Spatial cumulant models enable spatially informed treatment strategies and analysis of local interactions in cancer systems

Sara Hamis[⊛1,2], Panu Somervuo[3], J. Arvid Ågren[4,5], Dagim Shiferaw Tadele[5,6], Juha Kesseli[7], Jacob G. Scott[5,8,9], Matti Nykter[7,10], Philip Gerlee[11,12], Dmitri Finkelshtein[13], Otso Ovaskainen[2,3,14].

1: Tampere Institute for Advanced Study, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
2: Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland.
3: Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.
4: Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden.
5: Department of Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, Ohio, USA.
6: Oslo University Hospital, Ullevål, Department for Medical Genetics, Oslo, Norway.
7: Prostate Cancer Research Center, Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Centre, Tampere, Finland.
8: Case Western Reserve School of Medicine, Cleveland, Ohio, USA.
9: Department of Radiation Oncology, Cleveland Clinic, Cleveland, Ohio, USA.
10: Foundation for the Finnish Cancer Institute, Helsinki, Finland.
11: Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden.
12: Mathematical Sciences, University of Gothenburg, Gothenburg, Sweden.
13: Department of Mathematics, Faculty of Science and Engineering, Swansea University, Swansea, UK.
14: Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Norway.

⊛ Corresponding author: sara.hamis@tuni.fi.

Author ORCIDs: SH: 0000-0002-1105-8078, **PS:** 0000-0003-3121-4047, **JAÅ:** 0000-0003-3619-556X, **DST:** 0000-0001-8319-678X, **JK:** N/A, **JGS:** 0000-0003-2971-7673, **MN:** 0000-0001-6956-2843, **PG:** 0000-0001-8503-0177, **DF:** 0000-0001-7136-9399, **OO:** 0000-0001-9750-4421.

Abstract

Theoretical and applied cancer studies that use individual-based models (IBMs) have been limited by the lack of a mathematical formulation that enables rigorous analysis of these models. However, spatial cumulant models (SCMs), which have arisen from theoretical ecology, describe population dynamics generated by a specific family of IBMs, namely spatio-temporal point processes (STPPs). SCMs are spatially resolved population models formulated by a system of differential equations that approximate the dynamics of two STPP-generated summary statistics: first-order spatial cumulants (densities), and second-order spatial cumulants (spatial covariances).

We exemplify how SCMs can be used in mathematical oncology by modelling theoretical cancer cell populations comprising interacting growth factor-producing and non-producing cells. To formulate model equations, we use computational tools that enable the generation of STPPs, SCMs and mean-field population models (MFPMs) from user-defined model descriptions (Cornell et al., 2019). To calculate and compare STPP, SCM and MFPM-generated summary statistics, we develop an application-agnostic computational pipeline. Our results demonstrate that SCMs can capture STPP-generated population density dynamics, even when MFPMs fail to do so. From both MFPM and SCM equations, we derive treatment-induced death rates required to achieve non-growing cell populations. When testing these treatment strategies in STPP-generated cell populations, our results demonstrate that SCM-informed strategies outperform MFPM-informed strategies in terms of inhibiting population growths. We thus demonstrate that SCMs provide a new framework in which to study cell-cell interactions, and can be used to describe and perturb STPP-generated cell population dynamics. We, therefore, argue that SCMs can be used to increase IBMs' applicability in cancer research.

Keywords: Individual-based models, spatio-temporal point processes, spatial moments, cancer eco-evolution, mathematical oncology.

Statements and Declarations: The authors have no competing interests to declare that are relevant to the content of this article.

Classification codes: 92Cxx: Biology and other natural sciences - Physiological, cellular and medical topics. **92Dxx:** Biology and other natural sciences - Genetics and population dynamics.

1 Introduction

All biological systems are composed of interacting parts. In cancer systems, cells can interact with each other by, for example, exchanging signalling molecules and competing for resources such as space and nutrients. Such cell-cell interactions have been identified as factors that drive eco-evolutionary dynamics of cancer cell populations (Merlo et al., 2006; Reynolds et al., 2020). Consequently, these interactions have been proposed as cancer treatment targets, where the general premise is that treatments can perturb cell-cell interactions and, by extension, disease trajectories. Such perturbations can lead to reduced tumour growths and disease burdens (Dominiak et al., 2020; Brücher and Jamall, 2014; West and Newton, 2019). Cell-cell interactions, and the impact that they have on cancer system dynamics, have been mathematically studied by researchers using a variety of mathematical models including evolutionary game theory (EGT) models (Kaznatcheev et al., 2019; Archetti et al., 2015; Farrokhian et al., 2022), other ordinary differential equation (ODE) models (Hillen et al., 2013; Poleszczuk and Enderling, 2018; Jenner et al., 2018), partial differential equation (PDE) models (Haridas et al., 2017; Lorenzi et al., 2017; Hinow et al., 2009), and individual-based models (IMBs) (Stichel et al., 2017; Campenni et al., 2020; Hamis et al., 2021).

In IBMs, cells can be modelled as individuals that evolve on a spatial domain and partake in interactions with other cells in their neighbourhood (Metzcar et al., 2019; Chamseddine and Rejniak, 2020). IBMs are naturally able to incorporate local cell-cell interactions, spatial population structures and stochastic variations between cells and biological events. However, theoretical and experimental studies that use IBMs have been limited by the lack of a mathematical formulation that enables rigorous analyses of these models. On the other hand, in mean-field population models (MFPMs), such as ODE and EGT models, the density dynamics of cancer cell subpopulations can be described by a set of mathematically tractable ODEs after imposing a set of simplifying modelling assumptions. In such models, cells are not considered as individuals but, instead, they are collectively viewed as part of a population or subpopulation. With their analytical and mathematically tractable formulation, MFPMs have been used to quantify cell-cell interactions in *in vitro* experiments (Kaznatcheev et al., 2019), and are currently being used to inform personalised treatment strategies in a clinical prostate cancer trial (Zhang et al., 2017) (see also matters arising in response to the reporting of this study (Mistry, 2021)). However, MFPMs notably assume that the modelled system is "well-mixed" so that each cell interacts with all other cells in the system with equal probability. Therefore, MFPMs can not faithfully capture the dynamics of spatially structured systems with localised cell-cell interactions, such as solid tumours (Waclaw et al., 2015). In fact, previous mathematical modelling work has shown that imposing spatial constraints on interactions between individuals, via IBMs, often results in system dynamics that vastly contradict that simulated by MFPMs (Durrett and Levin, 1994).

Due to the pre-clinical and clinical applications of mathematical oncology models (Rockne et al., 2019), there exists a need to formulate models that capture localised cell-cell interactions, maintain cell discreteness, and are both spatio-temporally resolved and mathematically tractable. In this article, we describe an approach to achieve this by using spatial cumulant models (SCMs) which, to our knowledge, have not previously been applied to study cancer cell systems. SCMs are spatially resolved population models that are translated from a specific family of IBMs, namely spatio-temporal point processes (STPPs). Following a mathematical manipulation that involves a perturbation expansion around mean-field equations in the limit of long-ranged interactions (as explained in Section 2), SCMs approximate two STPP-generated summary statistics: first-order spatial cumulants (densities) and second-order spatial cumulants (spatial covariances). If we let $u_i^{(1)}(t)$ denote the density of subpopulation i, and $u_{ij}^{(2)}(t,r)$ denote the spatial covariance of individuals (here cells) from subpopulations i and j, then SCMs dictate that

$$\underbrace{u_i^{(1)}(t)}_{density} = \underbrace{q_i(t)}_{\substack{mean\text{-}field\\density}} + \epsilon^d \underbrace{p_i(t)}_{\substack{correction\\density}} + o(\epsilon^d), \tag{1a}$$

$$\underbrace{u_{ij}^{(2)}(t,r)}_{spatial\ covariance} = \epsilon^d \underbrace{g_{ij}(t,\epsilon r)}_{\substack{leading\ term\ of\ the\\spatial\ covariance}} + o(\epsilon^d), \tag{1b}$$

where t is the independent time variable and r denotes pair-wise distances between individuals. In Eq.1a, $q_i(t)$ is the density as computed by MFPMs and $p_i(t)$ is a correction term that depends on (1) the MFPM densities $q_i(t)$, (2) the leading spatial covariance terms $g_{ij}(t,\epsilon r)$, and (3) all corrections $p_i(t)$, where $i,j = 1,..,S$, and S denotes the number of subpopulations in the modelled system. Thus the correction terms account for the spatial structure of the population via their dependence on $g_{ij}(t,\epsilon r)$.

The modelled individuals inhabit a d-dimensional spatial domain, and the scaling factor $\epsilon = 1/\ell$, $\ell > 0$, results from the perturbation expansion, in which ℓ is an interaction length scale. Thus, SCM densities reduce to MFPM densities when $\ell \to \infty$. The terms $o(\epsilon^d)$ can be omitted in practical applications as they go to zero faster than does ϵ^d.

Since Eqs.1a,b are formulated in the limit of long-ranged interactions it follows that, in the context of cancer research, SCMs are appropriate to use when studying cell-cell interactions that extend beyond only immediately neighbouring cells. Such interactions include communication via diffusible substrates and resource competition. Formulating the right-hand sides of Eqs.1a,b for a specific biological problem can be a mathematically cumbersome task. However, Cornell *et al.* (2019) developed a mathematical framework and a computational software that enables the generation of rate equations for the quantities $q_i(t)$, $p_i(t)$ and $g_{ij}(t, \epsilon r)$ for a large family of biological processes. More specifically, the software enables the generation of (1) STPPs and (2) MFPM and SCM rate equations for any user-defined biological system that can be described by one or more reactant-catalyst-product (RCP) processes. An RCP-processes is formulated by a set of reactants (points that are consumed in the process), catalysts (points that catalyse the process without being consumed in it), and products (points that are produced in the process). In this study, these points represent the center of mass of cells. Cornell *et al.*'s framework thus unifies a class of IBMs, *i.e.,* STPPs, with a class of analytical and spatially structured population models, *i.e.,* SCMs. Accordingly, we refer to their software as the `Unified Framework (UF) Software` throughout this article.

SCMs were first presented as a rigorous mathematical framework by Ovaskainen *et al.* (2014). The models build upon earlier studies of spatial moment equations in theoretical ecology (Ovaskainen and Cornell, 2006; Bolker and Pacala, 1997; Law et al., 2003) and probability theory (Finkelshtein et al., 2009; Kondratiev et al., 2008), and on probability theorists' work on STPPs, also known as "Markov evolutions in the space of locally finite configurations", (Kondratiev and Skorokhod, 2011; Finkelshtein et al., 2011). Spatial cumulants are closely related to spatial moments (Section 2.4), and we remark that spatial moment techniques have been used to derive analytical, deterministic approximations for cell population dynamics generated by different types of IBMs. These include on-lattice models in which cells are restricted to inhabit pre-defined lattice points, that can be occupied by one or zero cells (Baker and Simpson, 2010; Markham et al., 2015) and STPPs (Binny et al., 2016; Surendran et al., 2018). On-lattice IBMs explicitly account for volume exclusion, as opposed to STPPs which do not. STPPs instead generate points that can represent the center of mass of deformable cells that can move in continuous space and interact with their environment via probabilistic kernels. This brief discussion exemplifies that all IBMs are associated with different caveats and benefits. Therefore, a bio-mathematical modeller's choice of IBM and, by extension, options for describing IBM dynamics analytically, should be considered on an application-specific basis. The aforementioned cell population studies that used spatial moment techniques all challenged the view that there exists a dichotomy between individual-based and analytical models. We conjecture that one of the reasons that such techniques are not more widely used in cell biology research is that the relevant model equations have only been derived for specific cell processes, as opposed to general processes. Accordingly, a modeller wanting to use spatial moment techniques for describing the dynamics of cell populations, in which specific cell actions and interactions occur, might face a mathematically technical and time-consuming task of deriving spatial moment (or cumulant) equations with appropriate closure and approximation methods. The `UF-Software` tackles this issue of lack-of-generality by allowing a modeller to formulate spatial cumulant equations for any RCP-process. Moreover, the pipeline codes developed in this study enable a modeller to easily implement and numerically solve SCM equations. These pipeline codes also enable straightforward comparisons between STPP, SCM and MFPM-generated population dynamics.

The remainder of this article is structured as follows: In Section 2, we provide an introductory description of the mathematical theory that underlies SCMs and the `UF-Software`. In Section 3.1, we propose a practical step-by-step pipeline for using the `UF-Software` to implement STPPs, MFMPs and SCMs. Thereafter, as proof-of-concept, we exemplify how SCMs and the `UF-Software` can be used in mathematical cancer research by modelling dynamic, spatially structured cancer cell populations in which cells interact via diffusible substrates and are subjected to drugs (Sections 3.2-4.2). We demonstrate that SCMs can be used to both describe and perturb density dynamics of simulated cell populations generated by STPPs.

2 Mathematical theory of spatial cumulant models

Sections 2.1-2.10 are intended to comprise an introductory and conceptually self-contained background to SCMs. For more rigorous details of this mathematical theory, we refer readers to previous works (Ovaskainen and Cornell, 2006; Ovaskainen et al., 2014; Cornell et al., 2019). For ease of presentation we, throughout this article, consider spatial domains that are two-dimensional and have periodic boundary conditions. Moreover, we consider systems that are translationally invariant, meaning that pair-wise distances between cells influence system dynamics, whereas absolute positions of cells do not. This is equivalent to modelling a space-homogeneous system that lacks background variations, such as a typical *in vitro* system. Note, however, that the `UF-Software` can be modified to handle one, two, and three-dimensional spatial domains, as well as space-heterogeneous systems (Cornell et al., 2019).

2.1 Spatio-temporal point patterns and spatio-temporal point processes

Consider a cell population that evolves in time and space. A snapshot of the population can be represented by a point pattern in which the center of mass of the nth cell is represented by a point $x_n \in D$, where $D \subset \mathbb{R}^d$ denotes the spatial domain that the population inhabits. In this article, we set D to be a plane spanning two spatial dimensions, so that $d = 2$ and $x_n = (x_n^1, x_n^2)$. A point pattern comprising N_t points at time t can thus be described by a set of points η_t in continuous space such that

$$\eta_t = \{x_1, x_2, ..., x_{N_t}\}. \tag{2}$$

In order to categorise the individuals in the population, each point in η_t can further be coupled with a descriptive mark. Here, we let the mark m_n denote the subpopulation to which the individual in position x_n belongs so that m_n can take the value $1, 2, ..., S$, where S denotes the number of subpopulations in the modelled system. When marks are included, a point pattern describing a population snapshot can thus be described by a set of marked points

$$\hat{\eta}_t = \{(x_1, m_1), (x_2, m_2), ..., (x_{N_t}, m_{N_t})\}. \tag{3}$$

Two summary statistics that can be used to describe point patterns are densities and spatial covariances. Instructions on how to calculate these quantities from static point patterns are provided in Section 2.9. A point pattern that evolves in time and space is called a spatio-temporal point pattern, and an STPP is a mathematical model that can generate spatio-temporal point patterns in continuous time and continuous space. In this article, we will denote spatio-temporal point patterns by η, whether they are marked or not. Intuitively, an STPP can be thought of as a collection of stochastic processes that makes points appear or disappear from η. In the context of cell biology, such processes can, for example, describe cell division, cell movement, and cell death.

2.2 A verbal description of the relationship between spatio-temporal point processes and spatial cumulant models

The STPP rules that govern the evolution of spatio-temporal point patterns are of a stochastic nature. Therefore, a general STPP can generate an ensemble of different realisations, *i.e.,* spatio-temporal point patterns, from a given initial cell configuration. It is, in general, impossible to predict how one specific realisation will evolve in time and space. However, if we generate multiple STPP realisations, then SCMs can be used to deterministically approximate how densities and spatial covariances are expected to evolve on average for the realisations.

2.3 Reactant-catalyst-product processes and interaction kernels

The `UF-Software` can be used to generate STPPs, MFPMs and SCMs for any biological system that can be described by one or more of RCP-processes. One-point RCP-processes include density-independent cell death, as the death of a cell in location x_v with mark m can be described by a single reactant, *i.e.,* point, (x_v, m). Such one-point RCP-processes are associated with a scalar model parameter that describes the rate (probability per time unit) at which the process occurs. Multiple-point RCP-processes include density-independent cell division and cell movement. More precisely, density-independent cell division can be described as an RCP-process in which a cell of mark m in location x_v produces a daughter cell with the same mark in location x_w, so that point (x_v,m) is a catalyst and point (x_w,m) is a product. Similarly, if a cell in location x_v with mark m moves to location x_u, then point (x_v,m) is a reactant and point (x_u,m) is a product. The rate at which such multiple-point RCP-processes occur are described by interaction kernels that depend on pair-wise distances between points.

The `UF-Software` includes two pre-defined, radially symmetric, interaction kernels, being the Gaussian kernel a_{G} and the Top-Hat kernel a_{TH} such that, on a two-dimensional spatial domain,

$$a_{\mathrm{G}}[I,\sigma](r) = \frac{I}{2\pi\sigma^2} e^{\frac{-r^2}{2\sigma^2}} \tag{4}$$

and

$$a_{\mathrm{TH}}[I,r_0](r) = \frac{I}{\pi r_0^2} \quad \text{if} \quad r \leq r_0, \tag{5a}$$

$$a_{\mathrm{TH}}[I,r_0](r) = 0 \qquad \text{if} \quad r > r_0. \tag{5b}$$

Here I denotes the total integral of the interaction kernel, σ is one kernel standard deviation, r_0 is a kernel interaction radius, and r is the distance between two interacting points, *i.e.,* reactants, catalysts or products. Note that the `UF-Software` can be modified to include any user-defined, non-negative and integrable interaction kernels. These kernels need not be maximised for zero cell-pair separations.

2.4 The relationship between spatial moments and spatial cumulants

The spatial cumulants are closely related to the spatial moments, which are also called correlation functions. The first spatial moment, $k^{(1)}(x)$, measures the expected density of individuals at location x. It follows that the expected number of individuals in an area Λ at a time point t can be calculated by

$$E(|\gamma_t \cap \Lambda|) = \int_{\Lambda} k_t^{(1)}(x)dx, \tag{6}$$

where γ_t is a point configuration. The second spatial moment, $k^{(2)}(x_1,x_2)$, can similarly be used to calculate the product of 'the number of individuals in area Λ_1' and 'the number of individuals in area Λ_2', such that

$$E(|\gamma_t \cap \Lambda_1||\gamma_t \cap \Lambda_2|) = \int_{\Lambda_1}\int_{\Lambda_2} k_t^{(2)}(x_1,x_2)dx_1dx_2 + \int_{\Lambda_1\cap\Lambda_2} k_t^{(1)}(x)dx, \tag{7}$$

where the last term has been included in order to handle self-pairs (Ovaskainen et al., 2020). The first and second-order spatial cumulants are respectively denoted by $u_i^{(1)}(x)$ and $u_{ij}^{(2)}(x_1,x_2)$, where

$$u_i^{(1)}(x) = k_i^{(1)}(x), \tag{8a}$$

$$u_{ij}^{(2)}(x_1,x_2) = k_{ij}^{(2)}(x_1,x_2) - k_i^{(1)}(x_1)k_j^{(1)}(x_2). \tag{8b}$$

Here, the subscripts i and j have been included to denote different subpopulations in the modelled system, so that $i,j = 1,..,S$, where S is the number of modelled subpopulations.

From Eqs.8a,b we can see that the first and second-order spatial moments and cumulants carry the same information. To then understand why the spatial cumulants have been introduced, note first that spatial moments can be defined up to any order, and that dynamical systems of spatial moments generally are unclosed. For example, in systems in which pair-wise interactions occur, the dynamics of the νth-order spatial moment depends on the $\nu+1$st-order spatial moment. Therefore, equations that describe spatial moment dynamics are generally impossible to solve without enforcing some sort of truncation or approximation closure scheme. The same closure problem arises for equations that describe the dynamics of spatial cumulants. However, in the limit of long-ranged interactions (Sections 2.6, 2.7), the spatial cumulant formulation enables the derivation of exact equations in closed form that describe first and second-order spatial cumulant dynamics of STPP-generated populations governed by RCP-processes (Section 2.8).

2.5 The relationship between Markov operators and operators for spatial cumulant rate equations

If we let F denote a real-valued function that describes some quantity of an STPP-generated realisation γ, and μ denotes a probability measure on Γ, then

$$\langle F,\mu\rangle := \int_{\Gamma} F(\gamma)d\mu(\gamma) \tag{9}$$

describes the value of F averaged over all realisations $\gamma \in \Gamma$. The rate of change of the pairing $\langle F, \mu \rangle$ can be obtained by

$$\frac{d}{dt}\langle F, \mu(t)\rangle := \langle LF, \mu(t)\rangle, \tag{10}$$

where L is Markov operator, that acts on F and describes a process that causes points to appear on, or disappear from, a set of points η (Eq.2). In order to describe a cell population that is governed by B biological processes (here, RCP-processes), the system's total Markov operator can be written as a linear combination of Markov operators L_b such that

$$L = \sum_{b=1}^{B} L_b. \tag{11}$$

For any such Markov operator L, there exists a corresponding operator L^{Δ} that describes the time evolution of spatial moments of any order such that

$$\frac{\partial}{\partial t}k(t,\eta) = (L^{\Delta}k)(t,\eta), \tag{12}$$

where L^{Δ} is a linear operator that depends on interaction kernels of the form $a(x)$ (Section 2.3). Similarly, a Markov operator L be mapped to an operator Q^{Δ} that describes the time evolution of spatial cumulants of any order such that

$$\frac{\partial}{\partial t}u(t,\eta) = (Q^{\Delta}u)(t,\eta), \tag{13}$$

where Q^{Δ} is a sum of the linear term L^{Δ} and a non-linear term M^{Δ},

$$Q^{\Delta} = L^{\Delta} + M^{\Delta}. \tag{14}$$

Deriving the operators L^{Δ} and Q^{Δ} from a Markov operator L can be a mathematically challenging and laborious task. However, for any system of RCP-processes, the `UF-Software` can handle the formulation of L, L^{Δ} and Q^{Δ} "under-the-hood".

2.6 The perturbation expansion

As a first step towards circumventing the problem of moment closure in population-dynamics problems, Ovaskainen and Cornell (2006) initially proposed using a heuristic perturbation expansion in which interactions between points (individuals) were considered to be long-ranged, but not global. Note that Markov operators L (Eq.11) generally can be described using interaction kernels $a(x)$ (Section 2.3). However, in the limit of long-ranged interactions, Ovaskainen *et al.* (2006) introduced scaled Markov operators L_ϵ that are described by scaled interaction kernels $a_\epsilon(x)$, such that

$$a_\epsilon(x) := \epsilon^d a(\epsilon x). \tag{15}$$

The integrals of $a(x)$ and $a_\epsilon(x)$ over $\mathbb{R}^d$ have the same value, but when $\epsilon \to 0$, the scaled interaction kernels become increasingly long-ranged and flat so that each individual interacts with an increasing number of other individuals. In the perturbation expansion proposed by Ovaskainen and Cornell (2006), population densities $k^{(1)}(t)$ were expressed as

$$k^{(1)}(t) = q(t) + \epsilon^d p(t) + \epsilon^{2d} s(t) + \dots\ , \tag{16}$$

where $\epsilon = 1/\ell$ and ℓ is a characteristic interaction length-scale. In Eq.16, the term $q(t)$ denotes the density as calculated by MFPMs, and the higher-order terms $p(t), s(t), \dots$ are correction terms that account for the spatial structure of the modelled population. The perturbation expansion in Eq.16 has two notable practical advantages: (1) the calculated density $k^{(1)}(t)$ tends to the MFPM density when $\epsilon \to 0$, and (2) if we calculate $k^{(1)}(t)$ with correction terms up to the nth order, then we know beforehand that the error will behave as $o(\epsilon^{nd})$. The correction terms $p, s, \dots$ in Eq.16 can in principle be derived up to arbitrarily high orders, but truncating the perturbation expansion after one correction terms yields

$$k^{(1)}(t) = q(t) + \epsilon^d p(t) + o(\epsilon^d). \tag{17}$$

Ovaskainen and Cornell 2006 derived the expressions for $q(t)$ and $p(t)$ for a specific biological system and, based on their mathematical results, they argued that including only one correction term in the density calculation (Eq.17) is expected to suffice to capture STPP dynamics for many practical applications. Building on the heuristic work presented in 2006, Ovaskainen *et al.* later (2014) developed a

rigorous mathematical framework for deriving spatial moment and cumulant rate equations from Markov operators, where the authors expected that spatial cumulants (of any orders) could be expressed as

$$u(t,\eta) = v(t,\eta) + \epsilon^d w(t,\eta) + o(\epsilon^d), \tag{18}$$

where $v(t,\eta)$ is a mean-field term and $w(t,\eta)$ is a first-order correction term. The authors showed that, for a specific model system (the spatial and stochastic logistic model) higher-order correction terms in Eq.18 can be omitted in the limit of long-ranged interactions. Making this result more generalisable, in 2019, Cornell *et al.* presented the `UF-Software` which enables the automatic generation of $v(t,\eta)$ and $w(t,\eta)$ for the first and second-order spatial cumulants in the limit of long-ranged interactions for any system that is formulated in terms of RCP-processes. Note that, in this study, only the number of points (or densities) and pair-wise distances between points in η are used in calculations, whereas other information regarding η is omitted. Therefore, dropping the explicit η-dependency in Eq.18, we may write the first and second-order spatial cumulants as $u^{(1)}(t)$ and $u^{(2)}(t,r)$ respectively (Eq.1a,b).

2.7 The model rescaling

Recall that the spatial moment dynamics generated by a Markov operator L can be described by an operator L^Δ (Eq.12). Such a Markov operator L generally depends on interaction kernels of the form $a(x)$ (Section 2.3). If we, instead, consider a Markov operator L_ϵ that is described by the same (RCP) processes as L, but uses scaled interaction kernels of the form $a_\epsilon(x)$ (Eq.15), we must introduce a new, scaled operator L_ϵ^Δ to describe the spatial moment dynamics generated by L_ϵ. Thus, if we let $\hat{k}_{\epsilon,t}$ denote spatial moments that are generated by an STPP with scaled interaction kernels $a_\epsilon(x)$, then

$$\frac{\partial}{\partial t}\hat{k}_\epsilon(t,\eta) = (L_\epsilon^\Delta \hat{k}_\epsilon)(t,\eta). \tag{19}$$

Note that Eq.19 is unclosed for general Markov operators L_ϵ. In order to describe STPP-generated population dynamics by a system of closed equations, Ovaskainen *et al.* (2014) used a model-rescaling approach. To choose an appropriate rescaling, the authors remarked that the model should satisfy two important properties in order to be practical and relevant for biological applications: (1) when the interaction length ℓ goes to infinity, the model dynamics should tend to MFPM dynamics, as each individual would interact with an increasing number of other individuals. (2) The spatial clustering of the individuals should scale with ℓ, *i.e.,* the model should scale with respect to space. The authors therefore introduced a scaling operator S_c that rescales a point configuration η such that

$$(S_c k)(t,\eta) := k(t, c\eta), \tag{20}$$

where

$$c\eta := \{cx | x \in \eta\}. \tag{21}$$

A renormalised version of the operator L_ϵ^Δ can then be defined as

$$L_{\epsilon,\text{ren}}^\Delta := S_{\epsilon^{-1}} L_\epsilon^\Delta S_\epsilon. \tag{22}$$

If we let $k_\epsilon(t,\eta)$ denote the solution to the spatial moment equation that corresponds to this renormalised operator $L_{\epsilon,ren}^\Delta$, then

$$\frac{\partial}{\partial t}k_\epsilon(t,\eta) = (L_{\epsilon,ren}^\Delta k_\epsilon)(t,\eta) = (S_{\epsilon^{-1}} L_\epsilon^\Delta S_\epsilon k_\epsilon)(t,\eta) = (S_{\epsilon^{-1}} L_\epsilon^\Delta k_\epsilon)(t,\epsilon\eta) \Longrightarrow$$

$$\Longrightarrow S_\epsilon \frac{\partial}{\partial t}k_\epsilon(t,\eta) = (L_\epsilon^\Delta k_\epsilon)(t,\epsilon\eta)$$

and thus,

$$\frac{\partial}{\partial t}k_\epsilon(t,\epsilon\eta) = (L_\epsilon^\Delta k_\epsilon)(t,\epsilon\eta). \tag{23}$$

Note that the spatial moments are scaled in different ways in Eqs. 19 and 23. Together, these two equations enable the comparison of simulation-based results and analytical solutions. In other words, $\hat{k}_\epsilon(t,\eta)$ (which is a simulation-generated estimate of the spatial moment) and $k_\epsilon(t,\epsilon\eta)$ (which is an analytical expression of the spatial moment) should agree. Therefore,

$$\hat{k}_\epsilon(t,\eta) = k_\epsilon(t,\epsilon\eta). \tag{24}$$

Analogously to Eq.23, rate equations for scaled spatial cumulants take the form

$$\frac{\partial}{\partial t} u_\epsilon(t, \epsilon\eta) = (Q_\epsilon^\Delta u_\epsilon)(t, \epsilon\eta). \tag{25}$$

The assumed form of the spatial cumulants in the limit of long-ranged interactions (Eq.18) can be substituted into Eq.25. After such a substitution, the dynamical system is still unclosed so that the dynamics of the n-th order spatial cumulant $u_\epsilon^{(n)}(t, \epsilon\eta)$ depends on higher-order spatial cumulants. However, when solving for $v(t, \eta)$ and $w(t, \eta)$ (Eq.18), algebraic manipulation will show that $u_\epsilon^{(n)}(t, \epsilon\eta) = o(\epsilon^d)$ for $n > 2$, and thus can be omitted. This means that the model-rescaling approach (Section 2.7) together with the perturbation expansion (Section 2.6) enables the generation of closed, exact equations for the dynamics of the first and second-order spatial cumulants generated by STPPs.

The `UF-Software` can generate analytical expressions for $v(t, \eta)$ and $w(t, \eta)$ for the first and second-order spatial cumulants from any model description formulated by one or more RCP-processes. Thus, when using the `UF-Software`, a user does not need to derive any operators, or perform any algebraic calculations to obtain mathematical expressions for $v(t, \eta)$ and $w(t, \eta)$.

2.8 Spatial cumulant model equations

The `UF-Software` formulates $v(t, \eta)$ and $w(t, \eta)$ in terms of rate equations for three auxiliary functions: $q_i(t)$, $p_i(t)$ and $g_{ij}(t, \epsilon r)$ which are used to formulate the first and second-order spatial cumulants. By comparing Eqs.1 and 18 we can see that, for the first-order spatial cumulant, $v(t, \eta) = q_i(t)$ and $w(t, \eta) = p_i(t)$. And, for the second-order spatial cumulant, $v(t, \eta) = 0$ and $w(t, \eta) = g_{ij}(t, \epsilon r)$. For all subpopulations $i, j = 1, 2, .., S$, the `UF-Software` generates the right-hand sides of the differential equations

$$\begin{cases} \frac{\partial}{\partial t} q_i(t) = H_{q_i}, & \text{(26a)} \\ \frac{\partial}{\partial t} p_i(t) = H_{p_i}, & \text{(26b)} \\ \frac{\partial}{\partial t} g_{ij}(t, r) = H_{g_{ij}}, & \text{(26c)} \end{cases}$$

where $H_{q_i}, H_{p_i}, H_{g_{ij}}$ are exact expressions (Cornell et al., 2019). The expressions H_{q_i} generally depend on all subpopulation densities $q_1(t), ..., q_S(t)$. Similarly, the expressions $H_{g_{ij}}$ generally depend on all densities and leading spatial covariance terms $g_{ij}(t, \epsilon r)$. Lastly, the correction expressions H_{p_i} generally depend on all densities, leading spatial covariance terms and correction terms $p_i(t)$.

In order to calculate SCMs in practice, we first use the `UF-Software` to generate the the expressions for $H_{q_i}, H_{p_i}, H_{g_{ij}}$. Thereafter, we use the pipeline codes developed for this study (Section 3.1) to numerically calculate $q_i(t), g_{ij}(t, \epsilon r)$ and $p_i(t)$, in that order. By inserting these numerical results into Eqs.1a,b, densities and spatial covariances can be obtained. Note that, for many RCP-processes, the expressions for $H_{g_{ij}}$ include convolution integrals, which are easier to work with in Fourier space than in real space. Therefore, `UF-Software`-generated expressions for $H_{g_{ij}}$ generally include Fourier transforms of interaction kernels and variables $g_{ij}(t, \epsilon r)$.

2.9 Calculating spatial moments and cumulants from point patterns

In the space-homogenous case, the static first-order spatial moment, $k_{i,T}^{(1)}$, and cumulant, $u_{i,T}^{(1)}$, of subpopulation i at time T can be directly estimated from a point pattern as

$$k_{i,T}^{(1)} = u_{i,T}^{(1)} = \frac{N_{i,T}}{A}, \tag{27}$$

where $N_{i,T}$ denotes the number of type i individuals on domain D with area A. Static second-order spatial moments at time T can be approximated from a point pattern by

$$k_{ij,T}^{(2)}(z) = \frac{1}{A} \int_D \rho_i(x)\rho_j(x + z)dx, \tag{28}$$

where ρ_i denotes the density of subpopulation i, and z is a spatial displacement so that $z = (z^1, z^2)$ when D is two-dimensional (Dieckmann et al., 2000). The pair densities $k_{ij,T}^{(2)}(z)$ are also called auto-correlations (for $i = j$) and cross-correlations (for $i \neq j$). When approximating static second-order spatial moments

for space-homogenous systems, the magnitude $r = |z|$ of the displacement in Eq.28 is of importance, whereas the direction of z is not. In practical applications, we can thus construct a vector

$$\mathbf{k}^{(\mathbf{2})}_{\mathbf{ij,T}} = \Big(k^{(2)}_{ij,T}(r_0), \quad k^{(2)}_{ij,T}(r_{\Delta h}), \quad k^{(2)}_{ij,T}(r_{2\Delta h}), \quad \dots \Big), \tag{29}$$

for some displacement step size Δh, where each vector element $k^{(2)}_{ij,T}(r_\zeta)$ describes the density of (type i)-(type j) cell pairs that are separated by a distance $r \in [r_\zeta,\, r_\zeta + \Delta h)$. Following Eq.8b, static second-order spatial cumulants can be approximated by subtracting densities (Eq.27) from pair-densities (Eq.28). Then, analogously to Eq.29, following Eq.1, static second-order spatial cumulants can, in translationally invariant cases, be compactly approximated by a vector

$$\begin{aligned} \mathbf{u}^{(\mathbf{2})}_{\mathbf{ij,T}} &= \Big(u^{(2)}_{ij,T}(r_0), \quad u^{(2)}_{ij,T}(r_{\Delta h}), \quad u^{(2)}_{ij,T}(r_{2\Delta h}), \quad \dots \Big) = \\ \epsilon^d \mathbf{g}_{\mathbf{ij,T}} &= \epsilon^d \Big(g_{ij,T}(\epsilon r_0), \quad g_{ij,T}(\epsilon r_{\Delta h}), \quad g_{ij,T}(\epsilon r_{2\Delta h}), \quad \dots \Big), \end{aligned} \tag{30}$$

where the error terms $o(\epsilon^d)$ have been omitted as they are set to zero in practical applications. In the computational pipeline codes developed for this study, we use the Fast Fourier Transform to calculate the integral in Eq.28, and thus approximate the spatial covariance (Eq.30). Note that the correction terms in Eq.1 generally depends on all vector elements in $\mathbf{g}_{\mathbf{ij,T}}$.

2.10 Computer software and code

The `UF-Software` software is freely available to download (Cornell et al., 2019). The software comprises two main components: the `ModelSimulator` and the `ModelConstructor` which, respectively, can be used to generate STPPs and SCMs from user-defined RCP-process model descriptions. Note that, as MFPMs are a reduced form of SCMs, the `ModelConstructor` can also generate MFPMs. The `ModelSimulator` uses the Gillespie algorithm (Gillespie, 1977) and is written in the programming language C. The `ModelConstructor` is written in the Wolfram Mathematica software. Instructions on how to download the `UF-Software` are provided in the Supplementary Information (SI.1).

3 Methods and Model

3.1 The model implementation pipeline

For this study, we developed a model implementation pipeline (and pipeline codes) that systematises the use of the `UF-Software` in a number of steps (A1, B1-B4, C1-C7, D1 in Fig. 1). The pipeline includes a model formulation using RCP-processes (A1), the implementation of STPPs (B1-B4) and SCMs (C1-C7), a visual comparison of STPP, SCM and MFPM-generated population dynamics (C7), and analysis of SCM and MFPM equations (D1). Note that, as MFPM equations are a reduced form of SCM equations, they are also formulated and calculated in the pipeline. Note also that the initial cell configurations (B2, C4) can be generated *in silico* or be obtained from biological data, and that spatial cumulants from point patterns from simulation data and biological data can be calculated (B4) and visualised (C3). If suitable for the application, analytical results from (D1) can be used in the model description. Instructions on how to access, run and modify the code files for each step in the pipeline are provided in the Supplementary Information (SI.1, SI.2).

3.2 Describing a biological system using RCP-processes

The first step towards creating STPPs, SCMs and MFPMs using the computational pipeline entails translating a verbal model formulation into an RCP-formulation (Fig. 1, A1). In this study, as proof-of-concept, we consider a theoretical biological system that can be verbally described as follows:

"Two types of cancer cells, s_1 and s_2, co-exist. Cells of type s_1 produce growth factors that are both internalised by the producer cell and secreted to the environment. Cells of type s_2 do not produce growth factors and must receive growth factors from s_1 cells in order to proliferate. Both s_1 and s_2 cells can internalise growth factors from (other) s_1 cells but, because s_2 cells waste no energy on growth factor production, they benefit more from growth factor uptake than do s_1 cells. Consequently, s_2 cells have a higher growth factor-mediated proliferation rate than do s_1 cells. Similarly, s_1 cells express a higher proliferation rate when they receive growth factors from other cells, compared to when they self-produce growth factors. If drugs are applied to the system, s_1 and s_2 cells die at rates δ_1 and δ_2, respectively."

The ways in which cells can interact (and self-interact via autocrine signalling) through growth factors are pictorially shown in Fig. 2a-c, and the key conceptual difference between modelling STPP, SCM and

MFPM interactions is pictorially described in Fig. 2d,e. The modelled system can be described using the RCP-processes listed in Table 1, and the default model parameter values are listed in Table 2.

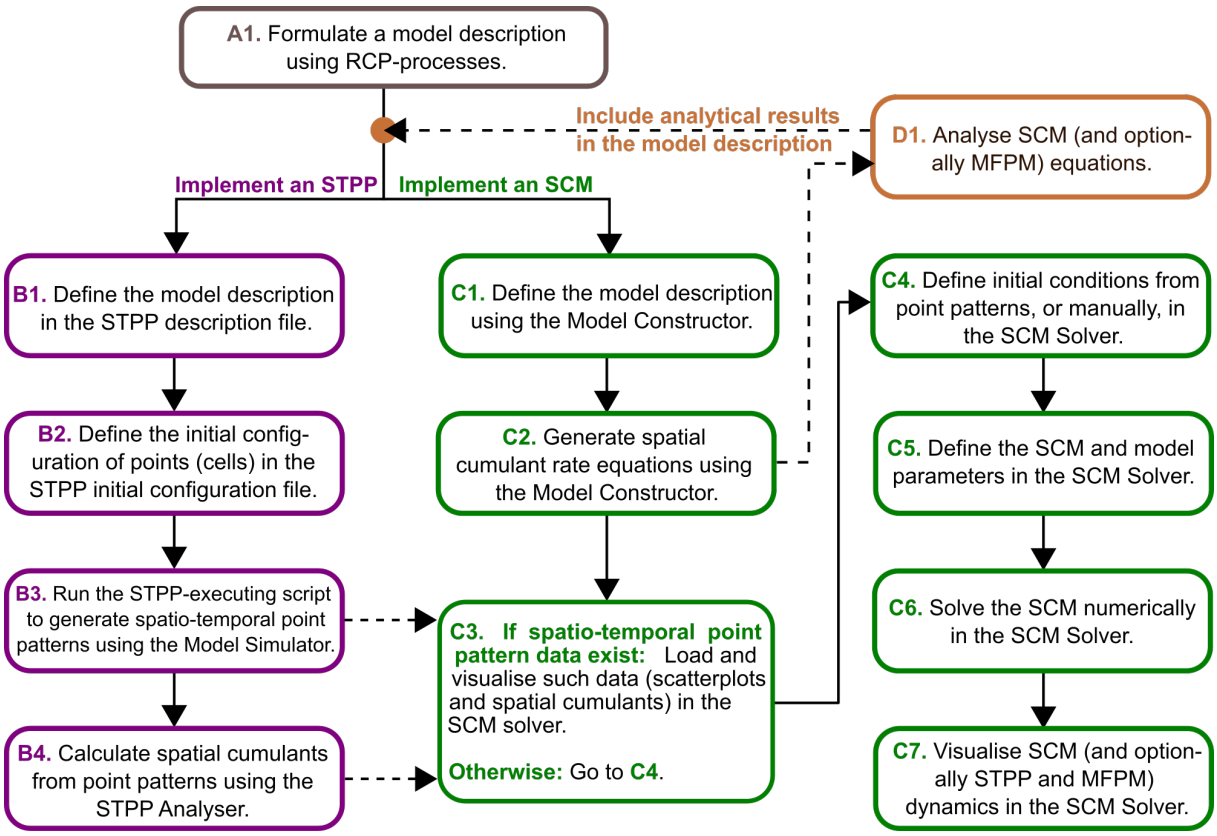

Fig. 1 We developed a computational pipeline that streamlines the implementation, analysis and output comparison of spatio-temporal point processes (STPPs), mean-field population models (MFPMs) and spatial cumulant models (SCMs). The pipeline uses the `UF-Software` (Cornell et al., 2019) to generate STPPs (B1-B3), MFPMs and SCMs (C1-C2) from reactant-catalyst-product (RCP) process model descriptions (A1). STPP (B4) and SCM (C3-C7) generated summary statistics are thereafter calculated using the pipeline codes developed for this study. Note that, as MFPMs are reduced forms of SCMs, they are also handled in steps (C1-C7)

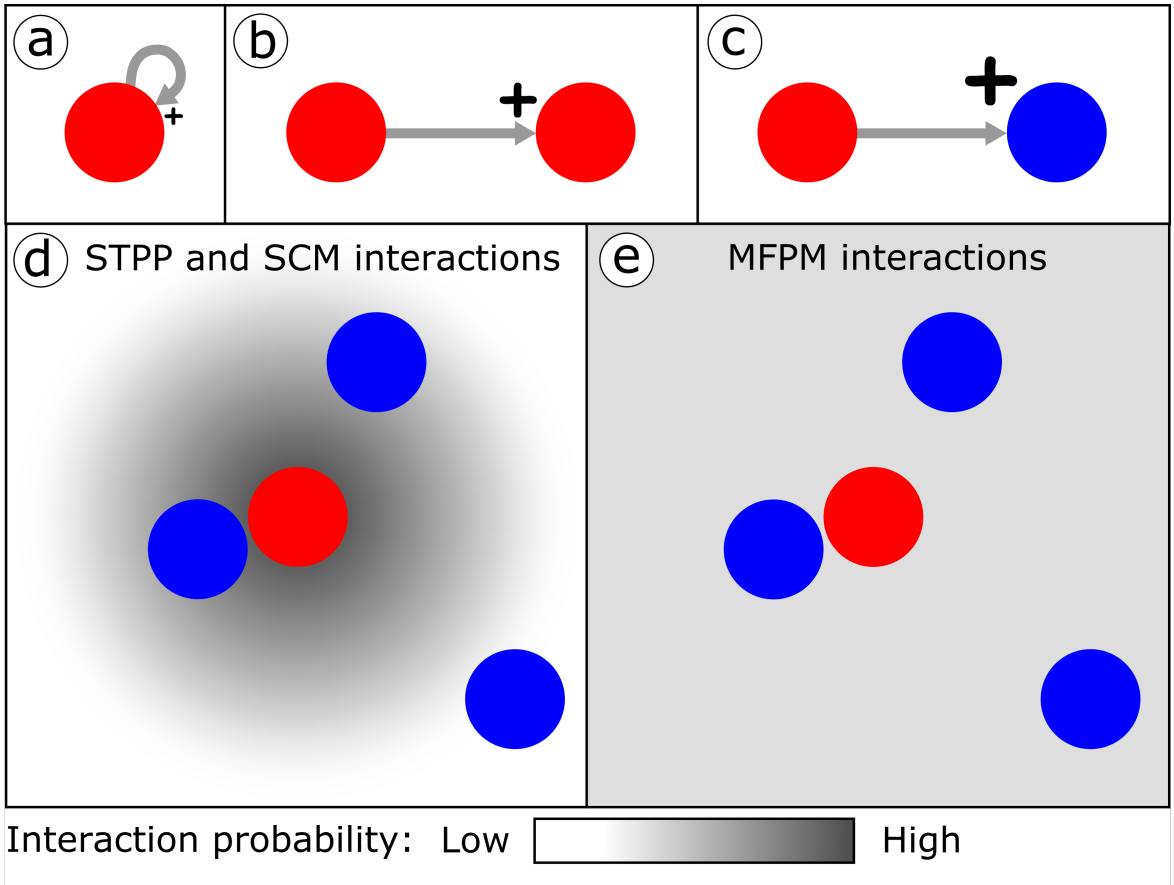

Fig. 2 We consider a theoretical system in which cells interact with each other through growth factor secretion and uptake. Panels (a-c) show the possible ways in which cells in the modelled system can interact (and self-interact via autocrine signalling). The size of the plus (+) mark indicates the benefit (in terms of proliferation rate) that the receiver cell gains from the interaction. (a) Type s_1 cells (red) can produce and internalise their own growth factors. (b) Type s_1 cells can receive growth factors from other s_1 cells. (c) Type s_2 cells (blue) can receive growth factors from s_1 cells. Panels (d,e) highlight the difference between modelling interactions in spatio-temporal point processes (STPPs), spatial cumulant models (SCMs) and mean-field population models (MFPMs). These panels show how likely the s_2 cells are to receive growth factors from the focal s_1 cell. In STPPs and SCMs, the probability that such a cell-cell interaction occurs depends on the distance between the two interacting cells, so that, for Gaussian interaction kernels, cells that are closer to each other are more likely to interact (d). In MFPMs, all cells in the system are equally likely to interact with each other, independently of spatial cell-cell separations (e)

Biological process	Reactants ($\mathbf{r}$), catalysts ($\mathbf{c}$), products ($\mathbf{p}$)	Markov operator	Description
Birth$[s_1, B_1(x_1 - x_2)]$	$\mathbf{r} = \varnothing$, $\mathbf{c} = \{x_1, s_1\}$, $\mathbf{p} = \{x_2, s_1\}$.	$L_B^{s_1}$	A cell with mark s_1 in location x_1 produces a cell of mark s_1 in location x_2 with kernel $B_1(x_1 - x_2)$.
Birth By Facilitation $[s_1, s_1, B_{11}(x_1 - x_2), C(x_1 - x_3)]$	$\mathbf{r} = \varnothing$, $\mathbf{c} = \{\{x_1, s_1\}, \{x_3, s_1\}\}$, $\mathbf{p} = \{x_2, s_1\}$.	$L_{BF}^{s_1 s_1}$	A cell with mark s_1 in location x_1 produces a cell of mark s_1 in location x_2 with kernel $B_{11}(x_1 - x_2)$. This process is facilitated by a cell with mark s_1 in location x_3 with kernel $C(x_1 - x_3)$.
Birth By Facilitation $[s_2, s_1, B_{12}(x_1 - x_2), C(x_1 - x_3)]$	$\mathbf{r} = \varnothing$, $\mathbf{c} = \{\{x_1, s_2\}, \{x_3, s_1\}\}$, $\mathbf{p} = \{x_2, s_2\}$.	$L_{BF}^{s_2 s_1}$	A cell with mark s_2 in location x_1 produces a cell of mark s_2 in location x_2 with kernel $B_{12}(x_1 - x_2)$. This process is facilitated by a cell with mark s_1 in location x_3 with kernel $C(x_1 - x_3)$.
Density Independent Death$[s_1, \delta_1]$	$\mathbf{r} = \{x_1, s_1\}$, $\mathbf{c} = \varnothing$, $\mathbf{p} = \varnothing$.	$L_{DI}^{s_1}$	A cell with mark s_1 in location x_1 dies at rate δ_1.
Density Independent Death$[s_2, \delta_2]$	$\mathbf{r} = \{x_1, s_2\}$, $\mathbf{c} = \varnothing$, $\mathbf{p} = \varnothing$.	$L_{DI}^{s_2}$	A cell with mark s_2 in location x_1 dies at rate δ_2.

Table 1: **Biological processes are formulated in terms of reactant-catalyst-product (RCP) processes.** The table includes the RCP-processes that are used to describe our theoretical system, in which growth factor producing (s_1) and non-producing (s_2) cells co-exist and interact.

Kernel	Parameter description	Parameter values	Biological description
B_1	Gaussian kernel with integral b_1 and standard deviation σ_B.	$b_1 = 0.001$, $\sigma_B = 25$.	Dispersal kernel of s_1 daughter cells, for cell division mediated via autocrine signalling of growth factors (Fig 2a).
B_{11}	Gaussian kernel with integral b_{11} and standard deviation σ_B.	$b_{11} = 0.025$, $\sigma_B = 25$.	Dispersal kernel of s_1 daughter cells, for cell division mediated via uptake of growth factors from other s_1 cells (Fig 2b).
B_{12}	Gaussian kernel with integral b_{12} and standard deviation σ_B.	$b_{12} = 0.050$, $\sigma_B = 25$.	Dispersal kernel of s_2 daughter cells, for cell division mediated via uptake of growth factors from s_1 cells (Fig 2c).
C	Gaussian kernel with integral c and standard deviation σ_C.	$c = 100$, $\sigma_C = 100$.	Interaction kernel describing growth factor secretion (from s_1 cells) and uptake (by s_1 or s_2) cells (Fig 2b,c).
Parameter			
ℓ, ϵ	Scaling parameters.	$\ell = 1$, $\epsilon = 1/\ell = 1$.	Scaling parameters that result from the perturbation expansion and model rescaling (Sections 2.6, 2.7).

Table 2: **Model parameter values.** The table shows the default model parameter values, in dimensionless form, that are used to produce the results in Figs. 3, 4, 5, 6.

3.3 Describing a biological system using Markov operators

Each RCP-process corresponds to a Markov operator (Table 1), and an STPP can be described as a sum of such Markov operators. For our regarded model system, when no drugs are applied, the total Markov operator becomes

$$L = L_B^{s_1} + L_{BF}^{s_1 s_1} + L_{BF}^{s_2 s_1} \,, \tag{31}$$

where all Gaussian birth kernels have the same standard deviations but different integrals such that

$$b_1 = 2\pi \int_0^\infty B_1(r) r dr < b_{11} = 2\pi \int_0^\infty B_{11}(r) r dr < b_{12} = 2\pi \int_0^\infty B_{12}(r) r dr, \tag{32}$$

for $r = |x_1 - x_2|$, and the total integral of the cell-cell interaction kernel is

$$c = 2\pi \int_0^\infty C(r) r dr \tag{33}$$

for $r = |x_1 - x_3|$. When drugs are applied to the system, the total Markov operator becomes

$$L^- = L_B^{s_1} + L_{BF}^{s_1 s_1} + L_{BF}^{s_2 s_1} + L_{DI}^{s_1} + L_{DI}^{s_2} \,, \tag{34}$$

where the superscript (-) has been introduced to distinguish between the Markov operators in Eqs. 31 and 34. In this study, we set the parameter values pertaining to the Markov operators $L_B^{s_1}, L_{BF}^{s_1 s_1}$ and $L_{BF}^{s_2 s_1}$ in step A1 (Fig. 1), and thus assume that these parameter values are inherent to the modelled cell population. Conversely, we assume that the parameters δ_1 and δ_2, pertaining to the Markov operators $L_{DI}^{s_1}$ and $L_{DI}^{s_2}$ can be modulated by drug doses. Accordingly, in Section 4.2, we use MFPM and SCM equations to analytically derive death rates that result in non-growing subpopulations at an arbitrary treatment time T (Fig. 1, D1). These death rates are then implemented in the STPP, as implicitly modelled treatment responses.

3.4 Describing a biological system using MFPM and SCM equations

Using the `UF-Software`, rate equations for the variables $q_i(t), p_i(t)$ and $g_{ij}(t, \epsilon r)$ in Eq.1 can be generated from RCP-formulated model descriptions. By implementing the model described by the Markov operator L^- (Eq.34) into the `UF-Software`, we obtain the following system of equations,

$$\frac{\partial q_1(t)}{\partial t} = H_{q_1} = \underbrace{b_1 q_1(t)}_{\text{from } L_B^{s_1}} + \underbrace{b_{11} c q_1(t)^2}_{\text{from } L_{BF}^{s_1 s_1}} - \underbrace{\delta_1 q_1(t)}_{\text{from } L_{DI}^{s_1}} \,, \tag{35a}$$

$$\frac{\partial q_2(t)}{\partial t} = H_{q_2} = \underbrace{b_{12} c q_1(t) q_2(t)}_{\text{from } L_{BF}^{s_2 s_1}} - \underbrace{\delta_2 q_2(t)}_{\text{from } L_{DI}^{s_2}} \,, \tag{35b}$$

$$\frac{\partial p_1(t)}{\partial t} = H_{p_1} = \underbrace{b_1 p_1(t)}_{\text{from } L_B^{s_1}} + \underbrace{W_{11}(t) + 2 b_{11} c q_1(t) p_1(t)}_{\text{from } L_{BF}^{s_1 s_1}} - \underbrace{\delta_1 p_1(t)}_{\text{from } L_{DI}^{s_1}} \,, \tag{35c}$$

$$\frac{\partial p_2(t)}{\partial t} = H_{p_2} = \underbrace{W_{12}(t) + b_{12} c \Big(q_1(t) p_2(t) + q_2(t) p_1(t) \Big)}_{\text{from } L_{BF}^{s_2 s_1}} - \underbrace{\delta_2 p_2(t)}_{\text{from } L_{DI}^{s_1}} \,, \tag{35d}$$

$$\begin{aligned} \frac{\partial \tilde{g}_{11}(t,k)}{\partial t} = H_{g_{11}} = & \underbrace{2\widetilde{B_1}(k) \Big(q_1(t) + \tilde{g}_{11}(t,k) \Big)}_{\text{from } L_B^{s_1}} + \underbrace{2\widetilde{B_{11}}(k) q_1(t) \Big(c + \widetilde{C}(k) \Big) \Big(q_1(t) + \tilde{g}_{11}(t,k) \Big)}_{\text{from } L_{BF}^{s_1 s_1}} \\ & - \underbrace{2\delta_1 \tilde{g}_{11}(t,k)}_{\text{from } L_{DI}^{s_1}}, \end{aligned} \tag{35e}$$

$$\begin{aligned} \frac{\partial \tilde{g}_{12}(t,k)}{\partial t} = H_{g_{12}} = & \underbrace{\widetilde{B_1}(k) \tilde{g}_{12}(t,k)}_{\text{from } L_B^{s_1}} + \underbrace{\widetilde{B_{11}}(k) q_1(t) \tilde{g}_{12}(t,k) \Big(c + \widetilde{C}(k) \Big)}_{\text{from } L_{BF}^{s_1 s_1}} \\ & + \underbrace{\widetilde{B_{12}}(k) \Big(c q_1(t) \tilde{g}_{12}(t,k) + \widetilde{C}(k) q_2(t) (q_1(t) + \tilde{g}_{11}(t,k)) \Big)}_{\text{from } L_{BF}^{s_2 s_1}} - \underbrace{\delta_1 \tilde{g}_{12}(t,k)}_{\text{from } L_{DI}^{s_1}} - \underbrace{\delta_2 \tilde{g}_{12}(t,k)}_{\text{from } L_{DI}^{s_2}}, \end{aligned} \tag{35f}$$

$$\frac{\partial \tilde{g}_{22}(t,k)}{\partial t} = H_{g_{22}} = \underbrace{2\widetilde{B_{12}}(k) \Big(c q_1(t) (q_2(t) + \tilde{g}_{22}(t,k)) + \widetilde{C}(k) q_2(t) \tilde{g}_{12}(t,k) \Big)}_{\text{from } L_{BF}^{s_2 s_1}} - \underbrace{2\delta_2 \tilde{g}_{22}(t,k)}_{\text{from } L_{DI}^{s_2}}, \tag{35g}$$

where $\tilde{\cdot}$ denotes the Fourier transform of $\cdot$, and k is the independent variable (the spatial frequency) in Fourier space. The k-notations have been introduced to denote the $\tilde{g}$-variables dependence on k and

$$W_{11}(t) = 2\pi b_{11} \int_0^\infty k\tilde{g}_{k,11}(t)\widetilde{C}(k)dk, \tag{36a}$$

$$W_{12}(t) = 2\pi b_{12} \int_0^\infty k\tilde{g}_{k,12}(t)\widetilde{C}(k)dk. \tag{36b}$$

The SCM is described by Eqs.35(a-g) and Eqs.36a,b, whereas the MFPM is described by Eqs.35a,b only. Note that, in the absence of drugs, the terms resulting from the operators $L_{DI}^{s_1}$ and $L_{DI}^{s_2}$ can be omitted so that the above system of equations describes the dynamics that corresponds to the Markov operator L (Eq.31). Equivalently, δ_1 and δ_2 can be set to zero.

3.5 Initial cell configurations and initial conditions

In this study, we consider three different initial cell configurations (ICCs). The ICCs have the same number of s_1 and s_2 cells, but different spatial configurations. Therefore, if we estimate the spatial cumulants from point patterns using the procedures outlined in Section 2.9, we find that the three ICCs have the same first-order spatial cumulants, but different second-order spatial cumulants (Fig. 3). In ICC.1, cells of type s_1 and s_2 are uniformly, randomly distributed across the simulated domain D. In ICC.2, cells of type s_2 are uniformly, randomly distributed across D, whereas type s_1 cells are randomly distributed within clusters. In ICC.3, both type s_1 and s_2 cells are clustered. In order to set the initial conditions for the SCM and MFPM models, we calculate densities and spatial covariances from the point patterns for all $i, j = 1, 2$. These calculations are used to set $q_i(t = 0)$ and $g_{ij}(t = 0, \epsilon r)$. The initial values for the SCM corrections are set to zero, so that $p_i(t = 0) = 0$, for $i = 1, 2$.

4 Results

4.1 SCMs capture STPP-generated cell population dynamics

We first set out to model the no-drug scenario with ICCs 1,2 and 3. To do this, we first implement the STPP (Eq.31) to generate 100 spatio-temporal point patterns from each ICC. We thereafter solve the MFPM (Eqs.35a,b) and SCM (Eqs.35a-g with Eqs.36a-b) equations numerically, using initial conditions for the spatial cumulants measured at the respective ICCs. In this proof-of-concept study, we will qualitatively say that a deterministic model "captures STPP dynamics", if the model's spatial cumulants lie within the minimum and maximum spatial cumulant values of 100 STPP-generated spatio-temporal point patterns.

Our results demonstrate that MFPM and SCM density dynamics coincide for ICC.1 only. For the simulated time duration, the MFPM is able to capture STPP-generated density dynamics resulting from ICC.1, where all cells are uniformly randomly distributed, but fail to do so for ICCs 2 and 3 (Fig. 3). As the cell-cell interactions via growth factors become more regional on the simulated domain, due to cell clustering, the MFPM performs worse in terms of describing STPP dynamics. This can be visually appreciated in Fig. 3, in which the deviation between MFPM and STPP dynamics increases for lower rows in the figure. On the other hand, the spatially informed SCM is able to capture STPP-generated density dynamics resulting from all considered ICCs in the simulated time duration (Fig. 3). This is due to the fact that SCMs account for spatio-temporal cell population structures by including equations that describe how the spatial covariances evolve in time. Indeed, our results show that spatial covariances calculated by SCMs capture those generated by STPPs (Fig. 4). In summary, the results in Figs. 3 and 4 show that SCMs can capture STPP-generated cell population dynamics, even when MFPMs fail to do so.

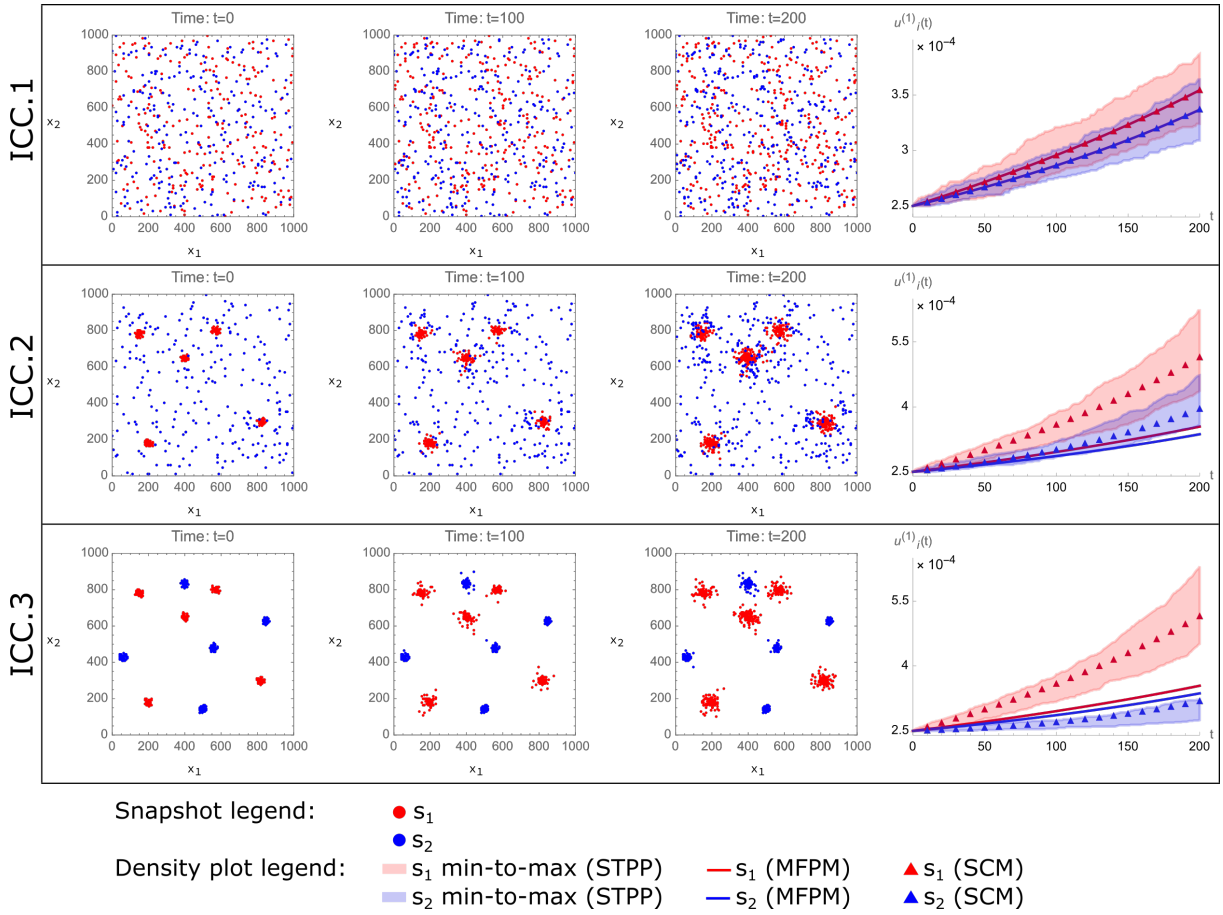

Fig. 3 Spatial cumulant models (SCMs) capture density dynamics generated by spatio-temporal point processes (STPPs), even when mean-field population models (MFPMs) fail to do so. The figure shows STPP, MFPM and SCM-generated cell population dynamics, where each row corresponds to a unique ICC (left-most column). Snapshots from one STPP-generated spatio-temporal point pattern per ICC are shown at time points $t = 0$, $t = 100$, and $t = 200$. Growth factor-producing cells (type s_1) are shown in red, and non-producers (type s_2) are shown in blue. The right-most column shows STPP, MFPM and SCM-generated density dynamics. In the density plots, the shaded STPP data shows a span of minimum to maximum (min-to-max) density values calculated from 100 spatio-temporal point patterns resulting from the same ICC. MFPM (solid lines) and SCM (triangles) results are obtained by numerically solving the MFPM and SCM equations with initial conditions calculated from the shown ICCs. The parameter values used to create the plots are listed in Table 2

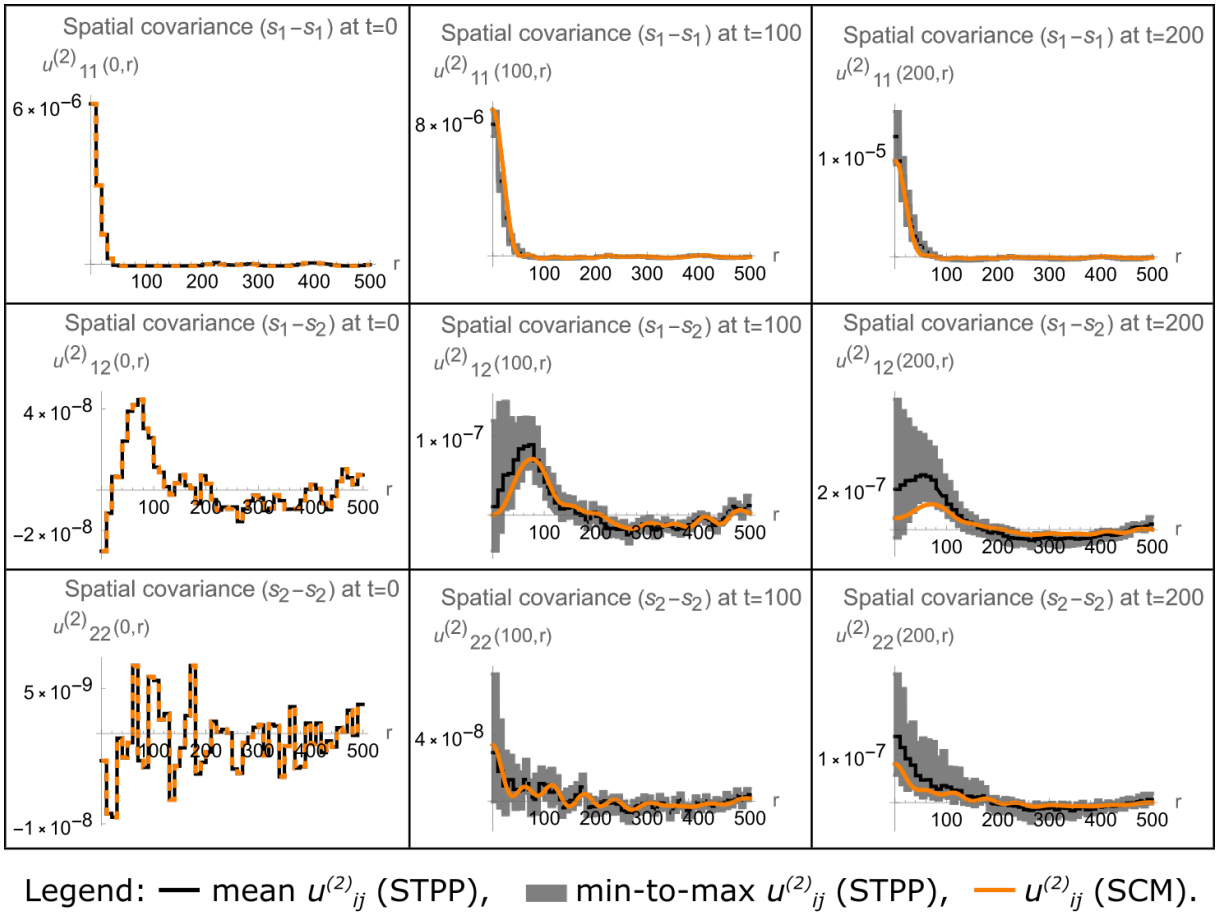

Fig. 4 Spatial cumulant models (SCMs) capture spatial covariance dynamics generated by spatio-temporal point processes (STPPs). The figure shows STPP and SCM-generated spatial covariance dynamics resulting from ICC.2 (as shown in Fig. 3). Each row corresponds to a unique pairing of cell types, and each column corresponds to a time point ($t = 0$, $t = 100$, or $t = 200$.). Mean and minimum-to-maximum (min-to-max) values from 100 STPP-generated spatio-temporal point patterns are, respectively, shown with solid black lines and shaded gray bands. SCM results (orange) in the two right-most columns are obtained by numerically solving the SCM equations, with the spatial covariances shown in the left-most column as initial conditions. The parameter values used to create the plots are listed in Table 2

4.2 SCM-informed treatment strategies outperform MFPM-informed treatment strategies in STPP-generated cell populations

We next set out to derive theoretical drug doses, here implicitly modelled via cell death rates, that stop both subpopulations from growing. In other words, we want to find the model parameters δ_1 and δ_2 that cause the derivatives describing population growths to be zero at the treatment time T. We thus derive MFPM-informed drug doses by setting

$$H_{q_1}\Big|_{t=T} = 0 \tag{37a}$$

$$H_{q_2}\Big|_{t=T} = 0. \tag{37b}$$

Solving Eqs.37a,b for δ_1^{MFPM} and δ_2^{MFPM}, where the superscripts denote that these are the MFPM-informed doses, yields

$$\delta_1^{\mathrm{MFPM}} = b_1 + b_{11} c\, \hat{u}_{1,T}^{(1)}, \tag{38a}$$

$$\delta_2^{\mathrm{MFPM}} = b_{12} c\, \hat{u}_{1,T}^{(1)}. \tag{38b}$$

Here, the hat-notation ($\hat{\cdot}$) has been introduced to denote quantities that will be measured from point patterns, and the variables $q_i(t)$ have been substituted with $\hat{u}_{i,T}^{(1)}$, *i.e.,* the measured density of subpopulation i at time T. Similarly, we derive SCM-informed drug by setting

$$\left(H_{q_1} + \epsilon^2 H_{p_1}\right)\Big|_{t=T} = 0, \tag{39a}$$

$$\left(H_{q_2} + \epsilon^2 H_{p_2}\right)\Big|_{t=T} = 0, \tag{39b}$$

$$\tag{39c}$$

and solving for δ_1^{SCM} and δ_2^{SCM} with all correction terms $p_i(t)$ set to zero upon measuring the densities and spatial covariances. Since all model parameters and variables are positive at time T, this yields

$$\delta_1^{\mathrm{SCM}} = \frac{b_1 \hat{u}_{1,T}^{(1)} + b_{11} c \cdot (\hat{u}_{1,T}^{(1)})^2 + \epsilon^2 \hat{W}_{11,T}}{\hat{u}_{1,T}^{(1)}}, \tag{40a}$$

$$\delta_2^{\mathrm{SCM}} = \frac{b_{12} c\, \hat{u}_{1,T}^{(1)}\, \hat{u}_{2,T}^{(1)} + \epsilon^2 \hat{W}_{12,T}}{\hat{u}_{2,T}^{(1)}}. \tag{40b}$$

To test these drug doses in the STPP, we, for each spatio-temporal point-pattern, measure the quantities $\hat{u}_{1,T}^{(1)}, \hat{u}_{2,T}^{(1)}, \hat{W}_{11,T}, \hat{W}_{12,T}$ at the treatment time and calculate the MFPM and SCM-informed doses. Note that the quantities $\hat{W}_{11,T}$ and $\hat{W}_{12,T}$ incorporate the spatial structure of the regarded cell population (via spatial covariances) and local effects of growth factor cell-cell interactions (via the kernel $C(r)$). After time T, we then let each spatio-temporal point pattern progress further in time with its individual death rates δ_1 and δ_2 applied to the STPP (Eq.34).

Our results show that both MFPM and SCM-informed treatment doses are able to stop subpopulations s_1 and s_2 from increasing for cell populations that started at ICC.1 (Fig. 5, top row). However, for populations that started with ICC.2, the SCM-informed doses drastically outperform MFPM-informed doses in terms of inhibiting population growths (Fig. 5, middle row). In this case, the MFPM-informed doses underestimate the required values of δ_1 and δ_2 due to the synergistic growth effects that arise from cell type $s_1 - s_1$ proximity, and cell type s_2 being close enough to s_1 cells to benefit from their secreted growth factors. In cell populations initiated with ICC.3, type s_1 cells are benefiting from being close to each other, whereas type s_2 cells are too far away the from s_1 cells to benefit from high s_1 densities. Accordingly, the MFPM-informed doses adequately inhibit population growths for s_2 but not for s_1 (Fig. 5 bottom row). However, the SCM-informed doses inhibit both s_1 and s_2 population growths for ICC.3. In conclusion, SCMs enable the derivation of treatment strategies that are informed by (1) the spatial structure of the targeted cell population, and (2) by localised cell-cell interactions. In our theoretical study, and in terms of inhibiting subpopulation growths, SCM-informed doses match MFPM-informed doses for ICC.1, and outperform MFPM-informed doses for ICCs 2 and 3.

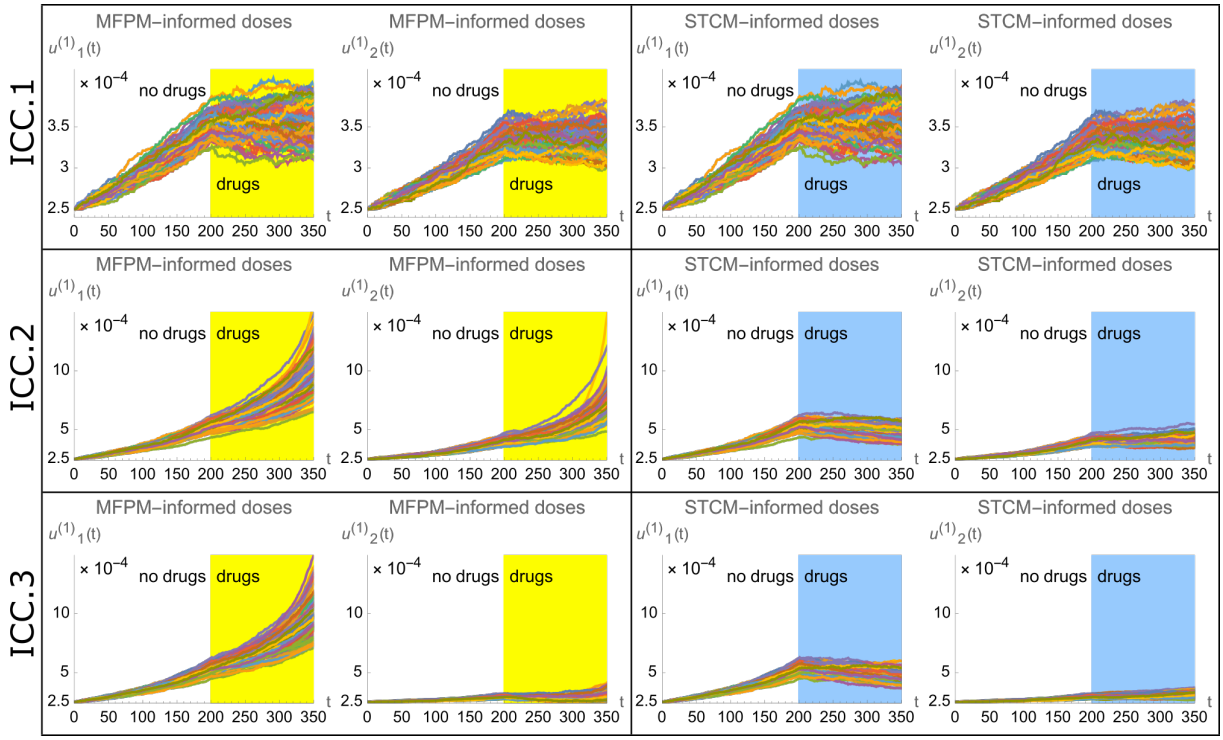

Fig. 5 Treatment strategies informed by spatial cumulant models (SCMs) outperform treatment strategies informed by mean-field population models (MFPMs) in terms of inhibiting population growths in cell populations generated by spatio-temporal point processes (STPPs). The figure shows STPP-generated density dynamics when drugs are applied at $t = 200$. Each row corresponds to the dynamics resulting from a unique ICC (as shown in Fig. 3). Each panel includes two plots, one of which shows the density dynamics of the growth factor-producing cells ($u_1^{(1)}(t)$), and one of which shows the density dynamics of the non-producer cells ($u_2^{(1)}(t)$). All plots show density dynamics from 100 spatio-temporal point patterns. Every spatio-temporal point pattern is assigned point pattern-specific drug doses (cell death rates) at $t = 200$, and is visualised with an individual line. Dynamics resulting from MFPM-informed drug doses (indicated by yellow areas) and SCM-informed drug doses (indicated by blue areas) are respectively shown in the left and right panels. The parameter values used to create the plots are listed in Table 2

4.3 Cell-cell interaction ranges and cell motility impact consistencies between STPP, MFPM and SCM density dynamics

Recall from Eq.1 that SCM subpopulation densities are expressed as MFPM densities plus a correction term. This correction term depends on the scaling factor $\epsilon = 1/\ell$, the densities and spatial covariances of the modelled population, and the RCP-processes that describe the actions and interactions of the modelled individuals via the variables $q_i(t), p_i(t), g_{ij}(t, \epsilon r)$. Consistencies (and deviations) between MFPM and SCM density dynamics thus depend both on the modeller's choice of ℓ, and on properties of the modelled population. To demonstrate these dependencies, we revisit the model system studied in Section 4.1 (Fig. 3), and include cell motility in the system. We then vary two system properties, specifically the range of cell-cell interactions and the magnitude of cell motility, to investigate how well MFPMs and SCMs capture STPP-generated density dynamics in response to these variations. In order to include cell motility to the model system, we first introduce two new RCP-processes in Table 3 that describe random movement of type s_1 and s_2 cells. The model parameter values that are associated with these RCP-processes are listed in Table 4.

Biological process	Reactants ($\mathbf{r}$), catalysts ($\mathbf{c}$), products ($\mathbf{p}$)	Markov operator	Description
Jump$[s_1, M_1(x_1 - x_2)]$	$\mathbf{r} = \{x_1, s_1\}$, $\mathbf{c} = \varnothing$, $\mathbf{p} = \{x_2, s_1\}$.	$L_M^{s_1}$	A cell with mark s_1 in location x_1 jumps to location x_2 with kernel $M_1(x_1 - x_2)$.
Jump$[s_2, M_2(x_1 - x_2)]$	$\mathbf{r} = \{x_1, s_2\}$, $\mathbf{c} = \varnothing$, $\mathbf{p} = \{x_2, s_2\}$.	$L_M^{s_2}$	A cell with mark s_2 in location x_1 jumps to location x_2 with kernel $M_2(x_1 - x_2)$.

Table 3: **Cell motility processes are formulated in terms of reactant-catalyst-product (RCP) processes.** The table includes the RCP-processes that are used to describe random cell movement of growth factor-producing (s_1) and non-producing (s_2) cells.

Kernel	Parameter description	Parameter values	Biological description
M_1	Gaussian kernel with integral m_1 and standard deviation σ_M.	$m_1 = m$, $\sigma_M = 25$.	Random movement kernel of s_1 cells.
M_2	Gaussian kernel with integral m_2 and standard deviation σ_M.	$m_2 = m$, $\sigma_M = 25$.	Random movement kernel of s_2 cells.

Table 4: **Model parameter values for random cell movement.** The table shows the model parameter values, in dimensionless form, that are used to produce the results in Fig. 6. Note that the value of m is set on an example-specific basis (Fig. 6).

By adding the Markov operators listed in Table 3 to the Markov operator L in Eq.31, we formulate a new Markov operator L^+ that describes a system in which growth factor-mediated cell division and random cell movement occur so that

$$L^+ = L_B^{s_1} + L_{BF}^{s_1 s_1} + L_{BF}^{s_2 s_1} + L_M^{s_1} + L_M^{s_2} \,. \tag{41}$$

In order to generate the MFPM and SCM equations that correspond to Eq.41, note first that cell motility does not contribute to the MFPM equations. Therefore the corresponding MFPM model equations are Eqs.35a,b without cell death, so that the death rates are $\delta_1 = 0$ and $\delta_2 = 0$. Similarly, cell motility only impacts the SCM correction equations implicitly, via their dependence on the spatial covariance variables, but not explicitly. Therefore the SCM correction terms that correspond to the Markov operator L^+ (Eq.41) are simply Eqs.35c,d with $\delta_1 = 0$ and $\delta_2 = 0$. The SCM rate equations for the spatial covariance

variables do, however, need to be explicitly updated (from 35e-g) to include cell movement. We let the plus-notation (+) denote these updated rate equations and get

$$\left(\frac{\partial \tilde{g}_{11}(t,k)}{\partial t}\right)^{+} = H_{g_{11}}\Big|_{\delta_1=0} + \underbrace{\left(2\widetilde{M_1}(k) - 2m_1\right)\tilde{g}_{11}(t,k)}_{\text{from } L_M^{s_1}}, \tag{42a}$$

$$\left(\frac{\partial \tilde{g}_{12}(t,k)}{\partial t}\right)^{+} = H_{g_{12}}\Big|_{\substack{\delta_1=0\\ \delta_2=0}} + \underbrace{\left(\widetilde{M_1}(k) - m_1\right)\tilde{g}_{12}(t,k)}_{\text{from } L_M^{s_1}} + \underbrace{\left(\widetilde{M_2}(k) - m_2\right)\tilde{g}_{12}(t,k)}_{\text{from } L_M^{s_2}}, \tag{42b}$$

$$\left(\frac{\partial \tilde{g}_{22}(t,k)}{\partial t}\right)^{+} = H_{g_{22}}\Big|_{\delta_2=0} + \underbrace{\left(2\widetilde{M_2}(k) - 2m_2\right)\tilde{g}_{22}(t,k)}_{\text{from } L_M^{s_2}}. \tag{42c}$$

The terms $H_{g_{11}}, H_{g_{12}}, H_{g_{22}}$ depend on spatio-temporal system dynamics and are listed in Eqs.35e-g. The Fourier transform of $\cdot$ is, again, denoted by $\tilde{\cdot}$ and k is the independent variable in Fourier space. In summary, the SCM equations that describe the population dynamics generated by an STPP that is governed by the Markov operator L^+ (Eq. 41) are Eqs.35a-d (with $\delta_1 = 0$ and $\delta_2 = 0$), Eqs.36a,b, and Eqs.42a-c.

Now that the MFPM and SCM equations that correspond to L^+ (Eq.41) have been formulated, we are ready to compare STPP, MFPM and SCM system dynamics for variations of the kernels C (Table 2) M_1 and M_2 (Table 4). In order for this example-specific comparison to yield qualitatively distinct outcomes, we use initial cell configuration ICC.3 in which the cells are strongly clustered (Fig. 3). Let us first consider the model system in which the cells are sessile, so that m=0 (Table 4), and vary σ_C, the standard deviation of the cell-cell interaction kernel C that governs growth factor secretion and uptake (Table 2). In Fig. 6a, we compare STPP, MFPM and SCM density dynamics for three different values of σ_C that we classify to yield strongly localised ($\sigma_C = 100$), moderately localised ($\sigma_C = 150$) and weakly localised ($\sigma_C = 200$) cell-cell interactions.

We next consider the model system with a fixed σ_C value ($\sigma_C = 100$), and vary m, the total integral of the interaction kernels M_1 and M_2 (Table 4) that govern random cell movements. In Fig. 6b, we compare STPP, MFPM and SCM-generated density dynamics for three different values of m that we classify to yield weak ($m = 0.001$), moderate ($m = 0.01$) and strong ($m = 0.1$) cell motility.

In summary, Fig. 6 provides an example-driven, qualitative three-part stratification (I-III) of MFPM and SCM performance, in terms of their ability to capture STPP-generated density dynamics: (I) Cases where both MFPMs and SCMs underperform due to strongly localised cell-cell interactions and no-to-weak cell motility (Fig. 6a,b, left column). (II) Cases where MFPMs underperform whilst SCMs perform well due to moderately localised cell-cell interactions and no cell motility, or strongly localised cell-cell interactions and moderate cell motility (Fig. 6a,b, middle column). (III) Cases where both MFPMs and SCMs perform well due to weakly localised cell-cell interactions or strong cell motility (Fig. 6a,b, right column). This stratification (I-III) emphasises that a modeller's choice of modelling method should be influenced by the model system at hand, and that the choice of $\epsilon = 1/\ell$ in applications should be influenced by how much the MFPM dynamics deviates from the STPP dynamics.

In our problem-specific examples for case (I), SCM densities lie within the minimum-to-maximum densities of 100 STPP simulations, but not within the mean plus/minus one standard deviation of the simulations. Although the SCMs perform better than the MFPMs, in terms of capturing STPP dynamics, they underestimate STPP densities towards the end of the simulation time where STPP population sizes have increased by up to a factor 28 (Fig. 6a, left column). As we have not included any crowding or competition effects in our model system, this gives rise to strongly localised effects in strongly clustered populations that cause the SCMs, as formulated in the limit of long-ranged interactions, to underperform. In order to analytically describe STPPs in such cases, a modeller might attempt to consider other limits in which to develop SCM equations, or to use other spatial moment or cumulant techniques in which higher-order correction terms are included or strongly localised interactions are accounted for. Alternatively, research question permitting, a modeller might choose to use simulation-based approaches only, without an analytical counterpart, to study the dynamics of STPP-generated cell populations.

On the other hand, in case (III), MFPMs are sufficient to capture STPP dynamics and thus the inclusion of SCM model equations provides an unnecessary complication that can be avoided. Note that, according to the mean-field assumptions that underlie MFPMs, each cell in the modelled system interacts with any other cell in the system with equal probability (Fig. 2e). It is, therefore, to be expected that the appropriateness of these assumptions, and by extension the MFPMs ability to capture STPP-generated population dynamics, improves with increasing cell-cell interaction ranges (Fig. 6a) and cell motility (Fig. 6b).

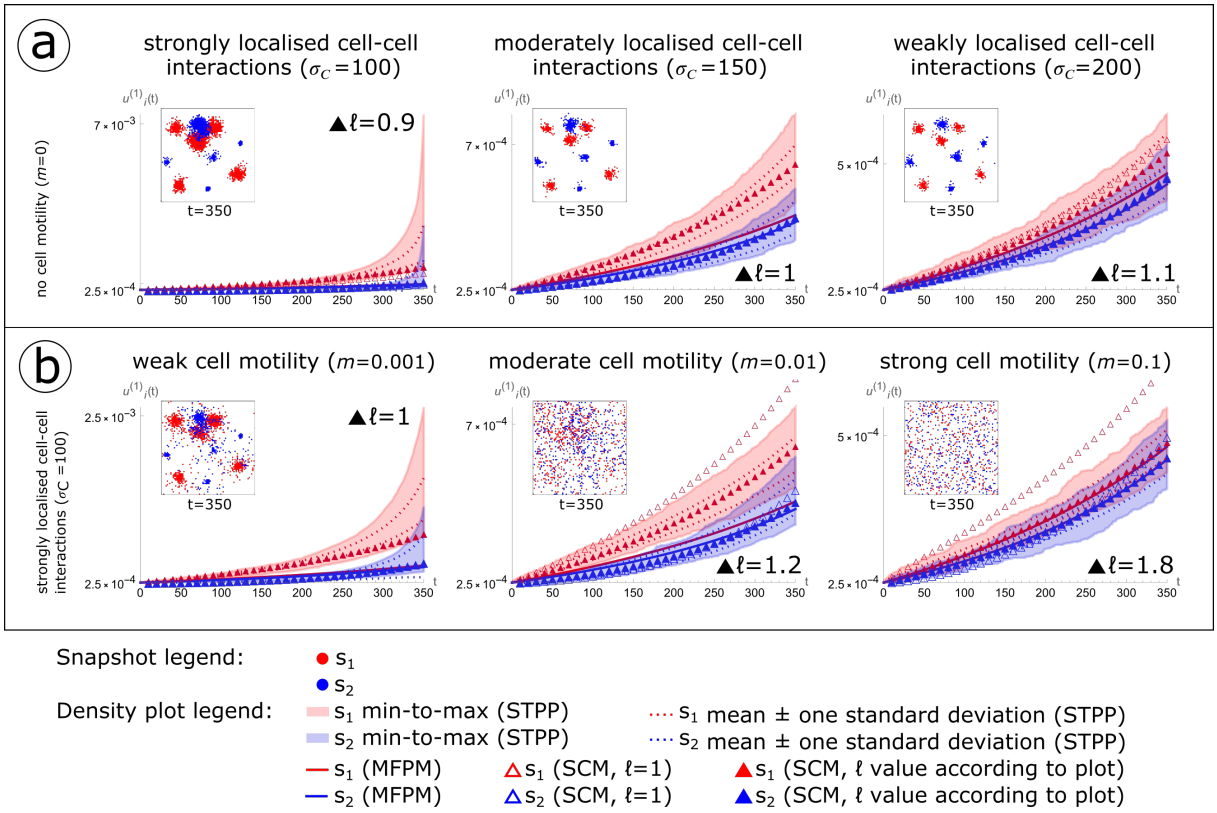

Fig. 6 Cell-cell interaction ranges and cell motility impact consistencies between density dynamics generated by spatio-temporal point processes (STPPs), mean-field population models (MFPMs) and spatial cumulant models (SCMs). The figure shows STPP, MFPM and SCM cell population density dynamics resulting from the initial cell configuration ICC.3 shown in Fig. 3. (a) The standard deviation σ_C in the Gaussian kernel that governs cell-cell interactions via growth factors (C in Table 2) is different in each column. (b) The integral m of the Gaussian kernel that governs random cell movement is different in each column. In both (a) and (b), the inserted scatter plots show one STPP-generated spatio-temopral point pattern at the simulation end time $t = 350$. Growth factor-producing cells (type s_1) are shown in red, and non-producers (type s_2) are shown in blue. In the density plots, the shaded STPP data shows the span of minimum to maximum (min-to-max) density values from 100 spatio-temporal point patterns. Density mean values plus/minus one standard deviation from these spatio-temporal point patterns are shown with dotted lines. Deterministic MFPM (solid lines) and SCM (triangles) results are obtained by numerically solving the MFPM and SCM equations with initial conditions calculated from ICC.3. Unfilled triangles show SCM dynamics when $\ell = 1$ (where the scaling parameter $\epsilon = 1/\ell$), and filled triangles show SCM dynamics when ℓ has been chosen according to the plots. Other parameter values used to create the plots are listed in Tables 2 and 4

Between the edge cases (I) and (III), lies case (II) in which SCMs provide an accurate and non-superfluous methodology to describe (Figs. 3,4,6) and manipulate (Fig. 6) STPP-generated cell population dynamics. We argue that there exist multiple applications in cell biology for which case (II) is relevant, and the use of STPPs and SCMs are appropriate. For example, STPPs and SCMs can be used to study cell systems in which cells are sessile or weakly to moderately motile, *i.e.,* not motile enough to warrant mean-field assumptions. SCMs can also be used to study systems in which cell-cell interactions are localised but extend nearest neighbour interactions. Such cell-cell interactions include localised competition for space or nutrients, and secretion and uptake of diffusible factors that can affect cell division, cell death and cell motility.

5 Discussion

Most current cancer treatments target genomic aberrations without consideration for eco-evolutionary tumour aspects that change over the course of treatment, such as cell-cell and cell-microenvironment interactions. The lack of understanding of such dynamical tumour aspects is one factor that may be hindering the development of efficient targeted treatment protocols (Gatenby et al., 2020; Wang and Deisboeck, 2019). However, clinical trials that are informed by mathematical modelling and take dynamical tumour processes into account have started to emerge (Zhang et al., 2017; Poels et al., 2021).

Although not applied in clinical settings, IBMs are widely used in mathematical oncology to study cancer eco-evolution in response to cell-cell interactions and treatments. The appeal of IBMs has increased with the resolution of state-of-the-art experimental and clinical cancer data revealing information on the single-cell level. With their spatio-temporal structure, IBMs, unlike MFPMs, have the ability to include both spatial information about how subpopulations are located in tumours, from spatial transcriptomics (Hunter et al., 2021), and temporal information about how genetic alterations in tumour cell subpopulations propagate over generations to evade treatments, from phylogenetic analysis (Schwartz and Schäffer, 2017). However, without an analytical mathematical framework in which to analyse IBMs, it is difficult to describe relationships between model variables and model parameters in a mechanistic way that lets us perturb the modeled system and better understand its inter-component dependencies. In the context of cancer, theoreticians, experimentalists and clinicians may be interested in perturbing cancer systems onto evolutionary trajectories that are favourable from treatment perspectives. Analytical mathematical models that account for treatment effects and cell-cell interactions can provide instructions for how to do so (Kaznatcheev et al., 2019).

In this article, we demonstrated that SCMs enable spatially informed treatment strategies and analysis of local interactions in cancer systems. Although the theoretical, biological systems that we modelled in this study were of a simple structure, only comprising up to five RCP-processes, the model systems could be expanded to include more biological processes. It is straightforward to use the `UF-Software` and the pipeline codes developed for this article to implement processes such as crowding effects, density-dependent cell death and cell-type conversions via cell-intrinsic or external factors. We remark that SCMs could be further improved to include more functionalities that are relevant in cell biology. For example, in this study, we categorised cells with discrete marks, so that each cell belonged to one of two subpopulations with set traits. However, in many cell biology scenarios, it would be more relevant to use continuous marks to describe *e.g.,* growth factor production, cell motility, drug resistance and gene expression in other forms. Moreover, in this study, all model parameters were cell-intrinsic constants. To allow for non-linear effects between a cell's proximity to other cells and its behaviour, the model parameters could be adjusted to be functions of *e.g.,* time, space or continuous marks. We leave these suggested improvements to the SCM framework as future challenges. Another future challenge will be to develop methods to parameterise the SCMs with experimental and, ultimately, clinical data. We emphasise that even though the mathematical theory that underlies SCMs may appear mathematically challenging at first sight, implementing SCMs (using the `UF-Software` and the pipeline codes developed for this study) is not. In order to make the methods discussed in this article accessible to the research community, we have provided clear instructions for how to access, run and modify these code files in the Supplementary Information (SI.1, SI.2).

Multiple research groups have mathematically modelled the effect that growth factor-mediated cell-cell interactions have on cancer progression and treatment outcomes. Notably, Zhang *et al.* are using an EGT model to inform treatment strategies in a clinical prostate cancer trial (2017). Their model aims to predict personalised treatment outcomes in response to drug schedules and assumed interactions amongst testosterone producing, dependent and independent cells (see also (Mistry, 2021)). Kaznatcheev *et al.* used an EGT model to show that both drugs and cancer-associated fibroblasts modulate cell-cell interactions in co-cultures of drug-sensitive and resistant non-small cell lung cancer cells (2019). Haridas *et al.* used a PDE model to quantify how interactions between melanoma cells and fibroblasts affect the

spreading of cells in *in vitro* co-cultures, finding no conclusive evidence of any cell-cell interactions via diffusible substrates (2017). Gerlee *et al.* used an IBM and an IBM-derived ODE model to demonstrate that autocrine signaling suffices to cause strong and weak Allee effects in cancer, in line with experimental observations from patient-derived brain tumor cell lines (2022).

We argue that SCMs provide a framework in which to study similar cancer research problems, with the added benefit of easy-to-produce duality: via the `UF-Software`, it is straightforward to generate both STPP-simulations and analytical SCM expressions describing user-defined biological systems. SCMs and can be especially helpful in studies when the locality of cell-cell interactions, via *e.g.,* diffusible factors or resource competition, can not be neglected. Moreover, we remark that there exist parallels between the SCM and EGT frameworks, which both are centered around interactions amongst individuals, and we theorise that the SCM methodology of incorporating space may facilitate the derivation of corrected, spatially resolved EGT replicator equations. One important contrast between EGT models and SCMs, however, lies in how the modelled interactions are described. In EGT models, the behaviours of cell populations in response to cell-cell interactions are quantified in terms of gain or loss in proliferative fitness (Farrokhian et al., 2022). On the other hand, SCMs build upon rule-based descriptions of how the individuals (cells) in the modelled system act and interact via RCP-processes. SCMs can therefore be used to define ensembles of allowable biological mechanisms, *i.e.*, candidate models, which could be evaluated against *in vitro* data, in line with a bottom-up modelling approach (Hamis et al., 2019). We also anticipate that the opportunity to analytically derive spatially informed cancer treatment strategies, as enabled via SCMs, will inspire new theoretical and applied research ventures.

Code access

Instructions on how to access, run and modify the code files used in this study are available in the Supplementary Information (SI.1, SI.2).

Funding

SH was funded by Wenner-Gren Stiftelserna/The Wenner-Gren Foundations (WGF2022-0044), the Tampere Institute for Advanced Study (2021-2023) and the Jyväskylä University Visiting Fellow Programme 2021. JAÅ was funded by Wenner-Gren Stiftelserna/The Wenner-Gren Foundations (WGF2018-0083). DST was funded by the Norwegian Research Council (NRC). JGS was funded by the National Institutes of Health (5R37CA244613-02) and the American Cancer Society Research Scholar Grant (RSG-20-096-01). MN was funded by the Academy of Finland Center of Excellence programme (project no. 312043). OO was funded by the Academy of Finland (grant no. 309581), the Jane and Aatos Erkko Foundation, the Research Council of Norway through its Centres of Excellence Funding Scheme (223257), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 856506; ERC-synergy project LIFEPLAN).

References

M. Archetti, D. A. Ferraro, and G. Christofori. Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer. *Proc Natl Acad Sci U S A*, 112(6):1833–1838, Feb 2015. doi: https://doi.org/10.1073/pnas.1414653112.

R. E. Baker and M. J. Simpson. Correcting mean-field approximations for birth-death-movement processes. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(4 Pt 1):041905, Oct 2010. doi: https://doi.org/10.1103/PhysRevE.82.041905.

R. N. Binny, A. James, and M. J. Plank. Collective Cell Behaviour with Neighbour-Dependent Proliferation, Death and Directional Bias. *Bull Math Biol*, 78(11):2277–2301, 11 2016. doi: https://doi.org/10.1007/s11538-016-0222-9.

B. Bolker and S. W. Pacala. Using Moment Equations to Understand Stochastically Driven Spatial Pattern Formation in Ecological Systems. *Theor Popul Biol*, 52(3):179–197, Dec 1997. doi: https://doi.org/10.1006/tpbi.1997.1331.

B. L. Brücher and I. S. Jamall. Cell-cell communication in the tumor microenvironment, carcinogenesis, and anticancer treatment. *Cell Physiol Biochem*, 34(2):213–243, 2014. doi: https://doi.org/10.1159/000362978.

M. Campenni, A. N. May, A. Boddy, V. Harris, and A. M. Nedelcu. Agent-based modelling reveals strategies to reduce the fitness and metastatic potential of circulating tumour cell clusters. *Evol Appl*, 13(7):1635–1650, Aug 2020. doi: https://doi.org/10.1111/eva.12943.

I. M. Chamseddine and K. A. Rejniak. Hybrid modeling frameworks of tumor development and treatment. *Wiley Interdiscip Rev Syst Biol Med*, 12(1):e1461, 01 2020. doi: https://doi.org/10.1002/wsbm.1461.

S. J. Cornell, Y. F. Suprunenko, D. Finkelshtein, P. Somervuo, and O. Ovaskainen. A unified framework for analysis of individual-based models in ecology and beyond. *Nat Commun*, 10(1):4716, 10 2019. doi: https://doi.org/10.1038/s41467-019-12172-y.

U. Dieckmann, R. Law, and J. A. J. Metz, editors. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, 2000. doi: https://doi.org/10.1017/CBO9780511525537.

A. Dominiak, B. Chełstowska, W. Olejarz, and G. Nowicka. Communication in the Cancer Microenvironment as a Target for Therapeutic Interventions. *Cancers (Basel)*, 12(5), May 2020. doi: https://doi.org/10.3390/cancers12051232.

R. Durrett and S. Levin. The Importance of Being Discrete (and Spatial). *Theor Popul Biol*, 46(3):363 – 394, 1994. doi: https://doi.org/10.1006/tpbi.1994.1032.

N. Farrokhian, J. Maltas, M. Dinh, A. Durmaz, P. Ellsworth, M. Hitomi, E. McClure, A. Marusyk, A. Kaznatcheev, and J. G. Scott. Measuring competitive exclusion in non-small cell lung cancer. *Sci Adv*, 8(26):eabm7212, Jul 2022. doi: https://doi.org/10.1126/sciadv.abm7212.

D. Finkelshtein, Y. Kondratiev, and M. J. Oliveira. Markov evolutions and hierarchical equations in the continuum. I: one-component systems. *J Evol Equ*, 9(02):197–233, 07 2009. doi: https://doi.org/10.1007/s00028-009-0007-9.

D. Finkelshtein, Y. Kondratiev, and O. Kutoviy. Semigroup approach to birth-and-death stochastic dynamics in continuum. *J Funct Anal*, 262(3):1274–1308, 09 2011. doi: https://doi.org/10.1016/j.jfa.2011.11.005.

R. A. Gatenby, Y. Artzy-Randrup, T. Epstein, D. R. Reed, and J. S. Brown. Eradicating Metastatic Cancer and the Eco-Evolutionary Dynamics of Anthropocene Extinctions. *Cancer Res*, 80(3):613–623, 02 2020. doi: https://doi.org/10.1158/0008-5472.CAN-19-1941.

P. Gerlee, P. M. Altrock, A. Malik, C. Krona, and S. Nelander. Autocrine signaling can explain the emergence of Allee effects in cancer cell populations. *PLoS Comput Biol*, 18(3):e1009844, 03 2022. doi: https://doi.org/10.1371/journal.pcbi.1009844.

D. T. Gillespie. Exact stochastic simulation of coupled chemical reaction. *J Phys Chem*, 81(25):2340–2361, 12 1977. doi: https://doi.org/10.1021/j100540a008.

S. Hamis, G. G. Powathil, and M. A. J. Chaplain. Blackboard to Bedside: A Mathematical Modeling Bottom-Up Approach Toward Personalized Cancer Treatments. *JCO Clin Cancer Inform*, 3:1–11, 02 2019. doi: https://doi.org/10.1200/CCI.18.00068.

S. Hamis, J. Yates, M. A. J. Chaplain, and G. G. Powathil. Targeting Cellular DNA Damage Responses in Cancer: An In Vitro-Calibrated Agent-Based Model Simulating Monolayer and Spheroid Treatment Responses to ATR-Inhibiting Drugs. *Bull Math Biol*, 83(10):103, 08 2021. doi: https://doi.org/10.1007/s11538-021-00935-y.

P. Haridas, C. J. Penington, J. A. McGovern, D. L. S. McElwain, and M. J. Simpson. Quantifying rates of cell migration and cell proliferation in co-culture barrier assays reveals how skin and melanoma cells interact during melanoma spreading and invasion. *J Theor Biol*, 423:13–25, 06 2017. doi: https://doi.org/10.1016/j.jtbi.2017.04.017.

T. Hillen, H. Enderling, and P. Hahnfeldt. The tumor growth paradox and immune system-mediated selection for cancer stem cells. *Bull Math Biol*, 75(1):161–184, Jan 2013. doi: https://doi.org/10.1007/s11538-012-9798-x.

P. Hinow, P. Gerlee, L. J. McCawley, V. Quaranta, M. Ciobanu, S. Wang, J. M. Graham, B. P. Ayati, J. Claridge, K. R. Swanson, M. Loveless, and A. R. Anderson. A spatial model of tumor-host interaction: application of chemotherapy. *Math Biosci Eng*, 6(3):521–546, Jul 2009. doi: https://doi.org/10.3934/mbe.2009.6.521.

M. V. Hunter, R. Moncada, J. M. Weiss, I. Yanai, and R. M. White. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat Commun*, 12(1):6278, 11 2021. doi: https://doi.org/10.1038/s41467-021-26614-z.

A. L. Jenner, C. Yun, P. S. Kim, and A. C. F. Coster. Mathematical Modelling of the Interaction Between Cancer Cells and an Oncolytic Virus: Insights into the Effects of Treatment Protocols. *Bull Math Biol*, 80(6):1615–1629, 2018. doi: https://doi.org/10.1007/s11538-018-0424-4.

A. Kaznatcheev, J. Peacock, D. Basanta, A. Marusyk, and J. G. Scott. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nat Ecol Evol*, 3(3):450–456, 03 2019. doi: https://doi.org/10.1038/s41559-018-0768-z.

Y. Kondratiev and A. Skorokhod. On contact processes in continuum. *Infin Dimens Anal Quantum Probab Relat Top*, 09(02):187–198, 11 2011. doi: https://doi.org/10.1142/S0219025706002305.

Y. Kondratiev, O. Kutoviy, and R. Minlos. On non-equilibrium stochastic dynamics for interacting particle systems in continuum. *J Funct Anal*, 255:200–227, 07 2008. doi: https://doi.org/10.1016/j.jfa.

2007.12.006.

R. Law, D. J. Murrell, and U. Dieckmann. Population Growth in Space and Time: Spatial logistic Equations. *Ecology*, 84(1):252–262, 2003. doi: http://www.jstor.org/stable/3108013.

T. Lorenzi, A. Lorz, and B. Perthame. On interfaces between cell populations with different mobilities. *Kinet Relat Mod*, 10(1):299–311, Mar 2017. doi: https://doi.org/10.3934/krm.2017012.

D. C. Markham, M. J. Simpson, and R. E. Baker. Choosing an Appropriate Modelling Framework for Analysing Multispecies Co-culture Cell Biology Experiments. *Bull Math Biol*, 77(4):713–734, Apr 2015. doi: https://doi.org/10.1007/s11538-014-0050-8.

L. M. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley. Cancer as an evolutionary and ecological process. *Nat Rev Cancer*, 6(12):924–935, Dec 2006. doi: https://doi.org/10.1038/nrc2013.

J. Metzcar, Y. Wang, R. Heiland, and P. Macklin. A Review of Cell-Based Computational Modeling in Cancer Biology. *JCO Clin Cancer Inform*, 3:1–13, 02 2019. doi: https://doi.org/10.1200/CCI.18.00069.

H. B. Mistry. On the reporting and analysis of a cancer evolutionary adaptive dosing trial. *Nat Commun*, 12(1):316, Jan 2021. doi: https://doi.org/10.1038/s41467-020-20174-4.

O. Ovaskainen and S. J. Cornell. Space and stochasticity in population dynamics. *Proc Natl Acad Sci U S A*, 103(34):12781–12786, Aug 2006. doi: https://doi.org/10.1073/pnas.0603994103.

O. Ovaskainen, D. Finkelshtein, O. Kutoviy, S. Cornell, B. Bolker, and Y. Kondratiev. A general mathematical framework for the analysis of spatiotemporal point processes. *Theor Ecol*, 7:101–113, 2014. doi: https://doi.org/10.1007/s12080-013-0202-8.

O. Ovaskainen, P. Somervuo, and D. Finkelshtein. A general mathematical method for predicting spatio-temporal correlations emerging from agent-based models. *J R Soc Interface*, 17(171):20200655, 10 2020. doi: https://doi.org/10.1098/rsif.2020.0655.

K. E. Poels, A. J. Schoenfeld, A. Makhnin, Y. Tobi, Y. Wang, H. Frisco-Cabanos, S. Chakrabarti, M. Shi, C. Napoli, T. O. McDonald, W. Tan, A. Hata, S. L. Weinrich, H. A. Yu, and F. Michor. Identification of optimal dosing schedules of dacomitinib and osimertinib for a phase I/II trial in advanced EFGR-mutant non-small cell lung cancer. *Nat Commun*, 12(1):3697, 06 2021. doi: https://doi.org/10.1038/s41467-021-23912-4.

J. Poleszczuk and H. Enderling. The Optimal Radiation Dose to Induce Robust Systemic Anti-Tumor Immunity. *Int J Mol Sci*, 19(11), Oct 2018. doi: https://doi.org/10.3390/ijms19113377.

B. A. Reynolds, M. W. Oli, and M. K. Oli. Eco-oncology: Applying ecological principles to understand and manage cancer. *Ecol Evol*, 10(16):8538–8553, Aug 2020. doi: https://doi.org/10.1002/ece3.6590.

R. C. Rockne, A. Hawkins-Daarud, K. R. Swanson, J. P. Sluka, J. A. Glazier, P. Macklin, D. A. Hormuth, A. M. Jarrett, E. A. B. F. Lima, J. Tinsley Oden, G. Biros, T. E. Yankeelov, K. Curtius, I. Al Bakir, D. Wodarz, N. Komarova, L. Aparicio, M. Bordyuh, R. Rabadan, S. D. Finley, H. Enderling, J. Caudell, E. G. Moros, A. R. A. Anderson, R. A. Gatenby, A. Kaznatcheev, P. Jeavons, N. Krishnan, J. Pelesko, R. R. Wadhwa, N. Yoon, D. Nichol, A. Marusyk, M. Hinczewski, and J. G. Scott. The 2019 mathematical oncology roadmap. *Phys Biol*, 16(4):041005, 06 2019. doi: https://doi.org/10.1088/1478-3975/ab1a09.

R. Schwartz and A. A. Schäffer. The evolution of tumour phylogenetics: principles and practice. *Nat Rev Genet*, 18(4):213–229, 04 2017. doi: https://doi.org/10.1038/nrg.2016.170.

D. Stichel, A. M. Middleton, B. F. Müller, S. Depner, U. Klingmüller, K. Breuhahn, and F. Matthäus. An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. *NPJ Syst Biol Appl*, 3:5, 2017. doi: https://doi.org/10.1038/s41540-017-0006-3.

A. Surendran, M. J. Plank, and M. J. Simpson. Spatial Moment Description of Birth-Death-Movement Processes Incorporating the Effects of Crowding and Obstacles. *Bull Math Biol*, 80(11):2828–2855, Nov 2018. doi: https://doi.org/10.1007/s11538-018-0488-1.

B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264, Sep 2015. doi: https://doi.org/10.1038/nature14971.

Z. Wang and T. S. Deisboeck. Dynamic Targeting in Cancer Treatment. *Front Physiol*, 10:96, 2019. doi: https://doi.org/10.3389/fphys.2019.00096.

J. West and P. K. Newton. Cellular interactions constrain tumor growth. *Proc Natl Acad Sci U S A*, 116 (6):1918–1923, 02 2019. doi: https://doi.org/10.1073/pnas.1804150116.

J. Zhang, J. J. Cunningham, J. S. Brown, and R. A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nat Commun*, 8(1):1816, 11 2017. doi: https://doi.org/10.1038/s41467-017-01968-5.