical details in favour of providing a very concise overview of the literature [12–14].

Mathematical models of GBM draw on a wide variety of molecular and imaging data: histopathological data, computerized tomography (CT), positron emission tomography (PET), single-photon emission computerized tomography (SPECT) and magnetic resonance imaging (MRI), such as T1 weighted (+/− gadolinium contrast), T2 weighted, T2-FLAIR, diffusion-weighted imaging, and most recently spatial and single-cell transcriptomics. The models have thus been employed to shed light on patient-specific data in vivo and ex vivo, as well as on data from animal models and in vitro experiments. These analyses have provided invaluable insights and deepened our understanding of glioma on molecular and structural levels. We address this issue in more detail in §5. The section also discusses the computational challenges posed by systematic inference of model parameters. Finally, §§6 and 7 provide a short overview of some clinical applications of these models and an overall summary, respectively. By including the topics addressed in §§5 and 6, we give our article a more holistic scope with which we aim to set our article apart from other recent reviews on cancer modelling.

While we discuss the different methods in light of GBM, we note that the same models are applied to other types of cancers. Indeed, the models build on concepts that are broadly used to study tissues. Therefore, we not only cite sources that deal with GBM but also occasionally refer the reader to illustrative papers from other areas of oncology and biology [15,16]. It is worth noting that the models are, in a broader sense, actually widely used across scientific disciplines. Throughout the article, we thus present ideas and concepts that are also employed in other fields, ranging from statistical mechanics to solid-state physics. For instance, the cellular Potts model (CPM) presented in §3.1.3 builds on the so-called Ising model used to describe ferromagnetism. We hence encourage the reader to ‘think outside the box’ when exploring the literature, and, in this spirit, we provide a few citations to areas outside the realm of biology.

Before commencing, we would like to point the reader towards other reviews and papers for further details. Lowengrub et al. [17] provide a detailed account of continuum models. For an elaborate discussion on discrete models, we refer the reader to Van Liedekerke et al. [18]. Both Metzcar et al. [12] and Weeransinghe et al. [13] give a brief overview of the topic and include a list of recent references. For insights into hybrid multi-scale modelling, we recommend Deisboeck et al. [19], and Chamseddine and Rejniak [20]. Falco et al. [21] present a concise overview highlighting their clinical implications. Ellis et al. [22] focus on mathematical models that address intratumour heterogeneity and tumour recurrence based on next-generation sequencing techniques. Alfonso et al. [23] highlight the challenges that mathematical models face when dealing with glioma invasion. Finally, for an overview
of the biology of the GBM from a clinical perspective, we refer the reader to the recent reviews by Finch et al. [11] and McKinnon et al. [24].

2. Continuum models

When dealing with cancer treatment, we face questions related to tumour size, shape and composition. These questions all address cancer on a macroscopic scale. While macroscopic tumour dynamics emerge from interactions on a cellular level, it is possible to construct informative mathematical models without tracking individual cancer cells. Instead, the tumour and its environment can be represented as continuous variables that are governed by partial differential equations (PDEs). Such continuum models capture many aspects of cancer in vitro, in vivo and in patients. They can account for the impact of heterogeneous brain tissue on tumour growth, for different invasive tumour morphologies and, to some extent, even for potential tumour recurrence [25,26]. Moreover, they have successfully been applied in studies on the impact of chemotherapy and repeated immunosuppression treatment [25,26]. Continuum models have thus been employed in diagnosis and treatment planning based on patient-specific data [29–33].

However, PDEs smooth out small-scale fluctuations, which implies that continuum models do not apply to small cell populations, such as those found in the tumour margin. For such small cell numbers, stochastic events play a crucial role, and the applicability and predictive power of continuum models are limited. To properly understand cancer invasion, we must track individual cells. But to do so comes at a high or currently insurmountable computational cost (see §3). Thus, the use of continuum models represents a trade-off that enables and supports scalability and mathematical insights. As a result, continuum models are widely used in the community [13,17,34].

This section introduces the basic mathematical concepts of continuum models and their biological motivation. Any such model in the literature builds on reaction–diffusion equations, describing variables such as the tumour cell density, tumour volume fraction, nutrient (oxygen, glucose) concentration, neovasculature, enzyme concentration, or other properties of the extracellular matrix (ECM) [25,26,35–38]. Considering any such variable, \( \rho(x, t) \), which is a function of position \( x \) and time \( t \), we have for its rate of change with time

\[
\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J} + S,
\]  

(2.1)

where \( \mathbf{J} \) is the flux of the considered variable, and \( S \) is the sources and sinks for this variable. Thus, equation (2.1) constitutes a conservation law. The exact expression for \( \mathbf{J} \) and \( S \), as well as the boundary conditions, will depend on the variable in question. For instance, when dealing with the quasi-steady diffusion of nutrients, equation (2.1) generally takes the following form [39–43]:

\[
0 = D \nabla^2 \rho + S.
\]

(2.2)

Here, \( D \) denotes a diffusion coefficient, while \( \rho(x) \) is the relevant nutrient concentration at the location \( x \) within the considered \( n \)-dimensional domain, which is often denoted by \( \Omega \).

Alternatively, let’s consider the (normalized) cancer cell density, \( \rho(x, t) \), at a time \( t \) and location \( x \). Many authors assume that the diffusion of cancer cells can be well approximated by Fick’s first law

\[
\mathbf{J} = -D \nabla \rho,
\]

(2.3)

where \( D \) denotes a diffusion coefficient, which we discuss in detail in §2.1. As regards the sources and sinks of the cancer cell density, it is commonly assumed that cell proliferation, i.e. tumour growth, is well described by a logistic growth term [44,45]. So, equation (2.1) takes the following form:

\[
\frac{\partial \rho}{\partial t} = \nabla \cdot (D \nabla \rho) + \lambda \rho(1 - \rho),
\]

(2.4)

where \( \lambda \) is the growth rate of the tumour cell population. Other authors assume exponential tumour growth, substituting the second term on the right-hand side by \( \lambda \rho \) [29]. Of course, more than one term might be necessary to summarize the relevant sources and sinks. Several more complex terms are needed, for instance, when considering differentiation between different interdependent tumour sub-populations (cf. §2.2).

It is worth stressing that equation (2.1) cannot stand on its own. Other relations and constraints, including (Neumann) boundary conditions, are needed. One might, for instance, naturally require that
there is no flux at the boundary of the brain domain, i.e. that the tumour does not penetrate the patient’s skull [45–47]. This being said, equation (2.1) is the backbone of any continuum model that spatially resolves the tumour (see also §4.1.1).

While we focus on spatio-temporal cancer models in this review, it is worth noting that equation (2.1) can be seen as a natural extension to non-spatial models for tumour growth. Indeed, if we drop any spatial dependence in equation (2.1), including the diffusion term, we end up with an ordinary differential equation (ODE) for the tumour cell density of the form

$$\frac{d\rho}{dt} = S.$$  

(2.5)

Depending on the source term, $S$, equation (2.5) might describe exponential, logistic or Gompertz tumour growth laws that serve as the foundation for non-spatial cancer models. Building on equation (2.5), one might thus construct a sophisticated network of coupled ODEs (or delay differential equations) that might differentiate between different tumour cell sub-populations, or account for immune responses or the effect of cancer treatment (e.g. [48–51] and the references therein). Non-spatial ODEs are thus widely used in mathematical oncology [52]. While such models do not spatially resolve the tumour, they have successfully been used to study the response to drugs and radiation treatments, recovering experimental constraints on the overall growth pattern [53,54]. Here, however, we only intend to mention equation (2.5) in passing to put equation (2.1) into perspective; any further exploration of non-spatial models lies beyond the scope of this review.

2.1. Anisotropic diffusion

By using a scalar for the diffusion coefficient in equation (2.4), we assume isotropic tumour growth. However, GBM spreads anisotropically, primarily expanding along pre-existing structures, such as blood vessels and white matter tracks [55–61]. To take the heterogeneous structure of the brain into account, Swanson et al. [29] hence proposed to adopt different values for the diffusion coefficients in grey and white matter. Concretely, on the basis of CT scans by Tracqui et al. [62], Swanson et al. [29] found the diffusion coefficient in white matter to be more than five times larger than in grey matter.

Taking the idea of heterogeneous diffusion further by including anisotropy, other authors [47,58,63–65] substitute the diffusion coefficient with an $n$-dimensional diffusion tensor, $D(x,t) \in \mathbb{R}^{n \times n}$. To evaluate $D(x,t)$, Painter and Hillen [47] deploy diffusion tensor imaging (DTI) data. DTI is an MRI technique that measures the anisotropic diffusion of water molecules and hereby maps highly structured tissue. This technique provides the diffusion tensor for water molecules, $D_H(x,t)$, throughout the brain. Of course, due to the size difference, the movement of cancer cells is more restricted than that of water molecules, which means that $D_H(x,t)$ does not adequately describe glioma growth, i.e. $D(x,t) \neq D_H(x,t)$. However, on the basis of a transport equation for individual cell movement, Painter and Hillen [47] establish a relation between $D(x,t)$ and $D_H(x,t)$ expressed in terms of the fractional anisotropy that is commonly used to quantify DTI data [66]. They do so based on a set of simplifying assumptions and parabolic scaling to a macroscopic model (see [67] for further details). Their final macroscopic model takes the following form:

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (D \rho) + \lambda \rho (1 - \rho),$$  

(2.6)

where the colon denotes a contraction [26]. The diffusion tensor $D \in \mathbb{R}^{n \times n}$ is symmetric and positive-definite, as it is related to the variance-covariance matrix of the probability distribution function that describes the velocity changes of individual cells [47,67]. In other words, when dealing with three-dimensional data, $D(x,t)$ is a symmetric and positive-definite $3 \times 3$ matrix that incorporates the impact of the local environment on cell migration. The model by Painter and Hillen [47] has been extended and applied by other authors [26,68,69].

Note that equation (2.6) is subtly different from equation (2.4). Apart from $D$ denoting an $n \times n$ tensor rather than a scalar, equation (2.6) includes an additional advective-type term since $\nabla \cdot (D \rho) = \nabla \cdot (D \nabla \rho) + \nabla \cdot (D \nabla \rho)$ [26]. Models of the form of equation (2.6) are referred to as Fokker–Planck models, while equation (2.4) is an example of a Fickian model. The additional advection term of the Fokker–Planck model has a demonstrable impact on the solution [70]. We also note that equation (2.6) would correspond to the Fisher’s equation if the first term on the right-hand side was substituted by $D \nabla^2 \rho$. Indeed, some authors employ a diffusion term of this kind to describe the cancer cell density [71].
2.2. Mechanical interactions, cell types, lineage and feedback

Tumours are hierarchically organized. GBMs harbour stem-like cells (GSCs) as well as proliferating (GCP) and differentiating (GTP) sub-populations (cf. [2,61,72,73]). Since these three cell types exhibit very different behaviours, it is insightful to differentiate between these sub-populations, even when dealing with continuum models.

Models that distinguish between viable and necrotic tumour tissue are the first step in this direction. Examples of such models can be found in the papers by Wise et al. [74,75] and Frieboes et al. [76,77]. Their work is based on reaction–diffusion equations of the form

$$\frac{\partial \rho_i}{\partial t} + \nabla \cdot (u_i \rho_i) = -\nabla \cdot J_{\text{mec},i} + S_i,$$  \hspace{1cm} (2.7)

where the index $i$ runs over all sub-populations, $u_i$ denotes the velocity of the considered cell species and $J_{\text{mec},i}$ is the flux that arises from mechanical interactions as follows:

$$J_{\text{mec},i} = J_i - \rho_i u_i,$$  \hspace{1cm} (2.8)

Note that while equation (2.7) appears to differ from the other reaction–diffusion equations listed earlier, this is merely a matter of notation. It is a Fickian model that can be derived by inserting equations (2.3) and (2.8) into equation (2.1). The advantage of phrasing the problem in this manner is that the mechanical flux reflects the mechanical interaction energy that can be obtained from an understanding of the underlying cell biology (see also [25,78–80]). By introducing $J_{\text{mec},i}$, it is thus possible to inform the model about the properties of the tumour and host without the need to construct a suitable diffusion tensor.

In articles that employ equation (2.8), $u_i$ is computed by imposing relations similar to Darcy’s law that links the velocity to a gradient in pressure [25]. For glioma, the relevant expression often takes the form $u_i = -\nabla p + F_i$, where $p$ is the solid pressure arising from the tumour proliferation, while $F_i$ reflects mechanical interactions.

We exemplify the source functions that enter equation (2.7) by listing the relevant terms for the necrotic tissue according to Wise et al. [74], for which

$$S_d = \lambda_A \rho_d + \lambda_N H(n_N - n) \rho_v - \lambda_C \rho_d.$$  \hspace{1cm} (2.9)

Here, the indices ‘d’ and ‘v’ refer to the dead and viable cancer cells, respectively, while $\lambda_A$, $\lambda_N$ and $\lambda_C$ denote the rates of apoptosis, necrosis and the clearance of dead cells, respectively. Moreover, $H$ is a Heaviside step function and $n_N$ is a viability limit for the nutrient concentration below which cells die—biologically speaking, this term encodes cell death through starvation and the use of the Heaviside step function is a simplifying assumption. For comparison, the source function for the viable tissue takes a similar form:

$$S_v = -\lambda_A \rho_v - \lambda_N H(n_N - n) \rho_v + \lambda_M \frac{n}{n_\infty} \rho_v,$$  \hspace{1cm} (2.10)

where $\lambda_M$ denotes the rate of mitosis and $n_\infty$ is the far-field nutrient level.

By distinguishing between GSCs, GCPs and GTPs, Kunche et al. [81] and Yan et al. [82–84] have incorporated the GBM lineage and hereby taken the discussed models one step further. Kunche et al. [81] use this to investigate feedback regulation of cell lineage progression. Furthermore, while previous articles only consider adhesion when computing the mechanical interactions, Chen et al. [85–87] include the impact of elastic membranes and the implications of the calcification of dead tumour cells.

2.3. Modelling the macroscopic environment

Understanding the microenvironment is essential for understanding GBM since the brain region and other properties, such as the patient’s age, have been shown to play a key role in tumour development and heterogeneity [61,88].

The concentrations of different chemicals, including (but not limited to) nutrients, drugs, matrix-degrading enzymes and ECM macromolecules, are commonly modelled using reaction–diffusion equations. The associated diffusion is often in the form $D \nabla^2 \psi$, but other second-order spatial derivatives can be found in the literature [89–91]. The source terms reflect the processes at play. For instance, the rate of oxygen consumption is often assumed to be proportional to the local oxygen concentration and might be proportional to the local cell density.
Overall, the use of continuum models to model, say, nutrient flows is justified through the different scales that define cancer. Continuum models reliably capture the spatial gradients and temporal variation of nutrients in a way that allows us to assess the behaviour of tumour cells. For some purposes, it might even be adequate to assume that the nutrient levels stay constant over time since cell proliferation takes place on a much longer time scale than nutrient diffusion (cf. equation (2.2)).

Initially, tumours do not possess their own vasculature but rather rely on the diffusion of nutrients and waste products. During this so-called avascular phase, the tumour may lack some features of malignancy and appear more benign. But beyond a certain size (1–3 mm in radius), the nutrient inflow can no longer sustain the growing cell population [92]. Hypoxia, i.e. oxygen shortage, sets in. Without stimulating angiogenesis, i.e. recruiting vasculature, tumour growth will stagnate. However, in response to the hypoxic conditions, the cancer cells release tumour angiogenic factors (TAFs), which stimulates the migration and proliferation of endothelial cells (ECs). New blood vessels sprout towards the tumour, and tumour growth resumes. Understanding the transition from avascular to vascular tumour growth is essential since it is a critical step towards malignancy. To study angiogenesis, many authors model both the diffusion of TAFs and ECs using reaction–diffusion equations [17,30,92–96]. As regards the ECs, we again encounter logistic or exponential source terms. Meanwhile, the gradient of the TAF concentration and that of adhesive molecules (e.g. fibronectin) enter through a flux term encapsulating chemotaxis and haptotaxis:

\[ J = -D \nabla \rho_{EC} + \sum_i x_i \rho_{EC} \nabla c_i, \]

where \( \rho_{EC} \) denotes the EC cell density and the sum runs over all relevant diffusive components, including TAFs. The concentration of and sensitivity to the \( i \)th component are denoted as \( c_i \) and \( x_i \), respectively [92].

Like continuum models for tumour cells, continuum models for ECs cannot capture the behaviour of individual cells. Some authors, therefore, rely on agent-based models to study the ECs during angiogenesis [97–101]. We refer to §3 for a detailed discussion of agent-based models [102]. Table 1 summarizes the properties of all continuum models discussed in this section.

### 3. Discrete (agent-based) models

Continuum models are based on the assumption that the behaviour of tumour cells is well approximated by macroscopic averages. However, this assumption breaks down for small cell populations since stochastic events dominate these. Therefore, continuum models cannot reliably capture the onset of tumour growth and they do not give a complete picture of the invasive tumour front or margin, i.e. the tumour–host interface, where the cancer cell density is low—although it is possible to address this drawback by linking different scales, as shown by Trucu et al. [103] (see also §4). Due to the aggressive invasion of healthy tissue by GBM, this shortcoming is especially problematic for understanding brain tumours. Moreover, continuum models cannot adequately deal with heterogeneity among cells of the same type. To address these issues, one needs to model cancer on a mesoscopic scale following individual tumour cells or cell clusters [104]. Mathematical models that do so are referred to by various names: discrete models, agent-based models, individual-based models or cell-based models. In all cases, cells or cell clusters are modelled as autonomous agents that follow a set of (stochastic) rules prescribing cell movement, death, division and growth. By doing so, discrete models capture the variability among cells that might, for instance, arise as a response to the microenvironment [89]. Note that while discrete models treat cells as autonomous quantities, continuum models are commonly used to depict the microenvironment (cf. §4.1). PDEs hence still come into play when dealing with, e.g., nutrient and drug concentrations [17,105].

Discrete models that describe spatial tumour growth or other cell populations, such as ECs (cf. §2.3), fall into two categories: lattice-based and off-lattice methods. While the former operates on a fixed lattice, the latter does not impose the same restrictions. Lattice-based models can be divided into three sub-categories [12,18]. Firstly, there exists a group of the so-called cellular automata (CA) models that treat all processes, including cell movement, as stochastic processes (cf. §3.1.1). For these models, each lattice site (e.g. pixel or voxel) can be occupied by at most either a single cell or a population of a limited size [106–109]. Each cell or cell cluster is thus characterized by its position on the grid. Secondly, there exist so-called lattice gas cellular automata (LGCA). While these models consider single cells and are conceptually very similar, they differ from other CA models by attributing a
Table 1. Overview of the reviewed continuum models. The references are not exhaustive.

<table>
<thead>
<tr>
<th>model type</th>
<th>features</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuum model (CM), tumour</td>
<td><strong>Concept:</strong> The tumour cell density can be modelled as a fluid using PDEs based on conservation laws.</td>
<td>Methods: Anisotropic diffusion tensor [26,47,68], cf. §2.1. Interaction</td>
</tr>
<tr>
<td></td>
<td><strong>Assets:</strong> CMs capture macroscopic tumour growth at low computational cost and simple models are analytically tractable.</td>
<td>energy-based models [25,74,84,87] accounting for lineage [82], cf.</td>
</tr>
<tr>
<td></td>
<td><strong>Drawbacks:</strong> They cannot capture the behaviour at low cell densities (e.g. at the tumour front).</td>
<td>§2.2. <em>Patient-specific studies:</em> [30,32,33].</td>
</tr>
<tr>
<td>CM, microenvironment</td>
<td><strong>Concept:</strong> Different properties of the environment, such as nutrient flow or vasculature, are modelled on a macroscopic scale.</td>
<td>Nutrient flow: [89–91].</td>
</tr>
<tr>
<td></td>
<td><strong>Assets:</strong> They are computed at a low computational cost and can be analytically tractable; spatial gradients and time variations, e.g. in nutrient levels, are sufficiently well resolved to study tumour cells.</td>
<td>Vasculature: [17,92].</td>
</tr>
<tr>
<td></td>
<td><strong>Drawbacks:</strong> CMs of ECs do not capture the behaviour of individual cells.</td>
<td>These models are widely used whether the tumour is represented by a CM or by agent-based models (cf. §§3 and 4).</td>
</tr>
</tbody>
</table>
velocity to each cell (cf. §3.1.2). After accounting for stochastic events, these models thus include a
deterministic evaluation of cell propagation at every time-step. Thirdly, there are so-called CPMs that
phrase cell interactions using concepts from statistical mechanics (cf. §3.1.3). CPMs attribute several
lattice points to each cell, whereby these models capture cell morphologies. Off-lattice models, also
known as lattice-free models, come in two flavours. Centre-based models (CBMs) describe cells as
simple particles whose interactions can be expressed as physical forces. Deformable cell models
(DCMs) or vertex models (VMs) extend this picture by resolving the cells using several nodes. Like
CPMs, DCMs can thereby account for cell morphologies.

One drawback of discrete models is their relatively high computational expense due to the large
number of cells that typically need to be simulated. After all, tumours reach $10^5$–$10^6$ cells per mm$^3$.
However, one benefit of discrete models is their ability to capture emergent properties. The simple
rules that govern the (nonlinear) interactions between individual agents will naturally give rise to the
dynamics of macroscopic tumour growth. As discussed in §3.1.2, reaction–diffusion equations might
thus naturally follow from up-scaling agent-based models [110–113]. Due to the ability of discrete
models to capture emergent properties, agent-based models can be calibrated based on patient-specific
data and used to predict macroscopic parameters [114,115].

Concerning the implementation of discrete models, it is worth mentioning that some software tools are
publicly available. There is no need for a research group to start from scratch. Examples of open-source
tools include CHASTE [116], PHYSICELL [117] and COMPUCELL3D [118]. For a recent overview of open-
source toolkits, we kindly refer the reader to table 1 in the article by Metzcar et al. [12]. In this
connection, we note that one must always select and adapt the numerical approach based on the
problem at hand. For instance, not all problems require the same geometry. While colon cancer cells
and cells in vitro move in two dimensions, GBM tumour growth in vivo is intrinsically three
dimensional, which affects the growth pattern [119]. In the following two sections, we will discuss
lattice-based models in more depth (§3.1) and then turn to a detailed discussion of off-lattice models (§3.2).

3.1. Lattice-based models

In lattice-based models, we follow a population of cells on a rigid grid. If a regular (e.g. Cartesian) lattice
is chosen, artefacts in the established tumour growth pattern might arise, reflecting the lattice symmetries
[18]. While unstructured meshes are harder to implement and combine with existing PDE solvers, they
do not suffer from this shortcoming [120].

3.1.1. Cellular automata with one or several cells per lattice site

In this section, we discuss two types of CA models: those with at most one and those with multiple cells
per lattice site. Both approaches have their advantages and drawbacks. Below, we will start by discussing
the single-cell CA models and then continue to frameworks that include many cells per lattice site.

When considering single tumour cells, we follow each of these through a sequence of time-steps.
During every time-step ($t \rightarrow t + \Delta t$), a cell might migrate to a neighbouring lattice site, it might die, it
might divide or it might grow, whereby it temporarily occupies two sites before dividing. All of these
events can be modelled as stochastic processes or be imposed through deterministic conditional
statements that encompass our knowledge about cancer—additional parameters, such as age,
mutation or phenotype, might be included in these rules [121]. For instance, while cells might die
spontaneously, external factors can make cell death increasingly likely. Thus, tumours develop a
necrotic core due to nutrient deprivation. One might, therefore, set the probability for individual cell
death to depend on the distance to the tumour front [109], the local nutrient concentration or the local
level of toxic metabolites [108].

Tumour cell migration is also affected by the location of the cell within the tumour. If we consider a
cell that is deeply buried within the tumour, all surrounding lattice sites are already occupied and there is
nowhere to migrate. This notion also plays a role in cell proliferation. To divide, a cell deep within the
tumour must push its neighbours away to make space for the daughter cell, which becomes
increasingly difficult as the distance to the tumour edge increases. For simplicity, many authors
account for this notion by introducing a sharp cut-off: if there are no free sites within a certain
distance, the cell will be unable to divide [122]. Other models assume a smooth decline in
proliferation with depth [112,123]. On a macroscopic level, this suppression of cell division leads to a
proliferative rim of constant width, below which we encounter a quiescent cell population
surrounding a hypoxic region with a necrotic core at its centre. In accordance with data, this model
predicts that a tumour spheroid enters linear growth after initially growing exponentially [124]. As regards cell division, we also note that experiments suggest that the cell cycle time follows a $G$-like distribution, e.g. an Erlang distribution [125]. The rules imposed on the individual cells can account for a broad range of cell properties. For instance, if the CA model accounts for the hierarchical nature of tumour cells (GSCs, GCPs and GTPs), the probability of cell proliferation will also depend on the attributed cell type [121].

When solving the system numerically, one must choose a suitable time-step, $\Delta t$. For this purpose, it is helpful to consider the cell population as a whole. Every time a cell dies, moves, divides or grows, the system as a whole will transition into a new state. The time-step should be chosen accordingly such that only one event is likely to occur within $\Delta t$, and thus $\Delta t$ is itself a function of time. To achieve this, one might distil the temporal dynamics of the system into a master equation and employ kinetic Monte Carlo algorithms [112,126–128]. Moreover, the order in which the individual cells are updated should be random to avoid grid artefacts [17].

Rather than considering individual cells, one might track cell clusters of a limited size, keeping track of the number of cells at each lattice site. The underlying concepts stay the same. The advantage of reducing the resolution in this manner is the reduced computational cost, which allows for modelling tumours of centimetre size [125,129]. Meanwhile, a coarser resolution might lead to artefacts affecting the predicted tumour growth. Another drawback of both single- and multiple-cell CA models is the fact that they do not properly account for cell morphologies. For models that address this issue, see §3.1.3.

### 3.1.2. Lattice gas cellular automata models

Like the models in §3.1.1, LGCA models follow a set of stochastic rules to determine whether individual cells die, grow or divide. But in addition to specifying the location of the cells on the lattice, LGCA models attribute a discrete set of velocity channels to each spatial location. Hence, the cells roam a discretized phase space.

Analogously to migration in the CA models presented earlier, cells might stochastically migrate in phase space: they can move to neighbouring velocity channels following probabilistic rules. At most, one cell can occupy any given velocity channel at a given location. This restriction is not to say that these channels are unique, i.e. at each location there might be one or more channels with the same velocity, including channels at which the cell is at rest.

When using LGCA models, the time evolution of the system is split into two distinct steps. Firstly, the model allows for a number of stochastic events that correspond to the number of cells—they can die, grow, divide or change velocity channel. Afterwards, the model updates the position of every cell based on its individual velocity complying with momentum and mass conservation. In this manner, the model alternates between accounting for probabilistic interactions and deterministic cell movement. These basic concepts of LGCA models are summarized in figure 1.

Within the mean-field approximation, Hatzikirou and Deutsch [110] show that the Boltzmann equation governs the macroscopic dynamics of LGCA models. (The Boltzmann equation can be used to describe the evolution of thermodynamic systems that are out of equilibrium. Applications range from quantum mechanics over cosmology to MRI [130].) Furthermore, through Taylor expansions and scaling, they arrive at a reaction–diffusion equation with a logistic growth term, giving a microscopic justification for the success of the models presented in §2.

LGCA models offer an intuitive implementation of cell migration. By comparing model predictions to experimental data, several studies have thus used such models to gain insights into cell–cell interactions and mechanisms that underlie tumour invasion [111,131,132]. In particular, the cited articles focus on the so-called ‘Go or Grow’ (GoG) hypothesis. This hypothesis addresses the fact that glioma cells exhibit an inverse correlation between cell migration and proliferation [133]. It states that migration and proliferation are mutually exclusive events, i.e. that moving cells cannot divide. This dichotomy stems from shared signalling pathways [134]. Since the transition from benign to malignant tumour growth is coupled with a transition from a highly proliferative to a highly migrative phenotype, it is paramount to understand the link between these two phenomena. Specifically, Monteiro et al. investigate the role of hypoxia [135].

### 3.1.3. Cellular Potts models

CPMs define a Hamiltonian function, $H$, to incorporate cell movement and growth, interactions between adjacent cells, and interactions between cells and their microenvironment (see [136] for a detailed
A discussion of Hamiltonian mechanics). Moreover, CPMs attribute multiple lattice sites to each cell, which allows the models to account for cell morphology. Again, CPMs do so via the Hamiltonian function. The basic concepts are illustrated in figure 2.

For simplicity, let’s consider a Hamiltonian of the following form \[97,137\]:

\[
H = \sum_{(x,x')} E(\tau(x), \tau(x'))(1 - \delta(x, x')) + \lambda V \sum_{\sigma} (V(\sigma) - V_T)^2.
\] (3.1)
The first term on the right-hand side of equation (3.1) summarizes the contact energy at the cell interfaces. The sum runs over all pairs of adjacent lattice sites \((x, x')\). Each cell (and the ECM) has a unique identifier \(\sigma\), i.e. lattice sites with the same value of \(\sigma\) are associated with the same cell (or the ECM). Moreover, each cell has a type \(\tau\) (e.g. GSC, GCP or GTP; see [138]). \(E(\tau(\sigma(x)), \tau(\sigma(x')))\) denotes the contact energy per unit surface area (in 3D) between a cell of type \(\tau(\sigma(x))\) and a cell of type \(\tau(\sigma(x'))\). The Kronecker delta, \(\delta(\sigma(x), \sigma(x'))\), is needed since we only encounter an interface if \(\sigma(x) \neq \sigma(x')\), while there will be a contact energy associated with cells of the same type. The second term on the right-hand side of equation (3.1) represents a penalty for deviations between the volume, \(V(\sigma)\), of a cell and its target size, \(V_T\). Here, the sum runs over all cells.

At every time-step, the CPM selects a random set of lattice sites at the cell boundaries and attempts to overwrite their cell affiliations \((\sigma)\) with those of neighbouring sites: \(\sigma(x) \rightarrow \sigma(x')\). In this manner, cells grow and move by hijacking adjacent sites, whereby they might have to push other cells aside. Whether these proliferation and growth attempts are successful or not is decided by a Metropolis algorithm based on the associated change in the Hamiltonian function (see [139] for a discussion of such algorithms). The proposed change in cell affiliation is accepted with the probability

\[
P(\sigma(x) \rightarrow \sigma(x')) = \min \left(1, \exp \left(-\frac{\Delta H}{\beta} \right) \right),
\]

where the motility of the cells increases with the increasing Boltzmann temperature, \(\beta\) (cf. [137]).

One can include additional terms in equation (3.1) to add more information about cell biology. For instance, in analogy to the volume constraint, Ouchi et al. [140] suggest including constraints on the surface area of the cell (in 3D) [141]. Other authors include chemotaxis, motility, haptotaxis and haptotaxis in this manner [98,142–146]. One immediate drawback of CPMs is that they are limited to phenomena and cell properties that can be formulated as terms in a Hamiltonian.

Meanwhile, when accounted for, cell division and death are treated as stochastic events or occur when certain conditions are met. For instance, Gao et al. [138] let cells attempt division when \(V(\sigma) > 2V_T\) and impose apoptosis for proliferating glioma cells after a pre-determined number of division attempts. Other authors impose (probabilistic) conditions on apoptosis and necrosis based on nutrient levels or overcrowding [97,141].

### 3.2. Off-lattice models

One of the main drawbacks of lattice-based models is that the lattice introduces a lower length scale. Off-lattice models overcome this problem at the expense of increased computational cost. We give a schematic overview of the two different types of off-lattice models in figure 3.

#### 3.2.1. Center-based models

CBMs treat cancer cells as physical particles whose trajectories can be deduced from their equations of motion [136]. The cells themselves are usually represented as spheres (in 3D) or viscoelastic ellipsoids that deform when subjected to external forces [147]. During mitosis, the mother cells are often modelled as dumb-bells, consisting of two overlapping spherical cells [148].

The equations of motion can be derived within an energy-based picture or by modelling the relevant interactions as physical forces that act on the particles. In the latter approach, the models draw on Newton’s second law and commonly assume that inertial forces and any average acceleration can be neglected (cf. Brownian dynamics). Put another way, we are dealing with friction-dominated overdamped motion, which means that friction forces balance out all the other forces [18,122,124]. For cell \(i\) on a substrate, we can then write (cf. [18])

\[
I_{ci}^s \frac{d\mathbf{x}_i}{dt} + \sum_{j \neq i} I_{ji}^c \left( \frac{d\mathbf{x}_i}{dt} - \frac{d\mathbf{x}_j}{dt} \right) = \mathbf{F}_{i}^{\text{sub}} + \mathbf{F}_{i}^{\text{mig}} + \sum_{j} \left( \mathbf{F}_{ji}^{\text{adh}} + \mathbf{F}_{ji}^{\text{rep}} \right).
\]

The first term on the left-hand side expresses cell–substrate friction forces, while the second term deals with friction between cells. Here, \(I_{ci}^s\) and \(I_{ji}^c\) are tensors that account for anisotropies and inhomogeneities. For instance, for spherical cells in a homogeneous and isotropic environment, \(I_{ji}^s = \gamma \mathbf{I}\), where \(\gamma\) is a damping coefficient and \(\mathbf{I}\) is the identity matrix [122]. Thus, the drag force that acts on a cell in an isotropic viscous environment is proportional to the cell’s velocity [149].
On the right-hand side of equation (3.3), \( F_{\text{adh}} \) denotes forces that emerge from adhesion and repulsion between cell \( i \) and the substrate. \( F_{\text{mig}} \) describes the migration forces. For instance, chemotaxis might introduce a preferred direction for migration [122]. One might account for the memory of cells, i.e. the persistence of a cell to move in a given direction. Some authors do so using inertial forces. In addition, some authors include cell polarity, giving the cell a preferred direction of motion. \( F_{\text{mig}} \) becomes a noise term that describes a random movement in the absence of such forces and assumptions. The last term on the right-hand side summarizes the adhesion and repulsion between cell \( i \) and all other cells. These forces will typically depend upon the distance between the cell centres or the amount of overlap, \( \delta_{ji} \). Contact mechanics offers different approaches to model these cell–cell interactions, including Hertzian formulations and the Johnson–Kendall–Roberts model of elastic contact (cf. [110,148]). Some authors simply assume that all forces involved in cell-cell interactions are well approximated by linear springs, generalized linear springs or polynomial expressions (e.g. [43,149,150] and the references therein).

To exemplify the use of equation (3.3), let’s consider spherical cells, whose movement is dominated by isotropic friction on the substrate and cell-cell interactions. Meineke et al. [150] and Drasdo [122] use this model to describe cellular mono-layers in the intestinal crypt and in vitro. For mono-layers, we might assume that all cells have roughly the same velocity, which leaves us with the first term on the left-hand side and the last term on the right-hand side of equation (3.3):

\[
\gamma \frac{dx_i}{dt} = \sum_j \left( F_{\text{adh}}^i + F_{\text{rep}}^i \right),
\]

which amounts to a set of coupled ODEs that can be solved using explicit or implicit schemes, such as a forward Euler scheme [149,150]. In general, equation (3.3) can be written as a set of linear equations and be solved using matrix manipulation, for which standard software packages are available.

The computational challenge of CBMs is twofold. Firstly, all interacting cell pairs must be identified. Secondly, the relevant interactions take place on chemical and dynamic time scales that are much shorter than the time scale that governs tumour growths. Besides their high computational cost, it should be noted that CBMs might suffer from artefacts when cells are densely packed [18]. Despite these shortcomings, CBMs have widely and successfully been applied to study tumour-related issues, such as drug response [151,152].

3.2.2. Deformable cell (vertex) models

DCMs are conceptually similar to CBMs: cell dynamics are distilled into equations of motion. However, rather than representing each cell by simple geometrical objects, DCMs depict cells as a number of...
connected nodes (cf. figure 3). In contrast to CBMs, DCMs hereby capture cell morphology and intracellular mechanics. The nodes within a cell are connected by viscoelastic elements that are part of either the cell membrane or the cytoskeleton. It is essential to include a cytoskeleton since the modelled cells might buckle without an internal scaffolding structure [153].

Assuming that inertial forces can be neglected, the equation of motion for any given node takes the same form as that of a single cell in CBMs (cf. equation (3.3)). Thus, frictional forces balance out all other forces acting on the node (see also [154]). Node–node interactions can be modelled by letting the viscoelastic elements, i.e. parts of the cell membrane or the cytoskeleton, act as damped linear springs. To better depict the behaviour of biopolymers, one might model the elements using nonlinear forces [155,156]. Besides node–node interactions, we must introduce forces that express area and volume constraints on the cell as a whole or on subdivisions spanned by several nodes. By considering subdivisions of the cell membrane, we might furthermore introduce forces that express the resistance of the cortex to bend [156].

Simulating realistic cell dynamics requires a large number of nodes per cell and the models can become rather complex. Due to the resulting high computational cost, it is often only feasible to simulate single cells or small cell populations on short time scales.

For completeness, it should be mentioned that there is a subgroup of DCMs called VMs. In such models, the cells are represented as polygons spanned by a grid of nodes, i.e. adjacent cells share vertices and edges. There is no space between cells. VMs are useful when dealing with densely packed cells and have thus found applications in tissue mechanics but have not been applied to GBM [157–159].

Table 2 gives an overview of all discrete methods discussed earlier.

### 4. Hybrid multi-scale models

In the literature, the term hybrid or multi-scale model might denote any combination of the different methods discussed earlier. A large variety of approaches thus fall into this category. Adopting the nomenclature from Deisboeck et al. [19], we distinguish between three types: composite hybrid models, adaptive hybrid models and calibrated models, such as continuum models with functional parameters. However, we note that the literature generally does not explicitly distinguish between these three types.

Composite hybrid models track and connect phenomena across various scales while keeping the scale associated with each individual aspect fixed (cf. §4.1). Thus, all tumour cells are treated on an equal footing but not in the same way as their environment. Since the coupling between a tumour and its environment must be considered in real-world applications, and since the environment can be modelled using PDEs, the agent-based models discussed in §3 are consequently often referred to as components of composite hybrid models when they are encountered in the literature.

In contrast, adaptive hybrid models dynamically adjust the local resolution based on the required level of detail (cf. §4.2). That is to say that not all tumour cells are represented in the same manner. Finally, the third type of hybrid model denotes, for instance, continuum models, whose parameters have been calibrated based on agent-based models, or biophysical considerations of microscopic or mesoscopic phenomena (cf. §4.3).

#### 4.1. Composite hybrid models

When using a discrete model to represent tumour cells, the extracellular biochemical players, such as oxygen, are commonly modelled as continuous fluids [91,164–166]. Moreover, intracellular changes, i.e. processes on a sub-cellular scale, might be tracked for each individual cell using ODEs as discussed in §4.1.1. These different microscopic, mesoscopic and macroscopic mechanisms are all coupled. For instance, the nutrient levels both depend on the cell density and impact the probabilities that are used to predict cell actions. When this interdependence is taken into account, the discrete tumour cell model becomes one of many components in a larger framework that spans and links multiple temporal and spatial scales. The framework as a whole then constitutes a composite hybrid model (cf. [20,65,89,151,167–174]). Such hybrid frameworks have been used to study a wide variety of phenomena, including drug resistance [152,175]. In each case, the focus of the study prescribes the resolution of different features. For instance, in some studies, discrete models of the vasculature are combined with continuum models of the tumour to better understand angiogenesis or drug delivery [101]. As another example, May et al. [176] couple discrete tumour models with biomechanical
Table 2. Overview of the reviewed discrete models. The references are not exhaustive. For other recent reviews, see [12,13,18,160,161].

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Features</th>
<th>References</th>
</tr>
</thead>
</table>
| **CA model** | **Concept:** All CA-type models are discrete and lattice based. Cell migration, death, movement, division and growth are modelled as stochastic processes.  
**Particularity:** Multiple cells or at most one are allowed per lattice site. | One cell per lattice site: [112,121,125,162].  
Multiple cells per lattice site: [107,109].  
**Bridging scales:** The limit on the number of cells per site (i.e. the resolution) can vary between different locations [18]. |
| **Assets:** Centimetre-scale simulations are feasible when tracking cell clusters. Individual cells can be represented.  
**Drawbacks:** They do not represent cell morphology, have a limited spatial resolution and are computationally heavy. For multiple cells per site, the individual cell positions are not tracked. | Deutsch, Hatzikirou and collaborators have investigated cell proliferation and glioma invasion [110,111,131,132]. |
| **LGCA model** | **Concept:** They are CA-type models. There are multiple velocity channels (and hence cells) per lattice site. A deterministic step evaluating movement is included (figure 1).  
**Assets:** Individual cells are represented. Convergence to CM has been shown.  
**Drawbacks:** They do not represent cell morphology, have a limited spatial resolution and are computationally heavy. | Deutsch, Hatzikirou and collaborators have investigated cell proliferation and glioma invasion [110,111,131,132].  
Authors have included a broad range of phenomena through a variety of terms in the Hamiltonian [97,98,138,140,143]. |
| **CPM** | **Concept:** They are CA-type models. There are multiple sites per cell. A Hamiltonian function governs stochastic events.  
**Assets:** CPMs draw on the physics behind the cell dynamics and represent cell morphology.  
**Drawbacks:** They have a limited spatial resolution. Centimetre-scale simulations are not feasible. | Cells can be spherical [110,148] or ellipsoidal [147]. Studies of tumour cell drug response: [152]. |
| **CBM** | **Concept:** CBMs model cells as physical particles governed by their equations of motion. No lattice is used. Cells are simple geometrical objects.  
**Assets:** The physics behind the cell dynamics is accounted for. Cells and their movement are resolved.  
**Drawbacks:** CBMs are computationally heavy. Centimetre-scale simulations are not feasible. Cell morphology is not modelled. | Single-cell adhesion and response to mechanical stress [155,156], tumour cell proliferation [163], vertex models [158]. |
| **DCM** | **Concept:** Cells are depicted as a collection of connected nodes described by their equations of motion.  
**Assets:** DCMs model cell morphology and stresses on a sub-cellular level. There is no lattice that would impose a limited spatial resolution.  
**Drawbacks:** Only simulations of small cell populations or single cells are feasible. | |
calculations of stresses and strains to arrive at a more accurate prediction of the tumour shape. Finally, we note that composite hybrid models have been employed in Bayesian inferences to establish posterior probability distributions for relevant model parameters [123] (see also §5).

4.1. Ordinary differential equations

When dealing with agent-based hybrid models, some authors include sub-cellular dynamics, modelling molecular networks, pathways and reactions [167,168,172,177]. Such processes are governed by a system of coupled nonlinear ODEs that represent mass balance equations and rely on the stoichiometry of the underlying biochemical reactions. If we consider \( N \) molecular species, the rate at which the concentration \( X_i \) of the \( i \)th species changes is thus a function \( f_i(X_1, \ldots, X_N) \) of the concentrations of the \( N \) different species such that we have a set of coupled differential equations of the following form:

\[
\frac{dX_i}{dt} = f_i(X_1, \ldots, X_N).
\] (4.1)

The right-hand side of equation (4.1) depends on the system in question but typically contains linear terms, quadratic terms, Michaelis–Menten terms or other nonlinear functions such as Hill function terms (see also [178] for an introduction to Michaelis–Menten kinetics). It is beyond the scope of the present paper to go into any further details. Rather, we kindly refer the reader to the references listed at the beginning of this section.

ODEs thus play a vital role in some composite hybrid models. In general, ODEs are a common tool in mathematical biology and oncology. For instance, they can capture global tumour growth patterns as briefly mentioned in §2.

4.2. Adaptive hybrid models

Detailed modelling of individual cells is required to capture tumour growth at low cell numbers. Agent-based models thus give unique insights into glioma invasion. However, at high cell densities, a lower resolution suffices. Continuum models realistically capture all relevant aspects of large chunks of bulk tumours. Adaptive hybrid models take advantage of this notion by describing different parts of the tumour tissue using distinct modelling approaches. Thereby, it becomes possible to draw on the strengths of discrete models when depicting centimetre-sized tumours without the insurmountable computational cost that would otherwise be involved.

The hybrid model presented by the groups of Kim and Stolarska [172,179,180] treats necrotic tumour zones, quiescent tumour tissue and the surrounding microenvironment as continuous fluids. The relevant PDEs are solved on a regular grid. The cells in the proliferative tumour rim (100–200 \( \mu \)m), on the other hand, are modelled using a CBM. They are represented as autonomous ellipsoids. At the boundary between the cell-based and continuous components, interpolation is used to communicate the forces that individual agents exert on the continuous component and vice versa (see also [181] for a related example from molecular dynamics).

Bearer et al. [182] and Frieboes et al. [183] couple an off-lattice approach with the continuum models by Wise et al. [74] discussed in §2. To study glioma invasion, they include dynamic transitions between the continuum and cell-based representations based on the microenvironment. In their model, both mutations and hypoxia might induce the development of a migratory phenotype [184].

The adaptive hybrid models discussed earlier allow for mass transfer between the different components and are carefully constructed to obey laws, such as mass conservation. However, it is worth noting that the continuum models are pre-assumed without guaranteeing that the imposed functional form emerges from up-scaling the cell-based representation (cf. §4.3). Moreover, due to the assumptions that enter the continuum and agent-based models, single-cell measurements might lead to different parameter values than those required for continuum models to fit patient-specific data [185].

Rather than coupling agent-based and continuum models, other authors [18,186,187] combine agent-based models with different resolutions. Such approaches are known as multi-scale agent-based modelling. For instance, one might couple CA models that consider single cells with models that deal with cell clusters at a coarser lattice resolution (see also [188] for a multi-resolution study of polymers).

It is an intricate task to couple models that represent the tumour on different scales. After all, the models might rely on very different paradigms—e.g. one might seek to create an interface between a stochastic and a deterministic model. As a result, a variety of approaches can be found in the literature. For a recent review, we refer the reader to the article by Smith and Yates [189]. But as a
final example, it is worth mentioning the article by Yates et al. [190] that demonstrates how to construct a transition region where the two models in question coexist with spatially varying contributions to the predicted dynamics.

4.3. Calibrated models

Various authors encode information from cell-based models or biophysical considerations into heterogeneous and time-varying parameters of continuum models. This goal can be achieved in different ways. As briefly discussed in §3.1.2, one might up-scale agent-based models using the mean-field approximation (cf. [110]). Other authors draw on the emergent macroscopic properties of agent-based models to derive phenomenological relationships (cf. [17,114]).

Each method mentioned earlier relies on simplifying approximations to derive an analytical expression, which might limit their predictive power. To circumvent the need to explicitly state expressions for the macroscopic variables altogether, Kavousanakis et al. [191] use a form of equation-free modelling called coarse projective integration [192]. This method invokes the so-called coarse time-stepper that consists of four phases: lifting, simulation, restriction and extrapolation. These four phases are repeated for every macroscopic time-step. In the lifting step, an ensemble of cell distributions is drawn from the macroscopic tumour cell density. Each cell distribution is then evolved over a short time span using an agent-based model—Kavousanakis et al. use a CA model. In the restriction step, the macroscopic density is updated by computing the ensemble average of the agent-based simulations. Under the assumption that the macroscopic cell density evolves more slowly than the microscopic variables, the observed change in the cell density yields an estimate of its time derivative. Using the forward Euler method, this information can be employed to project the cell density further into the future without considering individual cells.

We note that the general idea of calibrating coarse models based on more resolved representations is not only limited to continuum models. Recently, Van Liedekerke et al. [193] have thus calibrated the mechanical interaction forces of CBMs based on DCMs.

5. Data-driven modelling

Researchers draw on a wide range of experimental data to inform their mathematical models of GBM and do so following several different approaches. One can use data to select the underlying mechanism or inform the model parameters. For example, as discussed earlier, some researchers use DTI MRI data to determine the diffusion properties of cancer cells. Moreover, data allow for qualitative comparisons and constraints on the simulations. For instance, most models are set to recover generic properties of cancer, such as spherical avascular tumour growth. Finally, one can use data to infer model parameter values using a statistical framework, either maximum likelihood estimation or a Bayesian approach. Systematic incorporation of imaging and molecular data into the mathematical models of cancer via statistical inference is a promising and timely area of future development. This area of development is bolstered by the latest progress in imaging and omics approaches, which are producing quantitative data on an unprecedented scale. Furthermore, the latest progress in computational power, simulation-based inference methods and machine learning make this task computationally feasible. Several modelling studies aim to fit specific data from patients, animal models or in vitro experiments. Here, we provide a list of illustrative examples and summarize our perspective on important areas of future research. We also refer the reader to some recent reviews on the topic [14,194].

Two decades ago, Stamatakos et al. [195,196] used agent-based models to investigate the impact of chemotherapy in vivo based on patient-specific PET, SPECT, T1-weighted MRI, histopathologic and genetic data. They reduced the computational cost by clustering the cells into macroscopic (1 mm²) regions rather than simulating individual cells. More recently, Gallaher et al. [151] published a study in which they fit MRI and ex vivo cell-tracking data from rats to off-lattice agent-based models, simulating individual cells. They accomplish this fit by using random sampling because many fitting algorithms struggle with the stochastic nature of the agent-based models [197,198]. However, due to the high computational cost, agent-based models are most commonly found in studies focusing on small cell populations. For instance, Oraiopoulou et al. [199] use agent-based models to recreate the invasive morphologies of GBM observed in vivo. Meanwhile, many articles on agent-based models do not include a fit to real-world data but rather perform in silico experiments to make qualitative and quantitative predictions. For instance, on the basis of such analyses, Perez-Velazquez and Rejniak
The parameters used in state-of-the-art continuum models [185] to capture the migration and proliferation of GBM are derived from MRI and CT images, and the application of continuum models to patient-specific data in vivo is widespread in the literature. Thus, Swanson et al. [27,29] already used such data to inform their models 20 years ago [200]. Later studies that build on their work have included further imaging data, such as PET [30,32,201]. Isotropic diffusion reaction equations yield a detailed picture based on T1Gd and T2-FLAIR MRI images [31]. Other studies have employed histopathology [25]. The anisotropic continuum models discussed in §2.1 likewise built on MRI data, drawing on DTI images to compute the diffusion matrix [26,47,63,68,202].

Due to the relatively low computational cost of continuum models, they can be employed in Bayesian sampling schemes as illustrated by Lipkova et al. [203], who infer model parameters based on MRI and PET images using a Markov Chain Monte Carlo algorithm. Based on a clinical study conducted on eight patients, Lipkova et al. demonstrate that their Bayesian approach can be used to design personalized radiotherapy plans that harm less healthy tissue than the standard treatment protocol while achieving the same efficiency in terms of tumour recurrence within a predefined volume.

Ezhov et al. [204] address the same inverse problem using machine learning in tandem with continuum models. Ezhov et al. construct a library of 100 000 simulations with different parameter values and use their atlas to train a neural network. The trained neural network can infer the patient-specific model parameters based on MRI images within minutes, allowing for effortless model personalization. Today, continuum models are thus at the point where they have the potential to enter clinical settings and contribute to personalized treatments. While the same does not yet hold for agent-based models due to their high computational cost, machine learning might likewise help to overcome this obstacle in the future, as discussed by Jørgensen et al. [198]. Machine-learning algorithms would play the same role as they do in continuum models by circumventing the need to produce further simulations once a surrogate model or inference algorithm has been trained. Similar strategies are applied when dealing with agent-based models outside of cancer research [205].

Like imaging data, molecular data offer unique insights into GBM [206,207]. In particular, different sequencing techniques, including bulk RNA-seq [208], scRNA-seq [209–214] (recently reviewed in [215,216]) and spatial RNA-seq [217], have helped researchers to uncover different aspects of brain tumours. Sequencing data have thus contributed to our understanding of important cell types, the nature of the invasive tumour front, the spatial cellular organization of the tumours and the molecular basis of the heterogeneity of GBM. For instance, Ravi et al. [217] recently identified spatially distinct clusters of genes via spatial transcriptomics, and Neftel et al. [209] identified four main cellular states among malignant tumour cells by analysing data at the single-cell level. Furthermore, Neftel et al. demonstrated that the relative frequency of these four states varies between GBM samples. On the one hand, such information on the heterogeneity of GBM is valuable in its own right and can help to inform the underlying assumptions used in the mathematical models. As exemplified in §2.2, mathematical models thus present a unique tool to investigate the implications of such information on the tumour phenotype. On the other hand, the vast number of spatial GBM sequencing data that is becoming available can directly be used in parameter inference, model selection or validation of model predictions. Thus, mathematical models might be applied to fit the cell-type distributions that spatial GBM sequencing data uncover.

6. Clinical application of spatio-temporal modelling in GBM

GBM is a biologically complex and dynamic tumour that exists within the intricate and responsive brain environment. Parcelling and modelling discreet aspects of these biological processes and systems holds the potential for numerous direct clinical care applications and has the potential to address several areas of unmet need.

Mathematical diffusion tensor modelling is currently widely used in the context of MRI diffusion tensor-based tractography for the modelling of white matter tracts. Understanding patient-specific anatomy, and anatomo-pathological relationships, is critical for effective surgical planning and avoidance of iatrogenic injury [218]. Future refinements in data modelling can be expected to improve spatial resolution and reliability, which would directly translate to improved validity of neuro-imaging with direct clinical applications—as exemplified in the study by Lipkova et al. discussed in §5.
A key concern in clinical practice is the prediction of tumour growth patterns, sites of progression and the location of tumour recurrence. This triad is important for both therapeutic planning and prognostication [219]. Mathematical models have the potential to refine these facets of care and hence to offer highly tailored conformal radiation dosing and targeted surgical supra-marginal resection [220,221]. These therapeutic adaptations might maximize the prognostic benefits and lead to the destruction of the tumour while sparing normal neural tissue. Similarly, when combined with the modelling of growth rates, this information could be assimilated to generate more accurate prognostic maps with the prediction of anatomico-pathological functional deficits and overall prognosis.

Recent years have been characterized by increasing recognition of GBM’s molecular intra- and inter-tumoural heterogeneity and its functional consequences. This is exemplified in the evolution of the WHO classification of tumours from a purely histopathological to a fully integrated molecular classification [222]. As we have summarized, mathematical models of GBM hold the potential to illustrate novel prognostic aspects of tumour biology and the microenvironment not currently captured. These might include structural metrics, such as tissue density and cellular lineage, or functional ones, including oxygenation, nutrient supply and metabolism.

The potential for applying spatio-temporal mathematical modelling of GBM in clinical care is substantial but thus far remains largely unrealized. As summarized in this review, the foundation of knowledge in this area is now considerable. The stage is set for those who wish to advance the field from theoretical models to practical applications.

### 7. Summary and discussion

This article presents an overview of mathematical models that provide means to scrutinize and predict the spatial and temporal evolution of tumours. While we focus on models for GBM and highlight characteristics of this aggressive cancer type, we also attempt to give a rounded picture of state-of-the-art approaches across the field of mathematical oncology.

Spatially resolved tumour models fall into two main categories: they either treat the tumour cell density as a continuous variable (cf. §2), or deal with autonomous agents representing individual cells or cell clusters (cf. §3).

The computational cost associated with continuous models is relatively low. This property allows for the modelling of the entire tumour volume and the application to a wide variety of (patient-specific) data. Furthermore, continuous models can capture many of the characteristics of GBM. We summarize how continuous models achieve this goal in table 3. GBM spreads anisotropically following existing structures in the brain, including white matter tracks and blood vessels. On a macroscopic level, this behaviour is well described by reaction–diffusion equations when including an anisotropic diffusion term (cf. §2.1).
Moreover, by including source terms that couple a set of such reaction–diffusion equations, continuous models can link different cell sub-populations and account for interactions with the environment (cf. §2.3). Continuous models are hereby able to encapsulate the hierarchical nature that is a defining feature of GBM (cf. §2.2). The models also show great versatility: by including additional biologically motivated terms into the reaction–diffusion equations, continuous models can be painlessly extended to include phenomena ranging from immune responses to the impact of radiotherapy.

Tumour cells must always be understood in relation to their surroundings. In this connection, aspects such as nutrition flows, the ECM or the vasculature are often (if not always) represented using continuous models. Even when using agent-based models, continuous models thus play a vital role. However, continuous models have their drawbacks. They do not account for the stochastic events that dominate at low tumour cell densities. Consequently, they are not ideal for addressing some questions regarding the invasive tumour front or tumour recurrence that is a prevalent obstacle in treating GBM.

At low cell densities, agent-based models have an edge. Such models come in many different flavours. Some agent-based models borrow from the description of other phenomena. They thus lean on theoretical models, e.g. from solid-state physics, posing biological events in terms of physical forces and potentials (cf. §§3.1.3, 3.2.1 and 3.2.2). Other models rely on purely stochastic rules to encode biological information (cf. §§3.1.1 and 3.1.2). Moreover, agent-based models differ in the details that they encompass. While some models address aspects of cell morphology, others do not. While some models define a lower threshold for their spatial resolution by employing a lattice on which the agents move, others do not.

Like continuous models, agent-based models are able to capture the essential properties of GBM (table 3). Anisotropic diffusion along existing structures can be included by altering the rules that govern cell migration based on the environment. Moreover, by altering these rules based on the cell type, the hierarchical organization of GBM can be incorporated into the model.

The high computational expense of agent-based models constitutes their main drawback. It can be a significant stumbling block on the way to unleashing their full potential. However, this shortcoming can be partly circumvented. One might thus combine different approaches into a hybrid model bridging across the scales involved (cf. §§4.1 and 4.2). Alternatively, one might study small cell populations to derive emergent macroscopic properties that can be used to calibrate models on a coarser scale (cf. §4.3). Indeed, one of the main assets of agent-based models is their ability to link an understanding of individual cells to macroscopic events. To further lower the computational cost when confronting the models with data, one might use sophisticated statistical techniques or machine-learning methods that demonstrably reduce the number of models that need to be constructed [198,223], as discussed in §5.

Each of the available mathematical models has its strengths and weaknesses. Which model to employ will hence be dictated by the research question and the available data. Firstly, some approaches are better suited for depicting certain phenomena than other methods are. One might drastically simplify the work associated with the practical implementation by choosing the method wisely. Secondly, not all models are equally informative. The question is whether the additional information can be exploited and for what purpose [224]. On the one hand, the chosen model sets limits on the insights that one might gain. This notion favours more complex models. On the other hand, one should avoid cracking nuts with a sledgehammer. To exemplify this, it is thus worth noting that non-spatial models make full use of certain types of (patient-specific) data while being unable to address any questions on tumour morphology. They are undoubtedly very valuable tools in their own right. The same holds for each spatio-temporal cancer model presented earlier. With this in mind, we hope that our review provides a useful guide and helps the reader select a suitable mathematical model for their research.

This article aims to give an extensive overview of state-of-the-art modelling techniques. However, we left out some ideas and concepts that go beyond the scope of this review. For instance, one could use agent-based models incorporating evolutionary dynamics [225] or even evolutionary game theory [226].

Mathematical models are only useful if they describe and relate to experimental data. Our review thus ends with a summary and discussion of the application of mathematical models to experimental data and its clinical relevance. A wealth of imaging and molecular data is becoming available, particularly due to advances in single-cell and spatial transcriptomics. At the same time, researchers have begun to implement spatial models into statistical frameworks to perform systematic inference based on these data. While few such inference analyses have yet been published, advances in machine learning and computing power promise to further enable the systematic incorporation of imaging and molecular data into the spatial models of GBM. We thus envision a greater impact of mathematical modelling in unravelling the basic biology of GBM and its applications in clinical treatment in the near future.
Data accessibility. This article has no additional data.

Authors' contributions. A.C.S.J.: conceptualization, investigation, project administration, visualization, writing—original draft and writing—review and editing; C.S.H.: investigation, project administration and writing—review and editing; M.S.: investigation, project administration and writing—review and editing; W.T.: investigation and writing—review and editing; S.R.K.: writing—review and editing; D.G.: writing—review and editing; M.F.L.: funding acquisition, project administration and writing—review and editing; S.P.: funding acquisition, project administration and writing—review and editing; S.M.: funding acquisition, project administration and writing—review and editing; V.S.: conceptualization, investigation, project administration, supervision and writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by The Oli Hilsdon Foundation through The Brain Tumour Charity (grant no. GN-000595), in connection with the program ‘Mapping the Spatio-temporal Heterogeneity of Glioblastoma Invasion’. Work at the Cancer Institute Genomics Translational Technology Platform was supported by the Cancer Research UK (CRUK) City of London Centre Award (C7893/A26233). C.S.H. was supported by a CRUK Pioneer Award (C70S68/A29787), an AMS Starter Grant (SGLO21/1034), a National Brain Appeal Innovation Award (NBA/NSG/BTB) and UCLH BRC NIHR funding. S.R.K. was supported by The Edinburgh-UCL CRUK Brain Tumour Centre of Excellence award (C7893/A27590). M.F.L. was supported by the Rosetrees Trust and the John Black Charitable Foundation (grant no. A2200). S.P. was supported by CRUK (awards C55501/A21203 and 755084).

References


