Master's Thesis

Cell Invasion Models: How Complex do They Need to Be?

Vorgelegt von:

Veronika Hofmann

München, 16.01.2024

Aufgabenstellerin und Betreuerin:

Prof. Dr. Christina Kuttler

Veronika Hofmann

veronika.anja.hofmann@tum.de

Ich erkläre hiermit, dass ich diese Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

München, 16.01.2024

. . . .

Ort, Datum   (Veronika Hofmann)

Abstract

The invasion of the extracellular matrix (ECM) by cancer cells is a highly complex process, but many mathematical models describing it are quite simple. This raises the question how these simpler models differ from the more complex ones regarding their dependence on the input parameters and their qualitative results, and whether simpler models are able to capture and reproduce the results of complex models.

To investigate these questions, three cell invasion models are examined: two simple to intermediate partial differential equation (PDE) models with six and eight parameters, respectively, and a complex hybrid model with 17 parameters. The hybrid model was originally developed for angiogenesis, featuring a Cellular Potts model and a finite element formulation, and is extended in this work to describe cell invasion as travelling waves.

The models' parameter sensitivity with respect to initial ECM density, the ECM degradation rate and the cell proliferation probability, is examined using a variance-based method at various points in time. At first, the PDE models exhibit mainly first-order effects from the initial ECM density, and later in time from the ECM degradation rate. At all points in time, the hybrid model is affected the most by the proliferation probability, and by interaction effects between all three parameters.

The data fitting abilities are tested by generating data using the hybrid model and estimating the corresponding PDE model parameters. This is done under consideration of various phenomena of the hybrid model, and with different numbers of parameters to be estimated. Using the fitted parameters, the shape approximations and the velocities of the travelling invading waves are compared between the models. The best approximations in shape are performed by the PDE model with eight parameters, the best approximations in wave speed are obtained using the model with six parameters.

Overall, the main finding is that even though the model outputs are relatively similar and can be approximated adequately by the PDE models, the dependence of the hybrid model on its input parameters differs substantially from the one of the PDE models.

Acknowledgements

I would first like to thank my thesis advisor, Prof. Dr. Christina Kuttler of the professorship of mathematics in life sciences at the Technical University of Munich (TUM), for her excellent supervision of my work. She enabled me to explore the topic at my own speed and let me decide which aspects I wanted to delve into, while always being at my side to hand out expert advice on any problem or question I had during this journey. Also, she read the first version of this thesis in record time and made very helpful and constructive suggestions for improvement. Her own interest in the topic and her enthusiasm for my results gave me additional motivation, for which I am particularly grateful.

I wish to express my sincere thanks to Dr. Pirmin Schlicke, also of the professorship of mathematics in life sciences at TUM, for his great advice on the variance-based sensitivity analysis, and for lending me the computational power of his laptop. His help saved me an extra two weeks of waiting time for the results of the sensitivity analysis.

I would like to thank Rebecca Crossley of the mathematical institute of the University of Oxford for the nice and helpful exchange regarding the numerical methods she used for solving the partial differential equations.

I am thankful to my proofreader and friend Julius, who revised my grammar and the comprehensibility of my work, and spared no details. Hopefully, the acknowledgements are put to his satisfaction as well.

I would like to thank my friend Martin for his helpful advice on correcting and optimizing my code translations. That must have taken a lot of patience, which he thankfully had. My sincere thanks also goes to my friend Adrian, who helped me to overcome technical difficulties that occurred in the generation of some of my figures.

I am also grateful to my family for their support and interest in my results. In particular, I would like to thank my father, Matthias, for sparking my curiosity about mathematics and natural sciences and for being my role model, my mother, Birgit, for her good advice when I was stuck, and my sister, Alexandra, because she always had an open ear for me. I take this opportunity to express gratitude to everybody, who directly or indirectly, have lent their hand in this venture.

Contents

List of Figures

List of Tables

1. Motivation

Cell invasion describes the process of cancer cells destroying the tissue that surrounds them, and using the resulting space to intrude further. Coupled with proliferation, this leads to a disruptive, growing mass of malignant cells – the tumor. Often, the invasion's first victim is the extracellular matrix, a scaffold-like structure that under normal circumstances keeps the cells in line and enables the organs to fulfill their functions.

Not only physicians are inclined to understand the mechanisms underlying this fatal process. While mathematical oncology is a young field, the mathematical modeling of cell invasion is considered an important task, too. If an adequate model is used, researchers are able to make accurate predictions of the course of the disease and find promising treatment options.

Of course, this often rather involves models of the efficacy of certain chemotherapeutic agents or models of the dependence of tumor size on a certain treatment plan. But modeling the invasion speed can be very useful in practice as well: for instance if it is possible to predict the tumor size development by modeling the velocity of the invasive cell front, physicians can estimate how long surgery can be delayed until the tumor has reached a critical size that increases the likelihood of metastases or puts the affected organ in too much danger. This helps organizing hospital resources in times of unusually high demands, as it was the case during the most difficult times of the COVID-19 pandemic.

Cell invasion models could step in earlier in the medical process as well: the predictions of mathematical models can be used as an extension to medical trials, where they filter which drugs or treatment schedules are worthy being tested in an animal trial. This procedure is already used in certain medical fields such as antibiotics research [1].

The models that could – at least in theory – be used for predictions and filters like this are numerous and vary a lot regarding their complexity. "Complexity" in this context refers to the number of considered parameters and processes.

Indeed, cell invasion is a highly extensive phenomenon as it will be shown in this work, and it is easy to get lost in the biochemical interactions and physical feedback loops. Luckily there are many simple models, too, which provide some guidance through the world of cell invasion modeling by focusing only on what is absolutely necessary to pass as cell invasion model.

Despite varying complexity, all models describe the same process, hence they all share some common features. Between most models, equivalent processes and parameters can be identified. Nevertheless, this does not mean that these parameters impact the models' outputs in the same manner. One objective of this work is to gain an understanding of how the model-inherent processes differ considering models of various complexity. How different do they react depending on the choice of input parameters?

Also, the qualitative differences of the model outputs depending on complexity will be studied. How versatile are rather simple models, is it possible to obtain similar results with them as with the more complex ones?

And, as an implication of the answers to all of these questions: what might be the best fields of application for the respective model category?

Starting this journey, an overview of the biological processes involved in cell invasion is provided in chapter 2. It is followed by the introduction of two low- and intermediate-complexity partial differential equation (PDE) models as well as the rededication of a high-complexity hybrid angiogenesis model, consisting of a Cellular Potts model and a finite element formulation, to a model for cell invasion in chapter 3. The relevance and interplay of the model parameters is investigated in chapter 4, where a sensitivity analysis is performed with the three models. Chapter 5 focuses on the model outputs and tests how phenomena exhibited by the most complex model can be captured by the two simpler models.

2. Cell Invasion: Actors and Reactions

Before delving into the mathematical modeling, the process of cell invasion and the parties involved should be well understood. What is the extracellular matrix (ECM)? What is its purpose and how does it interact with the cells? How is it involved in cancer? What exactly is cell invasion, and what determines its direction and speed?

These are the questions that are answered in this chapter. While not every detail is important for the mathematical models that are introduced later in this work, it is worth having an impression of how many different factors and mechanisms *could* be considered in the process of modeling cell invasion.

2.1. Properties and Functions of the ECM

As the name already hints, the *extra*cellular matrix is the non-cellular substance surrounding the cells in every tissue and organ [2]. In mammalian ECM – which is the kind that is considered here – the main components are water, proteins and polysaccharides. One differentiates between two classes of macromolecules that shape the ECM, proteoglycans and fibrous proteins. Proteoglycans are a type of polysaccharides, and as hydrogels (meaning that they cannot be dissolved in water, and bind water molecules so that they appear as viscous liquids) they serve mainly as fillers and lubricants. Furthermore, they possess growth factor receptors allowing them to participate in cell communication [3]. Fibrous proteins are the main operands building the ECM structure. Among others, these fibrous proteins comprise collagen, elastin and fibronectin. In terms of protein mass, collagen is the major fibrous component of the ECM and as the so-called "collagenous backbone" [3] of the ECM it determines its mechanical properties through its fiber orientation [4]. Also, collagen plays a role in cell adhesion, chemotaxis, migration and differentiation [2]. Elastin fibers are highly extensible and can be found in high concentrations in tissues that are subject to repeated stretching, but their elasticity is limited by the surrounding collagen fibers [2]. Fibronectin serves as "biological glue" [3], regulating cell attachment and migration.

The components and structure of the ECM vary depending on the given cells and tissue, and even within the same tissue there can be differences. For instance, the ECM present

in cartilage tissue exposes three different zones to provide the optimal mix of shear and compression resistance as well as lubrication [4]. Illnesses, especially cancer, alter the organization of the ECM as well by destroying its structure and making it stiffer by changing its composure [2].

2.1.1. Mechanical Characteristics of the ECM

In most tissues, the ECM is a nonlinear elastic material, and it increases its stiffness with increasing deformation. This "strain-stiffening" behavior allows the ECM to transmit low forces and to be robust under stress at the same time [5]. The ECM achieves this through fiber reorientation: fiber alignment and densification make it more resistant against load [6]. The nonlinear elasticity is not only important for the physical protection of the tissue, it also serves as an information transmitter: several studies have found that in biological matrices made out of fibrin (a biopolymer that is synthesized during wound healing and serves as temporal ECM substitute [2, 5]) or collagen (the usual main component of the ECM), displacements due to cell contractions travel considerably further than in artificial, linear elastic substrates [7, 8, 9]. The force transmission serves as a communication tool for the cells such that they are able to organise processes such as the sprouting of capillaries [6].

The exact mechanical behavior of the ECM depends of course on its composition: collagen fibers are either semiflexible or stiff, whereas elastin fibers are elastic. A matrix made out of elastin alone would be linear elastic, but the surrounding collagen produces the non-linearity in the relationship of stress and strain. Apart from the strain-stiffening, another interesting mechanical behavior of the ECM is "microbuckling": under repeated compression, stiff ECM components, such as collagen or – in case of wound healing – fibrin fibers, change their shape and, depending on their orientation, might not regain their original shape when the exertion of force stops. This effect provides a feedback to the strain-stiffening as the resulting fiber lengthening shifts the material's response to stress to greater deformations: after the microbuckling occurred, the same amount of compression lead to more strain than before. It has to be noted that microbuckling has only been observed in fibers that are not covalently cross-linked. The researchers who discovered this property (Stefan Münster and colleagues) assume that it has consequences for cell migration: travelling cells exert forces on the ECM, compressing collagen fibers which then either resist the force and support the cells' motions (if they are cross-linked) or bulk under the force and lengthen, inhibiting the cells (if they are not) [10]. Interestingly, it has been found that at least in the case of breast cancer, tumorigenesis leads to collagen cross-linking [11]. Therefore, Münster et al. suspect that "not only the stiffening of the ECM

caused by the cross-linking of collagen, but also the inhibition of fiber lengthening might play an important role for the increased ability of cancer cells to invade the surrounding ECM" [10]. Regarding cross-links, it should also be mentioned that the collagen cross-linking, which accompanies cancer, increases the elastic modulus of the tissue by a factor 10 and thereby makes it much stiffer [11].

2.1.2. Tasks of the ECM

Just like its composition, the functions of the ECM can be very different depending on the tissue it is situated in. To give some examples, the ECM surrounding muscle cells helps fulfilling their physiological functions by transmitting mechanical force [12], whereas the mineral-rich ECM in bones provides rigidity and strength at a low weight for the nearby osteoblasts [13]. The perhaps most obvious function of the ECM is the provision of structural support for its environment: as mentioned previously, the ECM serves as a scaffold to which the cells can attach via their cytoskeleton [14]. In this manner, the ECM, together with the cells, builds tissues and organs with specific physical properties, e.g. elasticity, tensile and compressive strength. This is especially important for stem cell differentiation: the composition and mechanical properties of the ECM instruct the stem cells to develop towards a certain type of cell [2].

The ECM's chemical composition mediates inter- and intracellular processes. The surrounding cells possess various receptors on their surfaces to communicate with the ECM. Especially the receptor family integrin should be mentioned since with its 24 members it is the most widespread and serves as a two-way-mediator: "Integrins activation can lead to either outside-in signaling, (...) leading to cell polarity differences, cell survival and proliferation and cytoskeleton rearrangement, or to inside-out signaling, (...) which affect the cell adhesion and migration" [14]. Various forms of integrins can be found on the surfaces of both normal and malignant cells: some bind several ECM molecules, others are more selective and only bind a certain molecule. The strength of a cell's expression of certain integrins determines its adhesion to the ECM, how much force it can generate to move, or how likely it is to die by apoptosis, the programmed cell death that is induced by the interaction of integrins and appropriate ECM molecules [15]. In the case of cancer, this variety of tasks which integrins are capable of fulfilling makes the protein involved in almost every step of cancer progression: integrins can be found in cancer initiation, proliferation, local invasion, entering the vascular system, survival of the travelling cancer cells that are in the circulation system, migration into the secondary site, and eventually the growth of the metastatic tumor [16].

Through ECM remodeling, tissue homeostasis – an equilibrium state of a property within a tissue, e.g. number of cells – and morphogenesis – the process during embryonic development in which tissues assume their functional shape – are being controlled. This remodeling can take place after an impulse is given by the surrounding cells, but also autonomously through stimulation by released ECM components [17].

While the cells need the ECM for structure, and as a mediator of the various processes listed above, the ECM needs the cells, too. For instance in connective tissue, the most common cell type are fibroblasts, whose main functions are the synthesis of ECM elements, such as collagen or elastin fibers [18].

2.2. Mechanisms of Cell Invasion

Tissue invasion and the subsequent forming of metastases is one of the "hallmarks of cancer", as denoted in the famous paper by Hanahan and Weinberg, and it is the deadliest aspect of the disease [19]. After uncoupling from the original solid tumor, cancer cells – individually or as a group of multiple cells – start spreading into the nearby tissue. They are able to reach sites further away by entering the circulation system. After escaping the blood circulation, the cancerous cells start to proliferate and build a secondary tumor at the new site, the metastasis [20].

During the initial growth of the cancer, the ECM is used as a scaffold onto which the cells can attach to. After this initial growth, the cancer degrades the ECM (a process known as proteolysis), in order to create space in which to invade into [20]. Here, the integrins on the aberrant cells' surfaces play an important role. As mentioned previously, integrins determine the adhesiveness of the cell to the ECM. Optimal conditions for cell migration are given when the adhesiveness is at an intermediate level, allowing for "traction at the cell front while releasing contacts at the rear, resulting in net forward movement" [15]. Adhesiveness can be altered by lowering the integrin expression or the attraction strength of integrins for their ligands, which are the molecules that can bind to the integrin. The simultaneous ECM degradation is promoted by the interaction of integrins and proteases, the enzymes that disaggregate ECM molecules. Proteolysis has to be carefully balanced, as excessive ECM degradation could reduce traction in a way that impedes cell migration. [15]

2.2.1. Migration versus Invasion

Cell migration describes the trajectories and velocities of cells across accessible surfaces, such as across a Petri dish, or along the ECM scaffold when there is space available. Cell invasion, on the other hand, is a form of migration that requires the cells first to

disintegrate a barrier which they are about to transit. Among these barriers are other cells, or – as it is the case in this work – ECM elements. Invasion is a necessary process that groups of malignant cells undergo when they proliferate or metastasize. [21]

Cell invasion does not only occur for cancerous cells, but also during normal cell morphogenesis and wound healing. After the cells adhere to the barrier, it is disintegrated to enable migration (if the barrier is the ECM, the disintegration corresponds to proteolysis). [22]

It should be mentioned that cell migration and invasion can take place both individually and in groups. In the case of collectively migrating or invading cells, the cells use integrin-induced cell-cell adhesion to form a bulk divided in two parts: the first part consists of a single cell, the leader, which dictates the direction for the cohort and degrades only the parts of the ECM that need to be removed for it to travel through. The second part consists of subsequent cells widening the gap created by the first cell. The leading cells have been found to exhibit higher viscosity than their followers [23]. In the case of very soft tissue, the leading part can consist of multiple cells instead of one single leader [24].

2.2.2. Mesenchymal and Amoeboid Migration

Cancer cells have been observed to travel mainly in two modes: mesenchymal or amoeboid. Mesenchymal migration is a process that requires the cell to be able to form strong adhesions to the substrate, which are alternately dissolved and reassembled, creating forward movement (to be more precise: the cell needs to build so-called focal adhesions, which is discussed in detail in section 2.2.4). ECM degradation is an important ingredient for mesenchymal migration, since otherwise the cell is not capable of creating space for itself to invade into [25]. Mesenchymal migration can be observed for any tumor type at a certain stage, but is particularly frequent for cancer migration in connective tissues. [24]. Amoeboid migration on the other hand is less dependent on adhesion or ECM degradation. The cells hardly adhere on the substrate and move forward by contractions of their cytoskeleton. Some amoeboid cells exhibit such a high contractility, so that they do not need to disaggregate the ECM, squeezing instead through the tissue gaps. Amoeboid cell movement is mostly observable in hematopoietic or neuroectodermal tumor cells, but most tumor types consist of several amoeboid cells. [24, 25]

Depending on the cells' environment, i.e. the composition of the ECM, cancer cells can switch between these modes of migration. If the ECM is three-dimensional (as is is the case in most *in vivo* scenarios), an inhibition of cell-ECM adhesion mechanisms and ECM-degrading enzymes forces mesenchymally migrating cells into the amoeboid migration mode. [16, 26].

2.2.3. Hapto-, Duro- and Chemotaxis

When talking about cell movements, several variables can be considered. So far, mostly the manner of locomotion has been discussed, i.e. whether the cells need to destroy the ECM on their trajectory. Apart from that, the direction of their movement is important as well. In theory there are situations in which no direction can be detected – the cells spread through diffusion only – but most of the time some kind of attractor or repellent is present and guides the cells. The form of this "pulling" or "pushing" substance defines the kind of directed movement: if the cells are following a concentration gradient of a substrate which is dissolved in its environment, then the migration happens on the basis of chemotaxis. Chemotaxis can be observed not only in cells, but in many creatures – for example the manner in which many animals find conspecifics by following increasing concentrations of pheromones in the air which have been emitted by potential mates. [27]

Returning to cell movements along the ECM, another way of attracting or repelling cells is haptotaxis: it works very similar to chemotaxis, but here the substrate has to be bound to the ECM in contrast to being dissolved. The exact definition of haptotaxis varies from author to author; originally, it has been introduced to describe directed cell movement up a gradient of surface-adhesiveness [28]. Sometimes it is still used in this sense, for instance by Hood et al. [15], but other authors use the term haptotaxis to describe "chemotaxis on gradients of surface-bound attractants" [29, 30]. The latter definition – which is also the definition that is used in this work – allows for a switch between the haptotaxis cell movement, and the chemotaxis cell movement, when the attractor uncouples from the surface.

The last form of directed cell movement which is important for ECM invasion processes, is durotaxis. It is a relatively new discovery and describes how cells are attracted by regions of higher surface stiffness and repelled by softer ones [31]. Originally, durotaxis has been observed in fibroblasts, but in 2019 it could also be attested for cells responsible for brain tumors, breast, skin and bone cancer [30]. The effect of durotaxis on the cancer cells' movements appears to be a saturation process, meaning that as soon as a certain value of stiffness is reached, the attraction by even stiffer surfaces is no longer increased [30]. The correlation between the cell velocity and stiffness gradient depends on the cell type: cancer cells were not observed to increase their cell velocity with an increase in steepness of the stiffness gradient [30], whereas stem cells do [32]. In cancer research, durotaxis is worth studying since a strong correlation has been observed between increasing stiffness of the tumor microenvironment and more aggressive tumor expansion [30].

2.2.4. Focal Adhesions

To perform durotaxis, there needs to be way for the cells to "feel" the stiffness of the substrate on which they are placed. It has been mentioned earlier that on the cell's membrane multiple receptors can be found, where the various integrins have been stressed to be among the most common and versatile ones. At certain sites within the cell, clusters of integrin can be found which directly link the cell's cytoskeleton to the ECM. These regions, where force transmission between the cell and the ECM are possible, are called focal adhesions. Integrins, the main cellular adhesion receptors, can sense not only chemical, but also physical properties of the ECM [16]. Being sensitive to mechanical impulses, focal adhesions grow and stabilize (i.e. become less likely to disassemble) under force, resulting in a stronger "grip" which the cells have on the ECM. It does not matter whether this force is generated by the cell itself or whether the tension is applied by its environment, for instance by a rigid, hardly deformable surface. Durotaxis can be explained by the behavior of the focal adhesions: on soft, flexible surfaces, the tension acting on the cell's cytoskeleton may be smaller than the force needed to sustain the adhesion site, leading to the formation of relatively few and small focal adhesions. On stiffer surfaces, one can observe the contrary effect, where focal adhesions increase in size and number. On a stiffness gradient, this behavioral difference leads to an imbalance of focal adhesion development within the cell, provoking cell polarization (i.e. creating a spatial asymmetry within the cell) and making it oriented toward the stiffer site. [33, 34]

This phenomenon also influences the shape of the cell: less focal adhesions make the cell appear small and rounded, and the more focal adhesions are developed, the flatter and more elongated is the cell's shape [35].

Of course, asymmetric adhesion to the substrate alone does not create movement. When the cell is elongated, the front part is responsible for the creation of focal adhesions, and the back part of the cell is responsible for removing the adhesive sites. There are several possibilities for this to happen: integrins can be extracted from the cell membrane and left behind on the substrate, they can diffuse and be placed randomly on the cell surface, dissolving the focal adhesion to which the belonged, or they are packed into a vesicle and get more or less randomly transported to the front end of the cell. Especially the first case makes biosynthesis of new integrins to obtain new focal adhesions at the front end necessary [33]. It has to be stressed, that the first scenario has not been observed in malignant cells, but the other two scenarios have been found to occur in cancer. There even is a motor protein which helps transporting integrin to the cell front, making the integrin recycling in cancer cells more efficient. [16]

2.2.5. Proteolysis

As mentioned before, the process by which ECM elements are degraded by malignant cells is called proteolysis and it can be seen as a form of ECM remodeling. The malignant cells produce several, ECM-specific membrane-degrading enzymes, the proteases, the protein integrin, which activates proteases, and proteins that inhibit the interaction between the proteases and integrin. This inhibition is important, as too much proteolysis removes the substrate that the cells adhere to, thereby impeding their movement [15]. The cancer cells also secrete factors which motivate non-malignant inflammatory cells within their environment to increase their expression of proteases, recruiting them as helpers for their invasion [36]. Ironically, another accomplice of the cancer cells in their mission of degrading the ECM is the ECM itself: the ECM stiffens as a result of the tumor-induced collagen cross-linking, and this stiffening is a signal for both cancer and healthy cells, to excrete more proteases, leading to accelerated ECM decomposition. Reducing the cancer-provoked ECM stiffness and related factors are a current object of research in oncology [37].

In many cases, only the basement membrane needs to be disintegrated to allow cancerous cells to invade, not the whole ECM. The basement membrane is part of the ECM and works like a layer that separates tissue compartments, but of course it also is involved in cell-communication processes. Before single or multiple cancer cells can start their movements away from the original tumor, they need to disintegrate part of the basement membrane that has developed around the tumor itself as well. [15]

Proteolysis can also help cancer cells during the process of migration, not only as a preparation. ECM degradation does not only create a pathway, it also releases growth factors that have been bound to the ECM, which could increase invasion speed, creating a vicious cycle that leads to faster cancer spread [38]. Additionally, the degradation of ECM molecules at the back of the cell facilitates its release from the substrate. Cancer cells are using this effect by having their focal adhesions synthesize proteases [39].

2.2.6. Collagen Cross-Linking

Next to proteolysis, collagen cross-linking is another form of ECM remodeling. It is an important process for healthy tissue function, as it provides more stiffness to the ECM which is necessary for certain organs. Skeletal deformities can be a consequence of a collagen cross-linking deficiency. However, excessive collagen cross-linking has unfortunate consequences as well and has been observed during tumorigenesis: pre-malignant tissue has been observed to exhibit unusually high amounts of the amine oxidase lysyl (LOX), which serves as a catalysator of the cross-linking of collagen. This surplus of LOX is

expressed by normal, non-malignant cells, hardening the tissue and thereby creating optimal conditions for the growth of pre-malignant cells. Stiffer tissue leads to more focal adhesions, manifesting the position of the pre-malignant cells. Once these cells have fully transitioned into tumor cells, they secrete excessive amounts of LOX themselves, further stiffening the ECM. [11, 40]

3. Mathematical Models of Cell Invasion

The previous chapter provided some insight into the highly complex world of cell and ECM interactions. The list of properties and mechanisms that was presented is not exhaustive, and many processes are yet to be discovered. This however does not prevent mathematicians from developing models, either working with what is observable in experiments and casting data into a set of formulae (the top-down approach), or by mathematically expressing a selection of biochemical and/or physical mechanisms which they consider most important, and combining them in a system of equations (the bottom-up approach).

All of the models that are discussed in this work in detail are bottom-up models. They bear the advantage that most behaviors of the system can be traced back to the underlying mathematical process, which allows to draw conclusions about the connected biological process. Which are these processes, that mathematicians, physicists or bioengineers consider most important for cell invasion? Which types of models prevailed, and how complex are they?

3.1. Historic Overview

Cell invasion as a matter of mathematical modeling is not a new topic, but the possibilities of research have not been exploited, given the numbers of publications. Just to give an example, a search in the PubMed-database [41] for the query ("cell invasion" OR "tumor invasion" OR "tumour invasion") AND "mathematical model" leads to the result statistics shown in Fig. 3.1. Of course it depends on the database, but given that PubMed tends to feature lots of publications from life science mathematics, it is interesting that only about 60 results could be found. However, it is not surprising that the first article covered by PubMed is from the mid-1990s, as it was the time when in general the number of publications of mathematical papers on cancer-related topics remarkably increased [42]. In the following, a small selection of mathematical cell invasion models is presented to provide some context to the present work. A more detailed timeline can be found in appendix A.1.
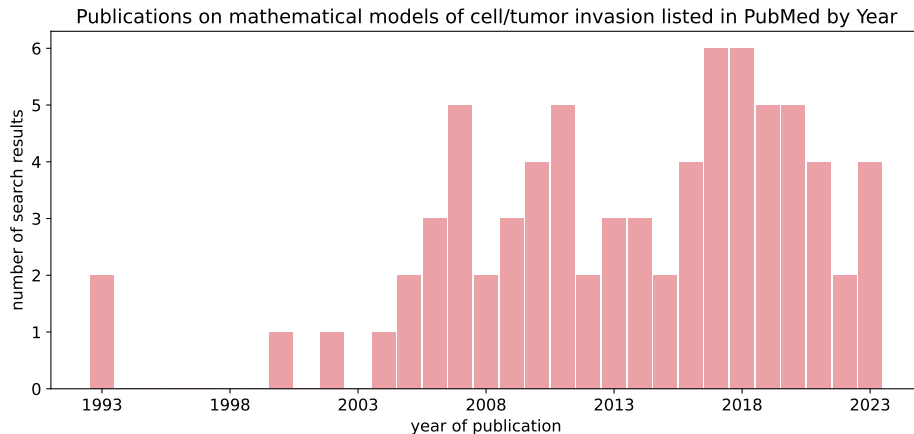
Figure 3.1.: Number of PubMed-listed articles per year that match the query ("cell invasion" OR "tumor invasion" OR "tumour invasion") AND "mathematical model". The data was collected in January 2024.

The history of cell invasion modeling starts with the modeling of something else – as it is so often the case. In 1937, Ronald A. Fisher developed a reaction-diffusion model describing the replacement of a gene by an advantageous mutation within multiple generations [43]. In the same year, his model was generalized by Andrei N. Kolmogorov, Ivan Petrovskii and Nikolai S. Piskunov, creating the famous FKPP model [44]. Even though its original field of application was genetics, this model has been – and still is – widely used for various biological applications.

The feature that makes this model so popular is that it allows for travelling wave solutions. Following James D. Murray's definition, travelling wave solutions are solution curves that move over time without changing their shape. The waves that are important for this work are of sigmoid shape, but other shapes such as bells are possible as well. Mathematically expressed, the solution $u(x, t)$ of a differential equation is called travelling wave if $u(x, t) = u(x - vt)$, where v denotes the constant velocity at which the wave moves in positive x-direction. Travelling wave solutions are characteristic for invasion processes and have been discovered for instance in chemical reactions, epidemics and species territorial invasion [27].

Perhaps the first mathematical model developed specifically for cell invasion was described by Robert A. Gatenby in 1995. It is a classic, two-dimensional logistic growth model that calculates the densities of healthy and malignant cells under competition for space and nutrients [20, 45]. Together with his colleague E. T. Gawlinski, he refined the model one year later and included the process of healthy tissue degradation through an acidic substrate generated by the malignant cells. Their model shows how an "interstitial gap"

builds at the invasive front of the tumor, creating space for the cancer cells to migrate into [20, 46].

Continuum models are not the only kind of cell invasion models. In 1999, Emma L. Stott used the so-called Cellular Potts model (CPM) for the first time in the context of tumor growth, together with her doctoral advisor Nicholas F. Britton and the physicists James A. Glazier and Marc Zajac. The CPM is a discrete stochastic model of cell movement where the cells' main objective is to save energy. This model is discussed in more detail in section 3.4.1. Stott and colleagues considered a spherical avascular tumor in a three-dimensional domain and included four types of cells: healthy cells, proliferating tumor cells, quiescent tumor cells and necrotic cells. Their model features tumor cell proliferation, nutrient availability and diffusion, and the decay of dead cells. The kind of growth they considered is called benign tumor growth, meaning that no healthy tissue is being destroyed, it is rather pushed away. With their simulation, they were able to reproduce the characteristic Gompertz growth curve. This curve displays "an initial exponential growth phase, followed by a deceleration of the growth rate and a final, steady-state size of the tumor due to exhaustion of growth resources" [47, 48].

In 2014, René F. M. van Oers and colleagues developed a model for angiogenesis (the sprouting of blood vessels), which bears a resemblance to cell invasion as it requires endothelial cells to detach from their original site, induce degradation of capillary basement membrane and to migrate toward an angiogenic stimulus [49]. However, in their model they neglected degradation and oriented attraction. Instead, they focused on cell-cell interactions and single cell responses to ECM mechanics, creating the first mathematical model of durotaxis. The model is hybrid, as it combines a two-dimensional CPM with a finite element method (FEM) description of the surroundings: the cells are discretely modeled using the CPM, and their movement is mimicked by augmenting or diminishing the space occupied by the cells, depending on the passive forces acting on them and on the active forces exerted by the cells themselves. The FEM model explains the cell-induced deformations of the ECM, which is considered to be an isotropic, uniform, and linearly elastic material that remains unchanged regarding its local density, but exhibits a cell-dependent stiffness gradient due to strain-stiffening [50]. Two researchers from van Oers' team, Elisabeth G. Rens and Roeland M. H. Merks, continued the study of the durotaxis model and published their extension in 2019: with the inclusion of focal adhesions, they can describe the cell spreading behavior governed by the ECM stiffness [35].

The relevance of the classic model by Gatenby and Gawlinski is proved by Alexander P. Browning and colleagues in 2019. They used a version of their model and calibrated it quantitatively to experimental data – another milestone in cancer invasion modeling, as it is the first such calibration. They estimated three parameters using a Bayesian sequential

learning approach: the cancer cell proliferation rate, their diffusivity and the rate at which the surrounding tissue is degraded. Even though they do not differ between healthy tissue cells and ECM elements, their model can be understood as a two-component model for cancer cell and ECM element density. [51]

Browning et al.'s model was simplified two years later by Chloé Colson and colleagues. They created a very simple continuum model: while Browning et al. assume cancer cell proliferation to be dampened by the presence of both, other malignant cells and healthy tissue, Chloé Colson and colleagues assume that the healthy tissue does not constitute an obstacle for the invading cancer cells. [52]

Continuum models ignore the individual particles (or cells) completely, and the CPM disassembles the particles into even smaller units – is there a middle ground that considers just the particles?

In 2023, Rebecca M. Crossley and colleagues did exactly that: they derived a continuum model that includes stochastics on the particle-level by coarse-graining an agent-based model. They considered this necessary, as probabilistic effects significantly impact diffusion in the case of rather few diffusing particles [53]. On the single-cell level, volume-filling is the most important factor of influence: the domain is assumed to be a one-dimensional grid, a lattice, and each lattice site is occupied by a certain number of cells and ECM elements. The cells can move to the adjacent lattice site or proliferate and put their offspring in the same lattice site they are in, but in each lattice site the number of occupants – may it be cells or ECM elements – must not exceed an upper bound N. The probability of a cell to change its location or to proliferate shrinks as the target site of this operation becomes more crowded. Cells can also degrade ECM elements at a certain rate, but only if the ECM element is located in the same lattice site. [54]

In the summary above, four levels of model complexity can be distinguished: the most simple form are the continuum models, as the one by Colson, to pick a more recent example. They are followed by agent-based models like Crossley's, and the CPM-based models, such as Stott's. The most complex type of cell invasion models are hybrid models such as the one by van Oers. In this work, three models of different complexity levels are compared and the relationship between their detailedness and the outputs they produce is studied. How much complexity is necessary to capture the main dynamics of cell invasion?

3.2. Crossley's Agent-Based Continuum Model

Rebecca M. Crossley and colleagues derived results similar to typical continuum models coming from the agent-based point of view [54]. While many other continuum models are derived from general invasion processes, this is a rather unusual approach of developing a relatively simple model while still incorporating mechanisms on the single cell level, making it worth to retrace its derivation at least partially.

Before delving into the equations, one word about agent-based modeling: in contrast to continuum models, agent-based modeling focuses on the objectives of the single particles (the agents) in the system, may it be cells, birds, or voters. There is no rule for the system's dynamics on a higher level – all major trends that become observable beyond a critical number of agents are a result of the sum of activities on the agent-level [55]. Assuming this critical number of agents is reached, these resulting collective effects can then be translated to continuum models.

In the following, some terms and explanations from Paul R. Taylor et al.'s article on compartment-based diffusion models [53] are used. As stated before, for the agent-based model the domain is assumed to be a one-dimensional evenly spaced lattice, and each lattice site i can have at most N occupying agents, i.e. cells or ECM elements. While ECM elements are immobile, cells can move but only to the neighboring lattice site and only if there is still space. Cells can also proliferate, but only if there is still space in their own lattice site for the daughter cell. And lastly, a cell can only degrade ECM that is situated in the same lattice site. Denoting a specific realization (a sample) of the model with j, the number of cells and ECM elements in lattice site i at time t is given by $u_i^j(t)$ and $m_i^j(t)$, respectively. With this notation, the scenarios described above can be quantified.

The probability of a cell to attempt a movement during a time step τ is given by $p_{\mathrm{m}} \in [0,1]$, and it is equally likely that the target site is the left lattice site $i-1$ or the right one $i+1$. The occupancy level at the target site decides whether the move is successful: movement success is inversely proportional to occupancy, leading to the following probability of movement to the left $(T_{i-}^{\mathrm{m}^j})$ or the right $(T_{i+}^{\mathrm{m}^j})$ in the time interval $[t, t+\tau)$:

$$T_{i\pm}^{\mathrm{m}^j}(t) = \frac{p_{\mathrm{m}}}{2}\left(1 - \frac{u_{i\pm1}^j(t) + m_{i\pm1}^j(t)}{N}\right).$$

The same principle holds valid for attempts of a cell to proliferate. The probability that a cell tries to create an offspring in time step τ is given by $p_{\mathrm{p}} \in [0,1]$, and therefore the probability of proliferation $T_i^{\mathrm{p}^j}$ during $[t, t+\tau)$ is

$$T_i^{\mathrm{p}^j}(t) = p_{\mathrm{p}} \left(1 - \frac{u_i^j(t) + m_i^j(t)}{N} \right).$$

The ECM degradation is not affected by volume-filling effects, it only depends on the number of cells present in lattice site i and the probability that one cell degrades one ECM element $p_{\mathrm{d}} \in [0, 1]$. Hence, the number of ECM elements $T_i^{\mathrm{d}^j}$ that are degraded during $[t, t + \tau)$ is given by

$$T_i^{\mathrm{d}^j}(t) = p_{\mathrm{d}} u_i^j(t).$$

Note that $T_i^{\mathrm{d}^j}$ is not defined as a probability, in contrast to $T_{i\pm}^{\mathrm{m}^j}$ and $T_i^{\mathrm{p}^j}$. Crossley does this to facilitate the subsequent derivation of the continuum formulation for the ECM degradation.

With these quantities, the agent-based model's formulation is concluded. Now the objective is to derive a system of partial differential equations (PDEs) using a so-called coarse-grained description of this model. Coarse-graining means that instead of considering every single cell or ECM element, one groups these agents – here conveniently by the lattice site they are in – leading to a reduced number of degrees of freedom in the system while maintaining the model's key physical features [56].

The exact steps performed in the coarse-graining can be found in appendix A.2. The method yields a parameterized system of PDEs, where the evolution of the cell and ECM density is described spatially for $x \in \mathbb{R}$ and temporally for $t \in (0, \infty)$. The equation for the cell density $u(x, t)$ is given by

$$
\begin{aligned}
\partial_t u(x, t) = D &\left(\partial_{xx} u(x, t) + \frac{u(x, t)\partial_{xx} m(x, t)}{K} - \frac{m(x, t)\partial_{xx} u(x, t)}{K} \right) \\
&+ r u(x, t) \left(1 - \frac{u(x, t) + m(x, t)}{K} \right),
\end{aligned}
\tag{3.1}
$$

where D can be interpreted as cell diffusivity, K is the carrying capacity and r is the proliferation rate or probability. The ECM density $m(x, t)$ can be described in a more simple manner, its equation is given by

$$\partial_t m(x, t) = -\lambda u(x, t)m(x, t), \tag{3.2}$$

where λ can be regarded as a degradation parameter.

To finalize their model, Crossley and colleagues remove the dimensions from the equations (3.1) and (3.2) by introducing the following substitutions:

$$\tilde{u} = \frac{u}{K}, \quad \tilde{m} = \frac{m}{K}, \quad \tilde{t} = tr, \quad \tilde{x} = \sqrt{\frac{r}{D}}x.$$

With these – and a bit of rearranging (3.1) using the product rule and the additional substitution $\tilde{\lambda} = \lambda K/r$ – one obtains the non-dimensional model

$$\partial_{\tilde{t}}\tilde{u} = \partial_{\tilde{x}}\left((1-\tilde{m})\partial_{\tilde{x}}\tilde{u} + \tilde{u}\partial_{\tilde{x}}\tilde{m}\right) + \tilde{u}(1-\tilde{u}-\tilde{m}),$$
$$\partial_{\tilde{t}}\tilde{m} = \tilde{\lambda}\tilde{m}\tilde{u}. \tag{3.3}$$

Crossley et al. constrain the diffusion term at $x = 0$, i.e. they postulate:

$$\left(1-\tilde{m}(0,\tilde{t})\right)\partial_{\tilde{x}}\tilde{u}(0,\tilde{t}) + \tilde{u}(0,\tilde{t})\partial_{\tilde{x}}\tilde{m}(0,\tilde{t}) = 0.$$

This can be interpreted as "building a wall" at $x = 0$, preventing any cells to leave the boundary towards negative x-values. In practice, this means that the domain is now constricted to $x \geq 0$.

Additionally, Colson et al. require $\tilde{u}, \partial_{\tilde{x}}\tilde{u} \to 0$ for $\tilde{x} \to \infty$ and set the initial conditions

$$\tilde{u}(\tilde{x},0) = \tilde{u}_0(\tilde{x}) \geq 0, \quad \tilde{m}(\tilde{x},0) = \tilde{m}_0(\tilde{x}) \geq 0, \quad 0 \leq \tilde{u}_0(\tilde{x}) + \tilde{m}_0(\tilde{x}) \leq 1 \; \forall \tilde{x} \in \mathbb{R}.$$

3.3. Colson's Minimal Model

Shortly before Rebecca Crossley constructed her model, another mathematician took on the challenge to model cancer cell invasion in one dimension [52]. Cloé Colson considers the same quantities as Crossley, cancer cells and ECM elements, and develops a reaction-diffusion model without the "detour" via the agents. Her goal is to set up a model that makes it easy to study the cross-diffusion effects of cells and ECM with analytical tools, leading to a relatively simple set of PDEs for $(x,t) \in \mathbb{R} \times (0, \infty)$:

$$\partial_t u(x,t) = \partial_x \left(D \left(1 - \frac{m(x,t)}{m_{\max}} \right) \partial_x u(x,t) \right) + ru(x,t) \left(1 - \frac{u(x,t)}{K} \right),$$
$$\partial_t m(x,t) = -\lambda u(x,t)m(x,t), \tag{3.4}$$

where u, m are the cell and ECM density, respectively, D is the diffusivity of cells when no ECM is present, $m_{\max}$ is the ECM density at which the cells are unable to degrade and invade it, r is the proliferation rate, K is the carrying capacity, and λ is the ECM degeneration rate. A summary of the parameters of both PDE models can be found in Tab. 3.1.

Parameter notation	Parameter name	Model
λ	ECM degradation rate	both
r	proliferation probability	both
D	cell diffusivity	both
K	carrying capacity	both
$m_{\max}$	unconquerable ECM density	Colson
$\hat{m}_0$	initial ECM density	both
α	initially invaded region	Crossley
σ	initially invaded region	Colson
ω	smoothness of initial invasion	Colson

Table 3.1.: Parameter definitions of the parameterized versions of Crossley's and Colson's models. The parameters in the last four rows correspond to the initial conditions of the numerical solutions, see appendix A.3 for more details.

The similarities and differences between Colson's and Crossley's models become more apparent when considering the dimensionless version of (3.4):

$$\partial_{\tilde{t}}\tilde{u} = \partial_{\tilde{x}}\left((1 - \tilde{m})\partial_{\tilde{x}}\tilde{u}\right) + \tilde{u}(1 - \tilde{u}),$$
$$\partial_{\tilde{t}}\tilde{m} = -\tilde{\lambda}\tilde{u}\tilde{m}, \tag{3.5}$$

where $\tilde{\lambda} = \lambda K/r$ [52]. While the equation for $\tilde{m}$ is the same as in (3.3), the reaction-diffusion equation for $\tilde{u}$ differs: in Crossley's model, both, the diffusion term $\partial_{\tilde{x}}((1 - \tilde{m})\partial_{\tilde{x}}\tilde{u} + \tilde{u}\partial_{\tilde{x}}\tilde{m})$ and the reaction term $\tilde{u}(1 - \tilde{u} - \tilde{m})$ include volume-filling effects due to cells and ECM, whereas Colson's model only incorporates the filling by the cell population [54]. Crossley compares the wave speed of the travelling wave solutions of the models (3.3) and (3.5), and finds that the additional volume filling effects through the ECM density slow down the invasion process, with the reaction term having a stronger influence on wave speed than the diffusion term [54]. The travelling wave solutions and their differing speeds are shown in Fig. 3.2. For more details on the numerical solution of the PDEs see appendix A.3.

3.4. Van Oers' Hybrid Model

While Crossley and Colson developed their models specifically for cell invasion, the biomedical engineer René F. M. van Oers built his hybrid model for the network-formation of endothelial cells which happens during the growth of blood vessels [50]. As it has been mentioned earlier, this process is similar to cell invasion, hence with a few extensions his model can be used for a very detailed description of mesenchymal, durotactically guided
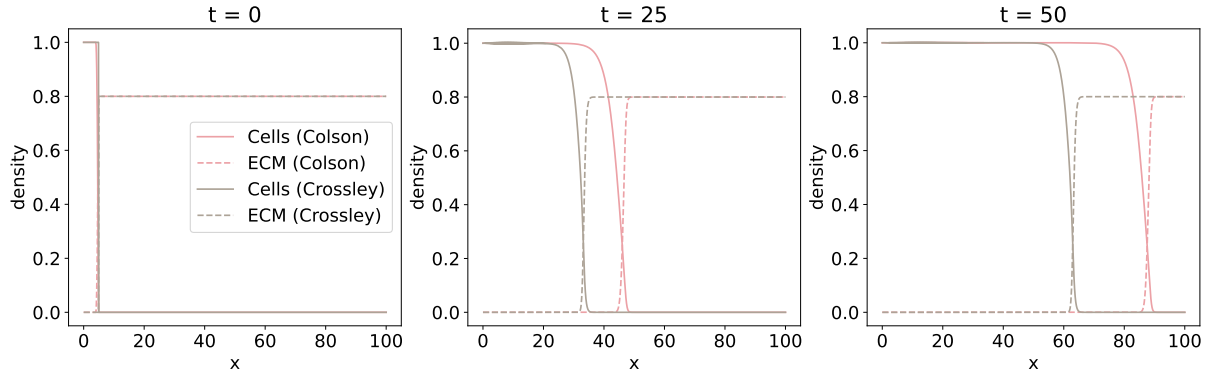
Figure 3.2.: Numerical solutions to the dimensionless models of Crossley and Colson for $t = 0$, 25, 50. Over time, the cell front in Colson's model becomes faster than the one in Crossley's model. (Parameter values: $\tilde{\lambda} = 10$; parameters of the initial conditions: $\hat{m}_0 = 0.8$, $\alpha = \sigma = 5$, $\omega = 1$. See appendix A.3 for more information about the parameters of the initial conditions.) Remark: At $t = 25$, Crossley's trajectory for the cells exposes slight oscillations close to the left domain boundary. This is because for certain $\tilde{\lambda}$-values, the model equations become stiff, making the numerical solver unstable. In the case of parameterized versions of the models, this effect is not as pronounced.

invasion in the two-dimensional domain. In this section, the original model by van Oers et al. are explained and the necessary extensions – ECM degradation and cell proliferation – are introduced.

3.4.1. The Origins of the Cellular Potts Model

A possibility to translate the interactions of cells into numerical operations was found by François Graner and James A. Glazier in the early 1990s [57]. Their Cellular Potts model (CPM) is a modification of the Potts model, which itself is a generalization of the (Lenz-)Ising model. The (Lenz-)Ising model was originally designed for ferromagnetism in cristals, using statistical mechanics and physics, but it has become very popular in the investigation of problems in the field of mathematical biology as well.

It might be helpful for this work's purposes to have a closer look at the development of the (Lenz-)Ising model: about 100 years ago, the electron's spin was discovered and was suspected to cause ferromagnetism, i.e. the property of certain metals to become magnets while they are in the proximity of a permanent magnet or lie inside a magnetic field. The idea was that as soon as the metal is put under the influence of such a magnetic field, all electron spins start to orient in the same direction – this alignment is the reason for magnetism in permanent magnets. The issue with this theory was how all electron spins

could be aligned in such a way, considering how each electron only influences its direct neighbors. The physicist Wilhelm Lenz proposed a model in which the molecules are located on the nodes of a (one- or multi-dimensional) grid and each molecule possesses a spin value $+1$ ("spin up") or -1 ("spin down"). He also assumed that the interaction energy between molecules depends only on the spin values of neighboring nodes of the grid, meaning that if a molecule is not a direct neighbor of another, the influence they can exert upon each other is zero. But if they are direct neighbors, there is a tendency that their spins will become the same, and the probability of this to happen depends on their interaction energy (and also on the strength of the external magnetic field). [58]

Lenz' student Ernst Ising analyzed the one-dimensional model in his dissertation and found that in the case of the one-dimensional grid, the uniform state (where all molecules have the same spin value) is unstable: if a single spin in the middle of the grid is changed by coincidence, as it can happen through thermal fluctuations, this could have enough power to change the spin value of the neighbors, and their neighbors, and so on, until no clear orientation which explains magnetisation is detectable [59]. However, this is only the case for the one-dimensional model: years later, Rudolf Peierls introduced the idea of boundaries between areas of $+1$- and -1-spins in the two- and three-dimensional model, which inhibit boundary-crossing change of spins at a sufficient probability. If the area or volume enclosed in these boundaries is small enough, there might be just enough molecules with the same spin value to create net magnetisation. Even though his proof contained an error, he provided a motivation for other scientists to review the model and soon after, Lars Onsager was able to solve the two-dimensional problem, proving that the (Lenz-)Ising model can be used as an explanation for ferromagnetism in the two-dimensional case. [60]

Some years later, in 1951, Renfrey B. Potts introduced the Potts model in his dissertation. He created two different versions of his model, the vector Potts model and the standard Potts model. In this work, the standard Potts model is considered since this is the one that is used for the CPM in most of the literature cited here. Both models assume that there are not only two spin values for each molecule, but $q \in \mathbb{N}$. If the number of possible spin values q is very large, one also talks about the "large-q Potts model". Once again, only the nearest neighbors are considered as factors of influence, and the likelihood for a specific molecule to change its spin value from q_i to q_j increases if there are many neighbors with the spin value q_j around (as it is the case for the (Lenz-)Ising model). [61]

As mentioned earlier, the physicists Graner and Glazier used the two-dimensional Potts model to describe cell sorting, a mechanism by which mixed embryonic cells of two different types find cells of the same kind and move closer to them. In the end, all cells of the same type are clustered together. As in the Potts model, their model assumes the grid

nodes to contain spin information, but since each node represents a partition of a cell, one usually finds clusters of nodes with the same spin value. More precisely, let there be N cells on the lattice, and let each cell have a size of five nodes, or more generally, of five lattice sites. Each cell has a unique spin value from 1 to N so that it can be identified (that is why spin values in the CPM can be interpreted as cell identification numbers). Then, one finds N clusters consisting of five neighboring lattice sites with the same spin on the grid. Lattice sites without any cells on them have spin value 0. Graner and Glazier also assign one of the two cell types to each of the occupied lattice sites. In the original CPM, energy is necessary for two things: the attachment between two cells of different type costs much more energy than the attachment between cells of the same type, and cell area changes are possible (e.g. from five lattice sites to four or six), but they cost more energy than maintaining the original, "target" cell size. Other factors that require energy can be introduced as well, such as resistance against an attracting substance. The goal of the system is to minimize the necessary energy by randomly choosing cell motions and size expansions or reductions and evaluating whether the chosen operation lowers the overall energy. Counterintuitively, if the system energy increases after the chosen operation, it is still accepted at a certain probability – the so-called Boltzmann probability – depending on the temperature. This is due to the fact that increasing temperature makes it more likely that the cells dissociate, meaning that they separate from the surface freeing them from dedicating any energy to the attachment to other cells or to the maintenance of their occupied area.[1] [57]

To summarize, with the CPM the relationship between cell motion and tissue properties such as pressure, temperature or the presence of attractors or repellents can be computationally investigated. [47]

3.4.2. Van Oers' Modification of the CPM

Now it is time to see how van Oers and colleagues combined the CPM with a finite element method (FEM) to simulate mechanical interactions between cells and ECM. Conveniently, both methods require a grid-like spatial domain so their combination does not need any transformations. The grid is assumed to be a square filled with square-shaped elements, each element having four nodes (one at each vertex). Note that in the following, "element" and "lattice site" are used synonymously to adhere to the common terms of the FEM and the CPM, respectively. In van Oers' version of the CPM, only one cell type is present, so there are no energy differences depending on which cells adhere

---

[1]Some authors, such as van Oers or Szabò, do not use temperature as a parameter, instead they use a dimensionless factor called "intrinsic cell motility" which accounts for random cell movements and has the same purpose as temperature in the original CPM. [47, 50]

to each other, but it is assumed that cell-cell adhesions require more energy than cell-ECM adhesions. The other factor that influences the system energy is the same as in the original CPM, namely changes in cell size. Van Oers et al. describe the overall system energy with the following Hamiltonian:

$$H = \sum_{\sigma \in \text{cells}} \lambda \left(\frac{a(\sigma) - A(\sigma)}{A(\sigma)} \right)^2 + \sum_{(x,x')} J(\sigma(x), \sigma(x')) \cdot (1 - \delta(\sigma(x), \sigma(x'))),$$

where the first sum captures the energy necessary to maintain a certain cell volume (or area), and the second sum describes the contact energy between the cells. In more detail, x and x' are pairwisely different lattice sites, $\sigma(x)$ is the spin (or cell identification number) of the cell that sits in lattice site x, λ is a parameter describing the elasticity of the cells, $A(\sigma)$ is the target volume of the cell (in the example above it would be $A(\sigma) = 5$ for all cells), $a(\sigma)$ is the volume of the cell in the current state of the system (it can be equal to $A(\sigma)$, but it can also be a bit higher or lower), $J(\sigma(x), \sigma(x'))$ is the necessary attachment energy between the cells at site x and x', or, if one of the sites, say x', is not occupied by a cell, i.e. $\sigma(x') = 0$, then J describes the cell-ECM adhesion energy. δ is the Kronecker Delta which makes sure that if x and x' are occupied by the same cell or if both are free of any cells, then the attachment energy between these lattice sites is zero. An overview of all parameters of the model and its extensions which are introduced later can be found in section 3.4.5.

As in the original CPM, after initially placing the cells on the grid, van Oers' model iteratively chooses a random site x and a neighbor x', and computes H for the scenario where $\sigma(x')$ becomes $\sigma(x)$, i.e. the cell with identification number $\sigma(x)$ extends the area it occupies and invades the area occupied by cell $\sigma(x')$. Of course it is also possible that one of the sites does not contain any cells, only ECM, or that both sites contain the same cell – all of these exceptions are covered by the formulation of H. The algorithm now compares the old Hamiltonian with the one for the potential new arrangement and evaluates this difference ΔH with the following rule:

$$\mathbb{P}(\Delta H) = \begin{cases} 1 & \text{if } \Delta H < 0 \\ e^{-\frac{\Delta H}{T}} & \text{if } \Delta H \geq 0, \end{cases} \tag{3.6}$$

where $\mathbb{P}(\Delta H)$ describes the likelihood for the switch from $\sigma(x')$ to $\sigma(x)$, and T is a constant for intrinsic cell motility, which can be imagined analogously to the diffusion coefficient in diffusion equations [62].

The last step of one iteration of van Oers' CPM consists of validating whether the switch violates a connectivity constraint: if the switch $\sigma(x') \to \sigma(x)$ splits the cell with iden-

tification number $\sigma(x')$, $\mathbb{P}(\Delta H)$ is set to 0. See Fig. 3.3 for a schematic depiction of the decision rule (3.6).
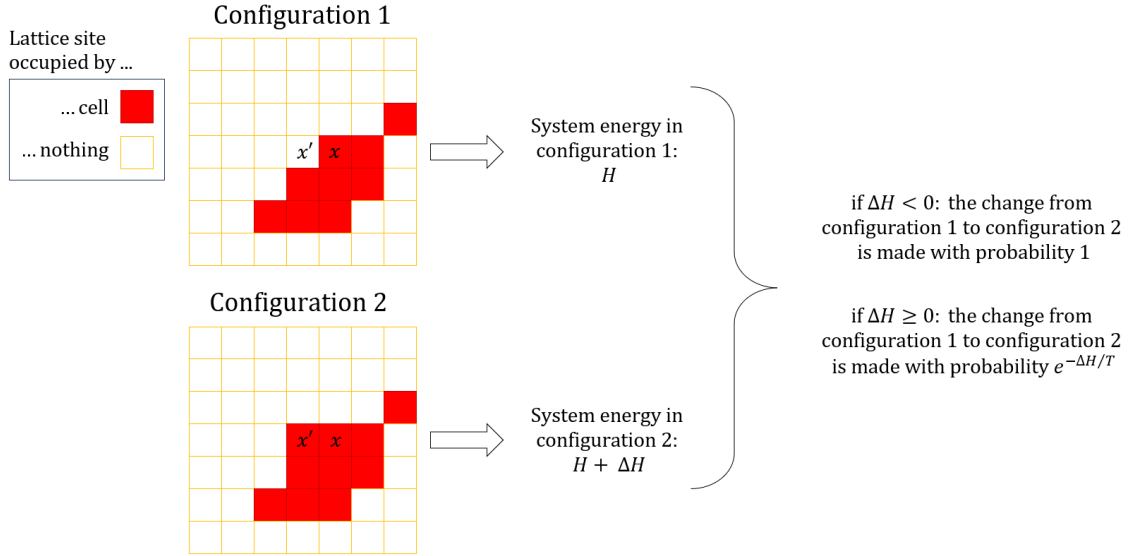
Figure 3.3.: Decision process if cell $\sigma(x)$ should expand to lattice site x', provided the expansion does not violate the connectivity constraint.

To gather a bit more theoretical background, the CPM is simulated using a Metropolis-Hastings algorithm. The same method is commonly used for the CPM's predecessors, the (Lenz-)Ising model and the Potts model. The Metropolis-Hastings algorithm is part of the class of Markov chain Monte Carlo methods which approximate complicated probability distributions by drawing large samples from simpler distributions that have a known relationship to the target distribution.

The random variables from this simpler distribution need to fulfill the Markov property, i.e. their probability needs to be memoryless: let $\{X^{(t)}\}_{t \in \mathbb{N}}$ be a family of random variables describing the state of a system, where $X^{(t)} \in \mathcal{X} := \{x_1, x_2, ...\}$. $\mathcal{X}$ is a finite or countably infinite set of states. Then, $\{X^{(t)}\}_{t \in \mathbb{N}}$ is called Markov chain if the transition probability from step $t - 1$ to step t is defined by

$$\mathbb{P}(X^{(t)} = x_1 | X^{(t-1)} = x_2, X^{(t-2)} = x_3, ...) = \mathbb{P}(X^{(t)} = x_1 | X^{(t-1)} = x_2),$$

i.e. the state $X^{(t)}$ only depends on the last state $X^{(t-1)}$. [63]

The CPM can be viewed as a Markov chain since the transition probability from the current configuration $X^{(t)}$ at time t to a different one $X^{(t+1)}$ in the next time step only depends on the state at time t, making it a memoryless process – hence satisfying the Markov property. Whereby "configuration" refers to the entity of spins of all lattice sites

in the system, and changing the spin of one site leads to a new configuration. Based on the current configuration, a possible new configuration is picked following a proposal density q. Sampling from q consists of two tasks, first selecting a lattice site from a uniform distribution, and secondly transferring its spin to a neighboring site. The last ingredient of the algorithm is the target density π, or at least a density proportional to it, which is denoted by $\tilde{\pi}$. The target density is equal to the Boltzmann-weighted configuration probability, i.e. the likelihood of a certain configuration to appear measured by the system energy induced by it (the lower the system energy, the more likely the configuration) [64]. The approximative target density is a function of the system energy $H(X^{(t)})$, (whereby $X^{(t)}$ is the configuration), and is given by $\tilde{\pi}(X^{(t)}) := \exp\left(-\frac{H(X^{(t)})}{T}\right) \propto \pi$. Eventually, by iteratively sampling from q and evaluating these samples using $\tilde{\pi}$, the goal is to find configurations that approximate samples from π, and thereby come closer to very low system energy.

With this interpretation and notation, it is possible to rewrite the above algorithm in a more compact form following Christian P. Robert's article on the Metropolis-Hastings algorithm [65]:

Algorithm Metropolis-Hastings CPM

Given $X^{(t)} = x^{(t)}$
1: Generate $Y^{(t)} \sim q(\ \cdot\ |x^{(t)})$

2: Choose $X^{(t+1)} = \begin{cases} Y^{(t)} & \text{with probability } \rho(x^{(t)}, Y^{(t)}) \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, Y^{(t)}) \end{cases}$

where $\rho(x, y) = \min\left\{\frac{\tilde{\pi}(y)}{\tilde{\pi}(x)} \cdot \frac{q(x|y)}{q(y|x)}, 1\right\}$

The decision rule defined by the function ρ is identical to (3.6): choosing the sample $Y^{(t)}$ starting from $x^{(t)}$ is just as likely as choosing $x^{(t)}$ starting from $Y^{(t)}$, so it is $q(x^{(t)}|Y^{(t)}) = q(Y^{(t)}|x^{(t)})$ and one obtains (using x for the old and y for the new configuration)

$$\frac{\tilde{\pi}(y)}{\tilde{\pi}(x)} \cdot \frac{q(x|y)}{q(y|x)} = \frac{\tilde{\pi}(y)}{\tilde{\pi}(x)} = \frac{\exp\left(-\frac{H(y)}{T}\right)}{\exp\left(-\frac{H(x)}{T}\right)} = \exp\left(\frac{H(x) - H(y)}{T}\right) \begin{cases} \geq 1 & \text{if } H(x) \geq H(y) \\ < 1 & \text{if } H(x) < H(y). \end{cases}$$

With the familiar definition of $\Delta H = H(y) - H(x)$ it can be concluded

$$\rho(x, y) = \begin{cases} 1 & \text{if } \Delta H \leq 0 \\ e^{-\frac{\Delta H}{T}} & \text{if } \Delta H > 0, \end{cases}$$

which is the same decision rule as van Oers defined it, with the only difference that the strictness of the inequalities is switched.

After one Monte Carlo step, by which n iterations of the above algorithm are meant (where n is the number of lattice sites), the first deformation in the ECM caused by the cells' movements is calculated using finite elements. In the following, some terms and explanations from Wolfgang A. Wall's lecture script on the FEM [66] will be employed.

As it has been mentioned earlier, the domain is a square filled with square-shaped elements. The setting is two-dimensional, hence each node has two degrees of freedom, one in x_1- and one in x_2-direction. A visualization of this local and global geometry of the domain is shown in Fig. 3.4. In total, assuming there are n_x elements per row (and $n_{x_2} = n_{x_1}$ elements per column), then in total there are $n_{x_1} + 1$ nodes per row (and column), leading to $(n_{x_1} + 1)^2$ nodes in the whole grid and hence to $2 \cdot (n_{x_1} + 1)^2$ global degrees of freedom. The objective of the FEM is to find the global vector $u \in \mathbb{R}^{2(n_{x_1}+1)^2}$ which contains the displacement of every degree of freedom, whereby u is the solution of the finite element equations given by

$$Ku = f, \tag{3.7}$$

where K is the global stiffness matrix and f is the force vector constituted by the forces from the cell motions. The global stiffness matrix is a collection of contributions from all elements: it is assembled by the local stiffness matrices of all elements (the local element stiffness matrix $k^{(e)} \in \mathbb{R}^{8 \times 8}$ is the same for each element).
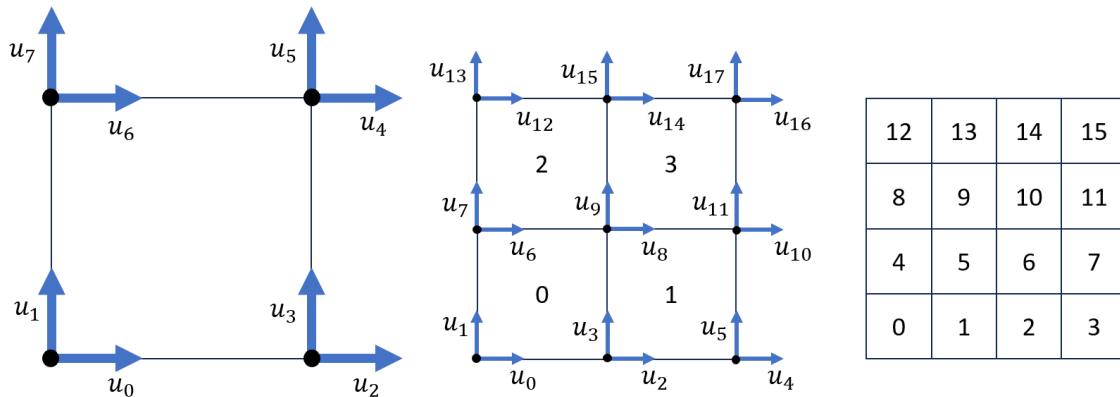
Figure 3.4.: Local and global FEM setting. Left: definition of the eight degrees of freedom in the local, element-wise domain. Middle: definition of the degrees of freedom in the global domain – note that when the elements share nodes, they also share the respective degrees of freedom. For instance, the horizontal displacement in the bottom right corner of element 0 is the same as the horizontal displacement in the bottom left corner of element 1, hence the shared displacement u_2. Right: the numeration of elements starts at the bottom left element and ends at the upper right element.

The matrix K in equation (3.7) is built in the setting of linear elasto-statics. The method is very common in mechanics and it is explained in detail in appendix A.4.

The definition of the force vector f in (3.7) is model specific, so it is explained here. f contains the externally applied loads, which can be interpreted as the cells' traction forces. As proposed by the biomedical engineer Christopher A. Lemmon and biologist Lewis H. Romer, cell shape and size are used to calculate the magnitude and direction of the forces they exert [67]. Let i and j be two nodes that belong to elements which are covered by the same cell. Then it is assumed that these two nodes pull on each other by a force proportional to their distance, $d_{i,j}$, such that the ensemble of all node forces within one cell point towards the center of the cell. Using μ as the proportionality constant, and Δx as the side length of one square element, a two-dimensional force vector F_i can be obtained for each node i:

$$F_i = \mu \Delta x \sum_j d_{i,j}. \tag{3.8}$$

After calculating F_i for all nodes i that belong to elements with cells on them and dissecting the F_is into the vertical and horizontal degrees of freedom, they constitute the global force vector f.

In summary, the cell forces f, together with the stiffness matrix K, dictate how cell movements deform the ECM. But how does the ECM influence the cell movements in return? For this purpose, van Oers and colleagues introduce durotaxis in their CPM: cells prefer ECM regions of higher stiffness, so it is assumed that migration towards high stiffness reduces the system energy compared to migration towards regions of lower stiffness. Regions in the ECM where stiffness is higher can by identified using the concept of strain-stiffening, a process during which material becomes stiffer the more it is deformed [68]. Once again, assume that a lattice site x and a neighbor x' were randomly chosen. Now the difference in system energy is calculated given that $\sigma(x')$ is changed to $\sigma(x)$, i.e. cell $\sigma(x)$ extends and invades the space of cell $\sigma(x')$. This difference not only consists in energy changes from cell-cell adhesions and size transformations, but now it also depends on the durotaxis factor:

$$\Delta H_{\text{durotaxis}} = g(x', x) \cdot \left(h(E(\varepsilon_1)) \cdot \langle v_1, v_d \rangle^2 + h(E(\varepsilon_2)) \cdot \langle v_2, v_d \rangle^2 \right).$$

The function g indicates whether cell $\sigma(x)$ expands or diminishes its size, multiplying $\Delta H_{\text{durotaxis}}$ with $g(x', x) = -1$ in case of expansion (the cell moves according to the durotactic gradient, a motion that requires less energy) and with $g(x', x) = 1$ in case of retraction (the cell opposes the durotaxis-induced motion, a resistance that needs more energy). The rest of the term combines three properties of the mechanical cell-

ECM interaction: preference for higher stiffness, strain-stiffening, and the orientation of stretch.

Stiffness is given by the function E (see below), and the preference for higher values of E is expressed via the sigmoid function

$$h(E) = \frac{\alpha}{1 + \exp(-\beta(E - E_{\text{tr}}))}.$$

α is the maximum value of h, β describes the steepness of the curve and E_{tr} provides the point at which $h(E_{\text{tr}}) = \frac{\alpha}{2}$. With this function van Oers captures how a certain level of ECM stiffness is necessary to motivate the cell to expand, but at some point a maximum is reached and higher values of E do not change the cells behavior anymore, as it is also described in section 2.2.3.

Strain-stiffening is described using the stiffness E. As mentioned before, the more deformed – or stretched – the material, the stiffer it becomes. Hence, stiffness is a result of stretch ε:

$$E(\varepsilon) = E_0 \cdot \left(1 + \frac{\varepsilon \cdot \mathbb{1}_{\varepsilon \geq 0}}{\varepsilon_{\text{st}}}\right),$$

where E_0 is the stiffness of the unstretched material, and ε_{st} describes how fast the ECM stiffens under stretch. The indicator function $\mathbb{1}_{\varepsilon \geq 0} = \{1 \text{ if } \varepsilon \geq 0, 0 \text{ else}\}$ ensures that only non-negative stretch, i.e. substrate extension, leads to stiffening, whereas negative stretch, i.e. compression, does not have any effect on the ECM.

The orientation of stretch plays an important role as "a strained ECM is stiffer along the strain orientation than perpendicular to it" [50], increasing the durotactic effect if a cell motion follows this orientation. Using the FEM solution, the strain tensor for each element is calculated. Its eigenvalues ε_1 and ε_2 yield the maximum and minimum stretch and they are used to compute the strain-induced maximum and minimum stiffness $E(\varepsilon_1)$, $E(\varepsilon_2)$. Their corresponding eigenvectors v_1 and v_2 yield their orientation inside the element. They are used for comparison with the direction of the cell motion which is given as a unit vector $v_d = (x' - x)/||x' - x||_2$, so that it can be said whether the motion follows the orientation of the stretch. The scalar products used for this comparison, $\langle v_1, v_d \rangle$ and $\langle v_2, v_d \rangle$, are squared so that only values greater or equal to 0 are possible. Cell movements v_d perpendicular to a stretch direction v_1 or v_2 are not influenced by the ECM stiffness.

A schematic depiction of the components of $\Delta H_{\text{durotaxis}}$ can be found in Fig. 3.5.

Closing the section on van Oers' original model, it should be mentioned that after each Monte Carlo step, the grid is assumed to be undeformed again, and the deformations are calculated starting from new with the next cell configuration. This simplification is
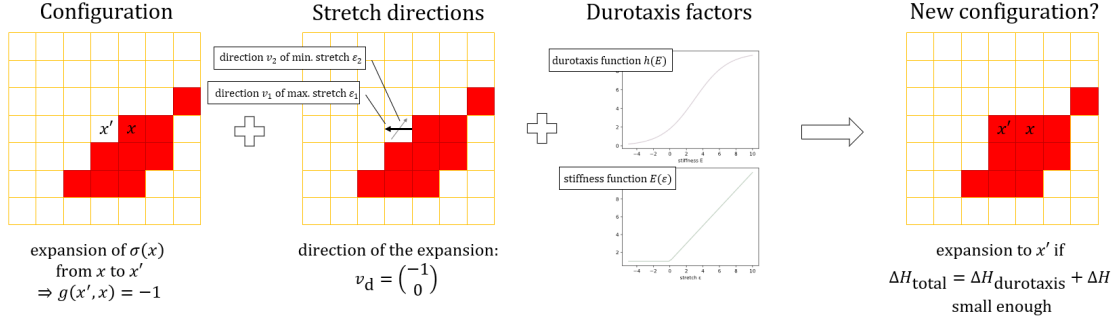
Figure 3.5.: Components of $\Delta H_{\text{durotaxis}}$ and their representation in the lattice. After $\Delta H_{\text{durotaxis}}$ has been computed, it has to be summed up with the Hamiltonian that captures energy expenditures or savings due to volume changes and cell-cell or cell-ECM adhesion. Then, the decision rule (3.6) can be applied to ΔH_{total}. The lattice coloring is the same as in Fig. 3.3.

used because otherwise, the stiffness matrix K can no longer be assumed to be constant, which would increase the computational expenses significantly. Nonetheless, the grid deformations still play an important role in the model as the deformation from the previous Monte Carlo step influences the Hamiltonian for the next step.

3.4.3. Extending van Oers' Model: ECM Degradation

So far, only one kind of actor is considered in van Oers model: the cells. In order to simulate ECM degradation, the ECM elements need to be included as well. They can be introduced in the CPM as a second type of cells with the properties of the ECM: for instance, the adhesion energy between a site occupied by a cancerous cell and another site occupied by an ECM "cell" is equal to the adhesion energy between the cell and the substrate. Also, ECM "cells" do not mechanically influence the surface.

Each ECM element occupies exactly one lattice site, making it possible to model the degradation process at a high resultion. When the ECM is initialized at the beginning of the simulation, its concentration can be regulated with a parameter $\hat{m}_0 \in [0,1]$, hence if $\hat{m}_0 < 1$, some lattice sites are still free of cells and ECM. In the following, let x denote a lattice site and $\tau(x)$ is the ECM element occupying it (analogous to the cell identification number $\sigma(x)$). Lattice sites where ECM elements are present cannot be visited by cells regardless of whether an expansion onto the ECM-occupied site leads to an energetic advantage. So, differently from the invasive CPM by Turner et al. (see a summary of their model in appendix A.1), the ECM is not regarded as a "soft barrier" by including its presence in the system energy Hamiltonian [62], but as an impenetrable wall that changes its shape as a consequence of each Monte Carlo step. Its shape is influenced by

the cells which secrete proteolytic enzymes. The update of the ECM element dissolution is guided by the following criteria, which are illustrated by Fig. 3.6:

- there is an area of influence $\mathcal{A}(\tau(x))$ surrounding each ECM element $\tau(x)$ in which the presence of cells and their proteases have an effect on the element. Cells outside of this area do not have any effect on the ECM element.

- within $\mathcal{A}(\tau(x))$, there is a gradient making cells far away from x less influential than cells closer to x. This gradient is determined by a function $f : \mathcal{A}(\tau(x)) \to [0,1]$ which maps the spatial cell distribution in the area of influence to the probability of degradation of $\tau(x)$.
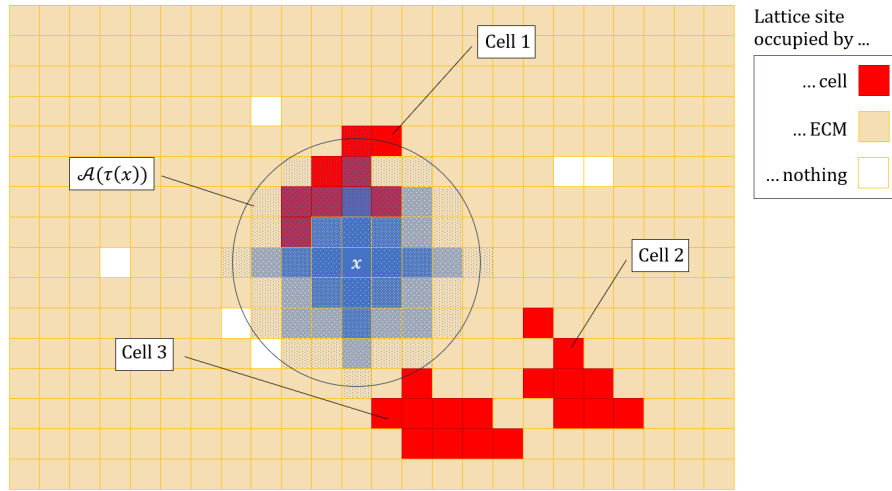
Figure 3.6.: The ECM element $\tau(x)$ at site x is surrounded by an area of influence $\mathcal{A}(\tau(x))$ with a radius of 4 layers (or elements). Lattice sites that are part of $\mathcal{A}(\tau(x))$ have a blue shading, and if they are part of the same layer, their shading is of the same intensity. Here, only lattice sites occupied by cell 1 influence the degradation probability of $\tau(x)$, cells 2 and 3 are too far away.

These properties are supposed to describe diffusion of proteases through the ECM and also through areas where cells are present. Of course it is a simplification to assume that these enzymes float through ECM just as they float through other cells. This limitation could be at least partially mitigated by the choice of f. f is not required to be a continuous function, hence one could select it such that it captures the different material properties of the entities present in $\mathcal{A}(\tau(x))$. Indeed, here f is not defined as a continuous function either. $\mathcal{A}(\tau(x))$ is divided in L "layers" of lattice sites, where the sites in each layer are assumed to have an equal influence on $\tau(x)$ if occupied by cells. In accordance with the rules defined above, this means that if site x_1 and x_2 are both part of the same

layer, their distance to x is equal or at least sufficiently similar. These layers can maybe best be understood as rings circling x. The influence of each layer i is represented by a weight $w_i \in [0, 1]$, increasing as the layers get closer to x, leading to the probability of degradation of ECM element $\tau(x)$

$$f(\mathcal{A}(\tau(x))) = \sum_{i=1}^{L} w_i \frac{\text{number of sites in layer } i \text{ occupied by cells}}{\text{total number of sites in layer } i}.$$

As f needs to produce values between 0 and 1, the additional restriction $\sum_i w_i = 1$ is made. Of course this still leaves a large variety of options for the choice of the weights. But as the function f is supposed to describe diffusion, it makes sense to choose weights that mimick the properties of diffusing enzymes. The concentration of diffusing particles decays exponentially as the layers are located further away from their source, hence the following choice of exponentially decaying weights seems reasonable:

$$w_i = a^{i-1} \frac{1-a}{1-a^L}, \quad i \in \{1, ..., L\},$$

where $a > 1$ is a parameter for the steepness of the decay. The sum of the w_i is equal to 1, which can be shown easily.

f is called after each Monte Carlo step for every ECM element and is then either degraded, making space available for nearby cells, or it remains unaffected.

3.4.4. Extending van Oers' Model: Cell Proliferation

To proliferate, cells must first double their size and then divide [69]. By their nature, cancer cells do not possess – or only possess very limited – endogenous mechanisms that control their growth and division, hence cancer cell growth and resulting division is mainly limited by space and nutrients.

Crossley et al. ignore nutrients and the changes in cell size altogether, making proliferation a purely space-limited, cell size independent process. The former kind of limitation might lead to similar results as the nutrient-dependent growth limitation: if the cells are spread on the surface with a lot of space between each other, diffusing nutrients can reach every cell equally well and all cells can easily divide. If the cells are clustered in a spheroid, only the cells at the outer rim have access to enough nutrients for proliferation, and also they are the only ones that have enough space to do so. The simplifying assumption that size is not of importance for cell division could however lead to quite different growth speeds and patterns than what is observed in nature.

To reduce computational expenses, no additional quantities are introduced, and proliferation is implemented to be only limited by space instead of nutrients, too. Just as Crossley and colleagues did, a proliferation probability p_{p} is included according to which each cell divides after every Monte Carlo step.

If a cell is selected to proliferate, it places a daughter cell next to it provided there is enough space right next to it. If there is not enough space available, it ceases to proliferate. Fig. 3.7 illustrates this decision rule. In the next Monte Carlo step, the same cell can be selected to proliferate again, and again it is tested whether there is enough space available. If eventually enough ECM has been degenerated or the cell has moved away far enough from its neighbors, a daughter cell can be put on the grid. This daughter is eligible for proliferation as of the next Monte Carlo step.
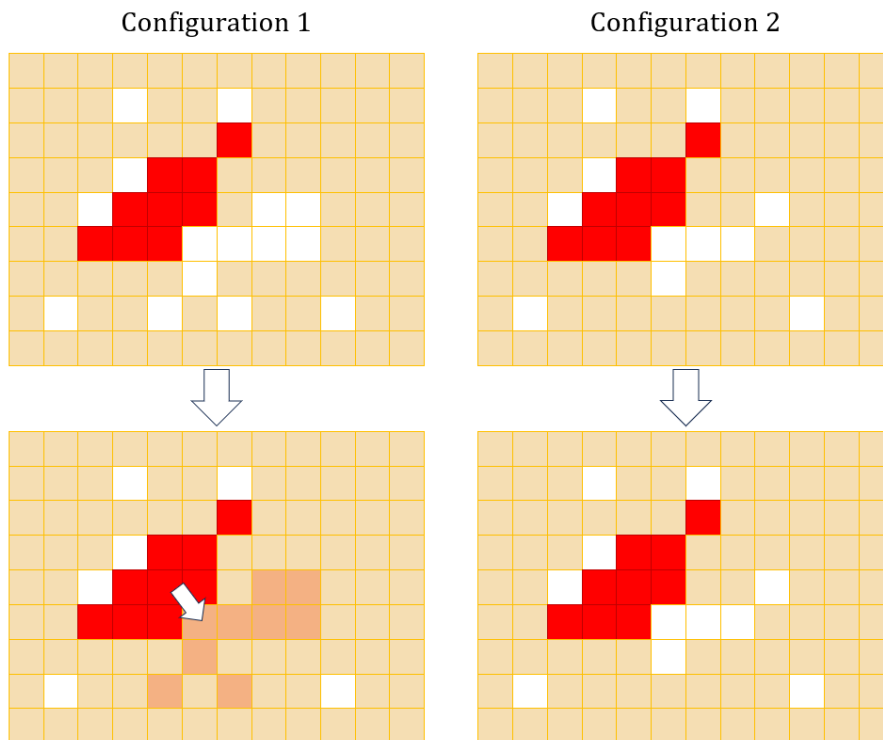
Figure 3.7.: In this example, the target size of a cell is $A(\sigma) = 9$ lattice sites. In both configurations, the cell in the top image is selected to proliferate. In configuration 1, there is a connected region of 9 unoccupied lattice sites right next to the cell, so it places a daughter cell there, see bottom image. In configuration 2, no such area is available and the cell does not proliferate, even though it was selected to do so. The lattice coloring corresponds to the one in Fig. 3.6.

3.4.5. Summary of the Extended Model and Exemplary Results

From now on, the modified van Oers-model is referred to as CPM-FEM model. An overview of its algorithm is given in Fig. 3.8, and the authorships of the various components of the code are listed in appendix A.5.
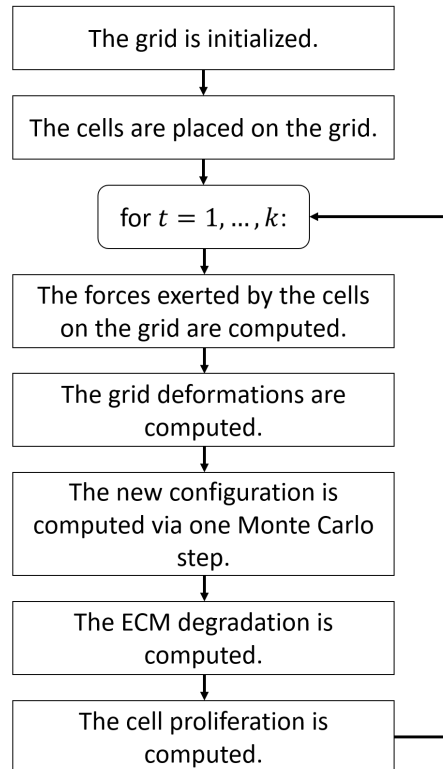
Figure 3.8.: Algorithm overview for the CPM-FEM model with k iterations.

The parameter notations and definitions of the CPM-FEM model are summarized in Tab. 3.2.

To provide an idea of what kinds of results are produced by the CPM-FEM model, the following examples are considered. First, the proliferation probability is set to zero such that the only assertive principle is the system energy. The grid is initialized with 100 by 100 elements and a spheroid consisting of 35 cells (its diameter is about 1/5th of the grid's side length) is placed in its middle. The rest of the parameters are chosen as in Tab. 4.1 on page 40 (more on the reasoning behind the parameter choices can be found in appendix A.6.1), the initial ECM density is set to $\hat{m}_0 = 0.8$, and the radius of the area of influence $\mathcal{A}(\tau(x))$ is chosen as one cell diameter.

The first row of Fig. 3.9 shows how the cells spread over the grid for 1000 iterations. Right after initialization, at $t = 0$, the cells' shapes are random as the only guiding factors are

Parameter	Parameter name	Parameter	Parameter name
A	target cell area	ν	Poisson's ratio
λ	cell elasticity	E	Young's modulus
J_{CC}	cell-to-cell attachment energy	E_0	reference stiffness
J_{CE}	cell-to-ECM attachment energy	$\varepsilon_{\mathrm{st}}$	strain-stiffening speed
T	intrinsic cell motility	α	max. durotactic guidance
μ	cell force constant	E_{tr}	durotaxis threshold stiffness
β	steepness of durotactic guidance	p_{p}	proliferation probability
$\hat{m}_0$	initial ECM density	initial number of cells
radius of the area of influence $\mathcal{A}(\tau(x))$		

Table 3.2.: Parameter definitions of van Oers' model (including the FEM parameters and the parameters from the extension).

$t = 0$ $t = 50$ $t = 200$ $t = 500$ $t = 1000$
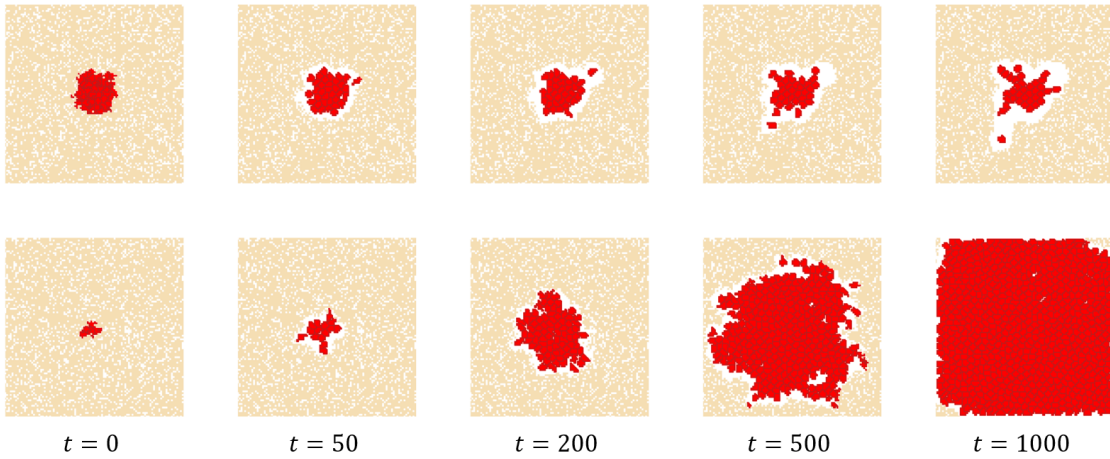
Figure 3.9.: Evolution of the CPM-FEM model for 1000 iterations. Top row: no proliferation. Bottom row: proliferation probability $p_{\mathrm{p}} = 0.001$. The lattice coloring corresponds to the one in Fig. 3.6.

connectedness and target size. At $t = 50$, their shape is already much more rounded because the inherent cell traction forces had time to act. The left hand side of Fig. 3.10 shows a close-up of this configuration including these forces, where the arrow length is proportional to the exerted force. It is visible that much stronger forces are acting within cells that have a lengthy shape (such as the cell at the top right that has escaped the spheroid), than within circular-shaped cells. This observation and its dependence on the substrate stiffness is one of the main findings of van Oers and colleagues [50]. As t progresses further, not only can one see how the surrounding ECM gets degraded more and more thoroughly, yielding more freedom to move for the cells, but it is also visible how the initial cell spheroid evolves towards a star by forming "fingers". These fingers are

the product of durotaxis, as it can be seen on the right hand side of Fig. 3.10: whenever cells are moving away from the spheroid, they leave a trace of strain behind them which motivates fellow cells to follow them. A lot of randomness is involved in this process, hence it does not always work as it can be seen in the renegade cell at the bottom left. It probably moved away from the initial site too quickly for other cells to follow.
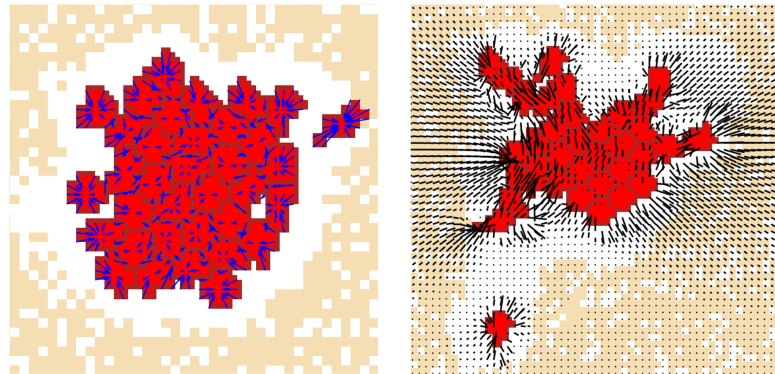
Figure 3.10.: Details of the example without proliferation. Left: close-up of the spheroid at $t = 50$ including the traction forces as blue arrows. Only forces above 0.001 nN are shown. Right: star-shaped spheroid at $t = 1000$ including the element-wise direction of maximum substrate stretch depicted as black lines. The length of the lines is proportional to the stretch magnitude.

Now, let the proliferation probability be elevated to 0.001. The number of cells that are on the grid at $t = 0$ is lowered to 5, and the rest of the parameter values is maintained. The result of this parameter choice is depicted in the bottom row of Fig. 3.9. Similarly to the scenario without proliferation it can be observed that the cells soon after initialization assume a rounded shape. While a little bit of durotactic finger formation could be assumed at $t = 500$, cell proliferation is without a doubt the factor that determines the outcome of this simulation.

Of course there are infinitely many parameter combinations to be tested, and the examples above are only an illustration of what kinds of results the CPM-FEM model can produce. The interested reader is referred to the Github repository `https://github.com/veronikahofmann/masters-thesis-cellinvasion`, where an implementation of the CPM-FEM model is available such that own parameter combinations can be tested.

4. Sensitivity Analysis

It is time to study the similarities and differences between the CPM-FEM model and the two PDE models that were introduced earlier. In this chapter, the effects of a selection of parameters on the model results are collected to see which parameter has the largest impact on the respective model, and how this impact changes over the temporal and spatial domain. This is done by performing a sensitivity analysis, which is a practice that studies "how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [70]. In other words, how strongly does varying one parameter change the resulting course of the invasion process? Are there interaction effects that transcend the impact of a single parameter? These questions are the focus of this chapter.

4.1. One-Dimensional Representation of the CPM-FEM Results

To make the results of the CPM-FEM model comparable to the ones by Crossley and Colson, a conversion of the two-dimensional grid to a one-dimensional domain is necessary. This is done by using a similar approach as for the ECM degradation in the two-dimensional model: starting from an element in the center of the plane, the surrounding elements are sorted into layers depending on their distance to the starting element. Each ring-shaped layer has the thickness of one element to make the resolution as high as possible. Of course, this method ignores elements close to the corners of the domain. As long as the domain is large enough, this limitation is accepted since the model initialization starts with a spheroid of cells in the middle of the grid, and the rim of the outermost layer is only reached after many iterations or in the case of extreme parameter choices. For each layer, the relative occupancy by cells and ECM elements is calculated, leading to a one-dimensional representation of cell and ECM density. This representation allows to perceive travelling wave solutions for the CPM-FEM model, as shown in chapter 5. See Fig. 4.1 for an explanatory chart and an example of the measure.
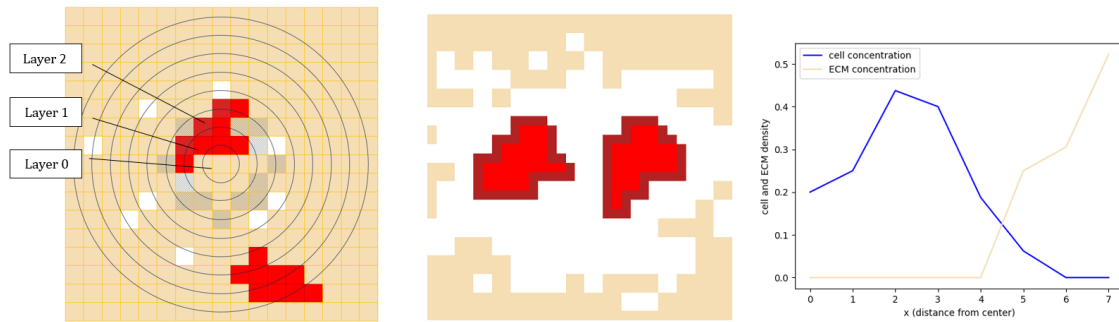
Figure 4.1.: Left: structural image of how the layers are constituted. The elements corresponding to layer 2 have a gray shading: if the middle point of an element lies inside the layer, then the whole element is part of the layer. Layer 2 comprises 16 elements, of which 3 are occupied by cells, 11 by ECM elements, and 2 are free. The resulting cell density in layer 2 therefore amounts to $3/16 = 18.75\%$ and the ECM density to $11/16 = 68.75\%$. Middle and right: plot of a simulated configuration in the two-dimensional domain (middle) which is transformed into the one-dimensional cell and ECM densities (right) using the layer-method. The lattice coloring is the same as in Fig. 3.6.

4.2. Parameter Choice

To focus on the essentials for the sensitivity analysis, only the effects of three parameters from the CPM-FEM model are studied.

To make comparisons between the models possible, these parameters need to appear in each of the three models investigated – at least to the extent that it is clear that the parameter in question is involved in the same process in all models. Three parameters which fulfill this criterion and are suspected to play an important part in the cell invasion process are the initial ECM concentration, the ECM degradation rate, and the cancer cell proliferation probability.

In Crossley's model, two of these parameters are considered the most influential, namely the initial ECM concentration $\hat{m}_0$ and the ECM degeneration rate λ [54]. In her dimensionless model (3.3), they are the only parameters left. In the sensitivity analysis, her parameterized model (3.1)-(3.2) is examined because it also features the proliferation rate constant r.

The same holds true for Colson's model [52], and for the same reason her parameterized model (3.4) is used.

Proliferation probability as the third parameter is selected since it was not investigated closely by neither, Crossley nor Colson, even though the unimpeded proliferation of cancer cells and the subsequent growth of the tumor is what characterizes the illness and makes

it so deadly. In addition, the parameter is quite accessible in terms of interpretation and all three models feature it.

At this point it should be discussed whether the statement about Crossley's model being more complex than Colson's still holds true. The assumption of Crossley's superior complexity is based on the fact that her model incorporates volume-filling effects among the cells and the ECM, while Colson's model only considers cell-caused filling. When using the parameterized model formulations however, Colson's model features eight parameters, versus six in the case of Crossley. Hence, from now on the complexity levels of the PDE models are assumed to be more or less equivalent.

Continuing with the parameter choice in the CPM-FEM model, all three mentioned parameters occur more or less analogously in the CPM-FEM model. They were introduced with the model extensions, see sections 3.4.3 and 3.4.4. The effect of initial ECM concentration in the CPM-FEM model can be investigated easily since there is the parameter $\hat{m}_0$ regulating its concentration directly. The proliferation probability is just as simple to manipulate, as there is a parameter p_{p} defined for it. The ECM degradation rate is the only parameter without a direct counterpart in the extended CPM-FEM model. There are two possibilities of interpreting one of the CPM-FEM parameters as such a rate:

- the radius of the area of influence $\mathcal{A}(\tau(x))$: changing the size of $\mathcal{A}(\tau(x))$ has effects that range from restricting the cells' proteolytic power to a very narrow strip around an ECM element $\tau(x)$, to allowing cells from the other end of the domain to have an influence on its degradation probability.

- the choice of the weight decay parameter a: it has to be $a > 1$, since otherwise the cells farthest away from an ECM element $\tau(x)$ have a stronger influence on its degradation probability than the ones right next to it. If $a > 1$ is very close to 1, then all layers within $\mathcal{A}(\tau(x))$ have a similarly strong influence on the degradation of $\tau(x)$. If $a \gg 1$, then only the occupants of the elements closest to $\tau(x)$ make a difference.

The size of $\mathcal{A}(\tau(x))$ is chosen as the parameter corresponding to the degradation rate λ from Crossley and Colson because its impact on the ECM degradation is more direct and it is therefore easier to interpret.

Now that is has been determined which parameters will be varied for the sensitivity analysis, fixed values are selected for the rest of the parameters. The choices for the CPM-FEM model can be found in Tab. 4.1 and the ones for the PDE models in Tab. 4.2. A justification for these choices is provided in appendix A.6.

Ranges for the three parameters that will be subject to the sensitivity analysis need to be selected in a way that does not lead to numerical instabilities in the solutions of Crossley's

Parameter	Value	Unit	Reference(s)
domain size	60 by 60	elements	chosen
element size	2.5	μm	[50]
Poisson's ratio ν	0.3	-	[71]
Young's modulus E	10	kPa	[50, 71]
cell diameter d	20	μm	[50, 72]
initial number of cells	2	-	[52, 54]
cell elasticity λ	500	-	[50]
cell-to-cell attachment energy J_{CC}	2.5	-	[50]
cell-to-ECM attachment energy J_{CE}	1.25	-	[50]
intrinsic cell motility T	1	-	[50]
cell force constant μ	0.01	nN/μm	[50, 73]
reference stiffness E_0	10	kPa	[50, 71]
strain-stiffening speed ε_{st}	0.1	-	[50]
max. durotactic guidance α	10	-	[50]
durotaxis threshold stiffness E_{tr}	14.5	kPa	[30]
steepness of durotactic guidance β	0.5	-	[50]

Table 4.1.: Parameter choices of the fixed parameters in the CPM-FEM model.

Parameter	Value	Unit	Reference(s)
spatial domain length L	30	length units	chosen
cell diffusivity D	0.5	μm²/s	chosen
carrying capacity K	1	-	[52, 54]
initially invaded region α	5	length units	chosen
unconquerable ECM density m_{max}	1	-	[52, 54]
initial invasion σ	5	length units	[52]
initial invasion ω	1	length units	[52]

Table 4.2.: Parameter choices of the fixed parameters in Crossley's and Colson's models.

and Colson's models, neither should they provoke extensive runtimes in the CPM-FEM model or produce results where "nothing happens". Experiments with the models yield reasonable computational costs and relatively non-stiff behavior, yet interesting results, for the following parameter ranges.

The challenge in choosing an appropriate interval for the **initial ECM concentration** $\hat{m}_0$ consists in the tradeoff between exploring a wide range of different initial conditions and avoiding coming too close to extreme cases which limit the options for the other parameters. Small $\hat{m}_0$-values lead to more or less unrestricted expansion of the cells, which would require either a larger domain or small degradation and proliferation rates. Large $\hat{m}_0$-values, on the other hand, can drastically slow down the invasion process and

expose numerical instabilities in the experiments. A broad and yet safe choice is given by $\hat{m}_0 \in [0.2, 0.8]$.

The **area of influence** $\mathcal{A}(\tau(x))$'s radius has to be at least of size $0.25d$, where d is the cell diameter. Otherwise, by construction of the algorithm, no element is close enough to $\tau(x)$ to have an influence on it. As a maximum value for the radius, $2d$ is chosen, which corresponds to 15 elements to both sides of $\tau(x)$. For the PDEs' **degradation rate** λ, no such considerations are necessary and a wide range can be assigned to it: $\lambda \in [0.2, 100]$. Proliferation is one of the main factors dictating computational cost in the CPM-FEM model. The **proliferation probability** p_{p} is therefore chosen to be small, yet large enough to cover the whole domain with cells if at the maximum of its interval: $p_{\mathrm{p}} \in [0.0001, 0.01]$. For Crossley and Colson, two different maximum values are chosen due to the different invasion speeds depending on the r-dependent reaction terms, whereby Colson's and Crossley's r values are set to $r_{\mathrm{Colson}} \in [0.00005, 0.1]$ and $r_{\mathrm{Crossley}} \in [0.00005, 0.4]$, respectively.

A summary of the parameter ranges can be found in Tab. 4.3.

Parameter	Interval	Unit	Reference(s)
initial ECM concentration $\hat{m}_0$	$[0.2, 0.8]$	-	chosen
radius of $\mathcal{A}(\tau(x))$	$[0.25d, 2d]$	µm	chosen
or degradation rate λ	$[0.2, 100]$	-	chosen
proliferation probability p_{p}	$[0.0001, 0.01]$	-	chosen
or r_{Colson}	$[0.00005, 0.1]$	-	chosen
or r_{Crossley}	$[0.00005, 0.4]$	-	chosen

Table 4.3.: Parameter ranges of the parameters for which the model sensitivity is investigated.

4.3. Variance-Based Method: Computation of Sensitivity Indices

Now that a one-dimensional measure for the CPM-FEM model has been defined, the parameters subject to the sensitivity analysis and their ranges have been identified and values were assigned to the remaining, fixed parameters, the analysis can start. To choose an appropriate method, the frame conditions and the goals of the analysis need to be considered. The input spaces that were defined above are quite large to capture the models' behaviors under extensive variation of the selected parameters. The PDE models are both nonlinear, and while at least semi-explicit solutions can be found for them analytically [52, 54], this is not possible for the CPM-FEM model. Also, the output of the

CPM-FEM model is non-deterministic, requiring a method that considers randomness to some extent, but is also capable of handling deterministic models. While the primary objective of the sensitivity analysis here is to investigate which of the three parameters is the most influential on each model outcome under consideration of the temporal and spatial development, gathering insights into the interplay between the parameters is interesting as well: if the same pairs of parameters interact strongly for all three models, it can be concluded that the relationships between the parameters are similar throughout the models. If not, there must be some differences inherent to the models or the choices of the fixed parameters.

In conclusion, the sensitivity analysis method of choice has to include the whole input space, handle nonlinearity and randomness, work without exact solutions or derivatives, and it should be able to capture interaction effects between the parameters. A method which provides all of these features is Sobol's variance-based sensitivity analysis.[1] Variance-based sensitivity analysis tries to answer the question "Which percentage of the variance in the model output is caused by the variance of each input parameter?" It can be answered by decomposing the output variance into the variances of the conditional expectations of the output under variation of a single parameter value or a parameter combination.

In detail, Andrea Saltelli describes variance-based sensitivity analysis in his book on global sensitivity analysis [74] as follows: the model is regarded as a function $f(X) = Y$, $X = \{X_1, ..., X_k\}$ being the uncorrelated input parameters and Y the one-dimensional output. The domain of f is given by the k-dimensional unit hypercube, i.e. each X_i fulfills $0 \leq X_i \leq 1$, $i = 1, ..., k$ – this does not restrict the parameter space, as all parameter ranges can be transformed to this interval. Saltelli describes how his colleague Ilya M. Sobol' decomposes the model function f as follows [75]:

$$f(X) = f_0 + \sum_{i=1}^{k} f_i(X_i) + \sum_{i=1,j>i}^{k} f_{ij}(X_i, X_j) + \cdots + f_{1,...,k}(X_1, ..., X_k), \qquad (4.1)$$

where f_0 is a constant and each of the other $f_{\mathcal{I}}$ is a function of the input parameters whose indices appear in $\mathcal{I}$, $\mathcal{I} \in \mathcal{S} := \{S : S$ tuple with ascending elements composed of $\{1, ..., k\}\}$, e.g. $f_{2\,4\,5}$ depends on X_2, X_4 and X_5. The number of terms per sum in (4.1) therefore increases and decreases again as there is one f_0-term, k f_i-terms, $\binom{k}{2}$ f_{ij}-terms, $\binom{k}{3}$ f_{ijl}-terms, and so on until there is one $f_{1,...,k}$-term left.

In the following, let $X_{\mathcal{I}} := \{X_{i_1}, ..., X_{i_j}\}$ and $\mathrm{d}X_{\mathcal{I}} := \mathrm{d}X_{i_1}, ..., \mathrm{d}X_{i_j}$, $i_1, ..., i_j \in \mathcal{I}$.

---

[1]Using a variance-based method and the corresponding Python-package `SALib` was personally recommended by Dr. Pirmin Schlicke, TUM.

Sobol' requires all terms of the form $f_{\mathcal{I}}$ to have zero mean in the sense that $\int_0^1 f_{\mathcal{I}}(X_{\mathcal{I}}) \mathrm{d}X_{\mathcal{I}} = 0$. In this case, it holds $\int_0^1 f_{\mathcal{I}}(X_{\mathcal{I}}) f_{\mathcal{J}}(X_{\mathcal{J}}) \mathrm{d}X_{\mathcal{I}} \mathrm{d}X_{\mathcal{J}} = 0$, $\mathcal{I} \neq \mathcal{J}$, meaning that the terms are pairwise orthogonal. With these prerequisites, one can determine the terms of the decomposition given the expected (conditional) model output Y:

$$f_0 = \mathbb{E}(Y),$$
$$f_i(X_i) = \mathbb{E}(Y|X_i) - f_0,$$
$$f_{ij}(X_i, X_j) = \mathbb{E}(Y|X_i, X_j) - f_0 - f_i - f_j,$$

and so on. f_i hence measures the effect of varying X_i, whereas f_{ij} measures the joint effect of X_i and X_j which arises additionally to the effects of varying X_i or X_j alone. The same holds true for terms with three or more variables: f_{ijl} measures the joint effects of X_i, X_j and X_l. To answer the initial question about the output variance, one last step is performed to arrive at a measure of sensitivity. For this, Sobol' assumes the model function f as well as the $f_{\mathcal{I}}$ to be square-integrable, then (4.1) can be squared and integrated to

$$\int_0^1 (f(X))^2 \,\mathrm{d}X - f_0^2 = \sum_{\mathcal{I} \in \mathcal{S}} \int_0^1 (f_{\mathcal{I}}(X_{\mathcal{I}}))^2 \,\mathrm{d}X_{\mathcal{I}}, \tag{4.2}$$

where the left hand side equals the variance of Y and the right hand side is a sum of variances. While the former comes directly from the definition of variance as $\mathrm{Var}(X) = \int x^2 f(x) \mathrm{d}x - \mathbb{E}(X)^2$, the latter might best be explained by exemplarily inserting the definition of one of the f_i and rewriting the integral by using the definitions of expected value and variance as well as the law of total expectation:

$$\int_0^1 (f_i(X_i))^2 \mathrm{d}X_i = \int_0^1 (\mathbb{E}(Y|X_i) - \mathbb{E}(Y))^2 \mathrm{d}X_i = \mathbb{E}\left((\mathbb{E}(Y|X_i) - \mathbb{E}(Y))^2\right) = \mathrm{Var}(\mathbb{E}(Y|X_i)).$$

Now, $V_i := \mathrm{Var}(f_i(X_i))$ is introduced and it should be noted that $\mathbb{E}(Y)$ can always be dropped as it is a constant:

$$V_i = \mathrm{Var}(f_i(X_i)) = \mathrm{Var}(\mathbb{E}(Y|X_i) - \mathbb{E}(Y)) = \mathrm{Var}(\mathbb{E}(Y|X_i)),$$
$$V_{ij} = \mathrm{Var}(f_{ij}(X_i, X_j)) = \mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j)).$$

With the new notation (4.2) can be rewritten as the so-called ANOVA-HDMR (analysis of variance – high-dimensional model representation):

$$\mathrm{Var}(Y) = \sum_{i=1}^k V_i + \sum_{i=1, j>i}^k V_{ij} + \cdots + V_{1,\ldots,k}.$$

The decomposed variance terms on the right hand side can be divided by the total output variance $\text{Var}(Y)$ to obtain a relative measure of the input variables' effects, leading to the first-order sensitivity indices (also known as first-order Sobol' indices) which capture the sole contributions of the variation of X_i:

$$S_i = \frac{V_i}{\text{Var}(Y)},$$

the second-order sensitivity indices which measure the contributions to the total variance by the interplay of X_i and X_j:

$$S_{ij} = \frac{V_{ij}}{\text{Var}(Y)},$$

and so on. One can also summarize the total contribution of the parameter X_i by defining the total-order index $S_{\text{T}i}$, which is the sum of all indices in which X_i appears. The total-order index captures the first-order effect of X_i plus all higher-order effects in which X_i is involved. For a model with three input parameters X_1, X_2, X_3 – as it is the case here – the total-order index of the first parameter is given by

$$S_{\text{T}1} = S_1 + S_{12} + S_{13} + S_{123}.$$

However, it is not necessary to consider all of the indices to capture the leading variance contributions. In most cases, the first- and second-order indices "have significant contributions to the overall variance of f" [76]. To be certain that one does not miss important effects, one can additionally compute the total-order indices and compare them with the ones of first and second order.

To compute these indices, data needs to be generated to be able to measure variance and means. The values for the three investigated parameters $\hat{m}_0$, $\mathcal{A}(\tau(x))$ (or λ), and p_{p} (or r) have to be chosen carefully such that they cover their ranges evenly (uniform distributions are assumed for all parameters on their respective intervals) and still allow for robust estimations of the indices above.

Let k be the number of parameters to be investigated (here, $k = 3$ for each model). For the parameters to be selected, the k-dimensional space is transformed from their original domains (here, the intervals in Tab. 4.3) to the k-dimensional unit cube $[0, 1]^k$. The goal is to select a set of points in this cube which leave no "holes" between them. True randomness – or pseudo randomness, as in scientific computing true randomness is hard to achieve – cannot guarantee the even covering of the space. Deterministic, low-discrepancy quasi-random sequences such as Sobol' sequences (named after the same mathematician who found the ANOVA-HDMR), are better at fulfilling this requirement. Without going further into detail, Sobol' sequences generate sets of cardinality 2^m, $m \in \mathbb{N}$ which "fill

the holes" that were left by the 2^{m-1} elements of the previous set in the sequence. As $m \to \infty$, the points in each set become more and more equidistant, which is the definition of low-discrepancy [77, 78].

It has to be noted that the points of this sequence are quasi-Monte Carlo points, meaning that they are not chosen independently from each other (new points are selected such that they are more or less equidistant to the points that already have been set). However, sample independence is required to be able to use Monte Carlo estimators for expected value or variance. These estimators can be used provided the sample is modified according to randomized quasi–Monte Carlo: with random transformations, the points are "jiggled" to loose their deterministic nature, but they still lie within their pre-defined ranges and preserve their even spread among the parameter space to some extend [78]. Saltelli does this by applying a random digital shift and a random linear scrambling to the points from the Sobol' sequence, which can be imagined like a combination of shift and permutation, where both operations include sampling from a uniform distribution over $[0, 1)^k$ [78, 79]. This method is referred to as "Saltelli scheme".

After the sampling and randomization is complete, the parameter samples are re-transformed into their original domains.

Of course the quality of the estimates for the indices does not only depend on the distribution of the parameter samples, but also on the number of model evaluations. The necessary number of evaluations depends itself on the number of parameters in the model. Let $n = 2^m$, $m \in \mathbb{N}$, be a number roughly determining the magnitude of model evaluations which is used for the cardinality of the Sobol' sequence. Usually, more than n model evaluations are necessary, so to obtain more individual parameter samples, one does not pick the Sobol' sequence from the k-dimensional unit hypercube, but from the $2k$-dimensional one. Then, these samples are put in an $n \times 2k$-matrix which itself is split into two $n \times k$-matrices A and B. k new $n \times k$-matrices $C_1, ..., C_k$ are built by setting $C_i = A$ and replacing the i-th column of C_i with the i-th column of B, $i = 1, ..., k$. Additional k $n \times k$-matrices $D_1, ..., D_k$ can be obtained by repeating the procedure vice versa, i.e. set $D_i = B$ and replace the i-th column of D_i with the i-th column of A. This way, one obtains $2k + 2$ matrices (all C_i and D_i, as well as A and B), each containing n parameter combinations, totalling to $n(2k+2)$ different parameter combinations. [74, 80]

Fig. 4.2 shows how n samples generated by a Sobol' sequence and their transformation into $n(2k + 2)$ samples using the Saltelli scheme including the above-described extension could look like for $k = 2$.

Saltelli found that with his scheme, the Monte Carlo estimates for the first- and total-order indices can be efficiently and robustly computed using the results of $n(2k+2)$ model
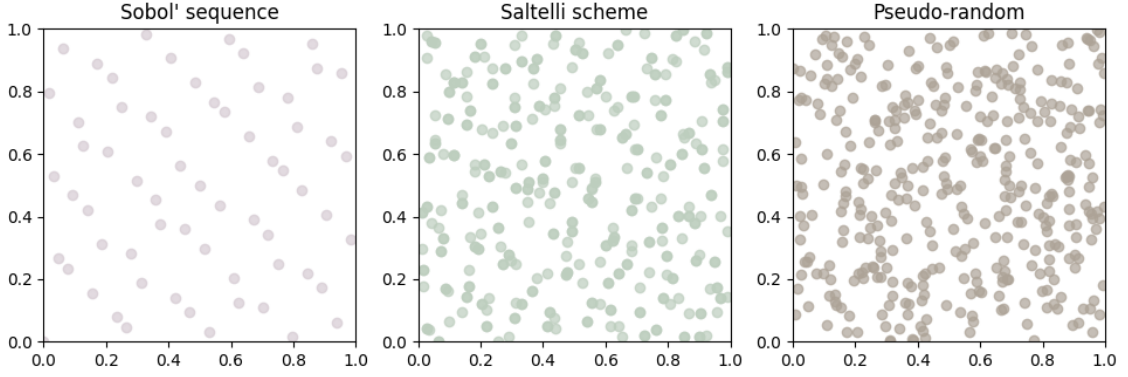
Figure 4.2.: Left: Sobol' sequence with cardinality $n = 64$. Middle: Sobol' sequence with cardinality $n = 64$ transformed into $n(2k + 2) = 384$ points using the Saltelli scheme. Right: for comparison, 384 points computed by a pseudo-random number generator.

evaluations [79]. The output variance $\text{Var}(Y)$ is estimated using $2n$ evaluations:

$$\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \approx \frac{1}{n} \sum_{j=1}^{n} y_A^j y_B^j - \left(\frac{1}{n} \sum_{j=1}^{n} y_A^j \right) \left(\frac{1}{n} \sum_{j=1}^{n} y_B^j \right) =: \hat{V}(Y),$$

where $y_A = f(A)$ is the length-n vector of model outputs resulting from the parameter sample matrix A, and y_A^j is the j-th entry of y_A. The notation works analogously for the other parameter matrices $B, C_1, ... C_k, D_1, ..., D_k$. Another nk evaluations are used for the first-order terms, with $i = 1, ..., k$ [80]:

$$S_i = \frac{\text{Var}(\mathbb{E}(Y|X_i))}{\text{Var}(Y)} = \frac{\mathbb{E}\left((\mathbb{E}(Y|X_i) - \mathbb{E}(Y))^2 \right)}{\text{Var}(Y)}$$

$$\approx \frac{1}{\hat{V}(Y)} \left(\frac{1}{n} \sum_{j=1}^{n} y_A^j y_{C_i}^j - \frac{1}{n^2} \sum_{j=1}^{n} y_A^j \left(\sum_{l=1}^{n} y_B^l \right) \right) =: \hat{S}_i.$$

Using the estimate for the variance of the conditional expectation by Toshimitsu Homma and Saltelli [81], the remaining nk evaluations are used for the total-order terms [80]:

$$S_{Ti} = 1 - \frac{\text{Var}(\mathbb{E}(Y|X_1, ..., X_{i-1}, X_{i+1}, ..., X_k))}{\text{Var}(Y)}$$

$$\approx 1 - \frac{1}{\hat{V}(Y)} \left(\frac{1}{n} \sum_{j=1}^{n} y_A^j y_{D_i}^j - \left(\frac{1}{n} \sum_{j=1}^{n} y_A^j \right) \left(\frac{1}{n} \sum_{j=1}^{n} y_B^j \right) \right) =: \hat{S}_{Ti}.$$

The same evaluations also allow to estimate the second-order indices [79]:

$$S_{ij} = \frac{\text{Var}(\mathbb{E}(Y|X_i, X_j))}{\text{Var}(Y)} - S_i - S_j$$

$$\approx \frac{1}{\hat{V}(Y)} \left(\frac{1}{n} \sum_{r=1}^{n} y_{D_i}^r y_{C_j}^r - \frac{1}{n^2} \sum_{r=1}^{n} y_A^r \left(\sum_{l=1}^{n} y_B^l \right) \right) - \hat{S}_i - \hat{S}_j =: \hat{S}_{ij}.$$

As mentioned earlier, the number of model evaluations determines the quality of these estimations. Two factors were respected when this number was chosen: the computational expense regarding the CPM-FEM model should be manageable, and if possible, all models should be evaluated the same number of times to make the estimations most comparable. Due to the limiting first factor, the number of model evaluations need to be relatively low, leading to a Sobol' sequence of length $n = 1024$, i.e. 8092 parameter combinations. Concerning the practicalities of the computations, the sensitivity analysis is conducted using the Python package `SALib` [82]. The parameter combinations are sampled with `SALib.sample.sobol`, and the indices are computed with `SALib.analyze.sobol`, both using the methods described above.

4.4. Results

While it turns out that the index estimates for 8092 model evaluations appear quite reliable in the case of the CPM-FEM model, they are at least slightly flawed for Colson's model, and unusable for Crossley's. For the latter, the estimates of the first-order indices are in some instances larger than 1, which is unacceptable. Therefore, $n = 2048$ is chosen for the PDE models, while $n = 1024$ is maintained for the other two models.

One flaw in the calculation of the indices for the two PDE models could not be fixed by elevating the number of samples: when investigating the parameters' effects on the ECM density, a few values in the first third of the domain could not be calculated which is visible in the plots as missing data. The reason for this could be that the output variance $\text{Var}(Y)$ at these points in the domain is equal to zero. Due to the initial invasion, i.e. the maximum cell density from $x = 0$ to $x = 4$ or 5 – depending on the model –, every model evaluation starts with a cell density of 1 and an ECM density of 0 in this interval. The sensitivity indices of the PDE models in this interval can thus be thought of as zero.

Fig. 4.3 and 4.4 show the first-order sensitivity indices for all three models and their evolution over the domain at three selected time steps, namely $t = 10$, 20 and 30. The influence of the parameters is evaluated separately for the progression of the cells and the degradation of the ECM.
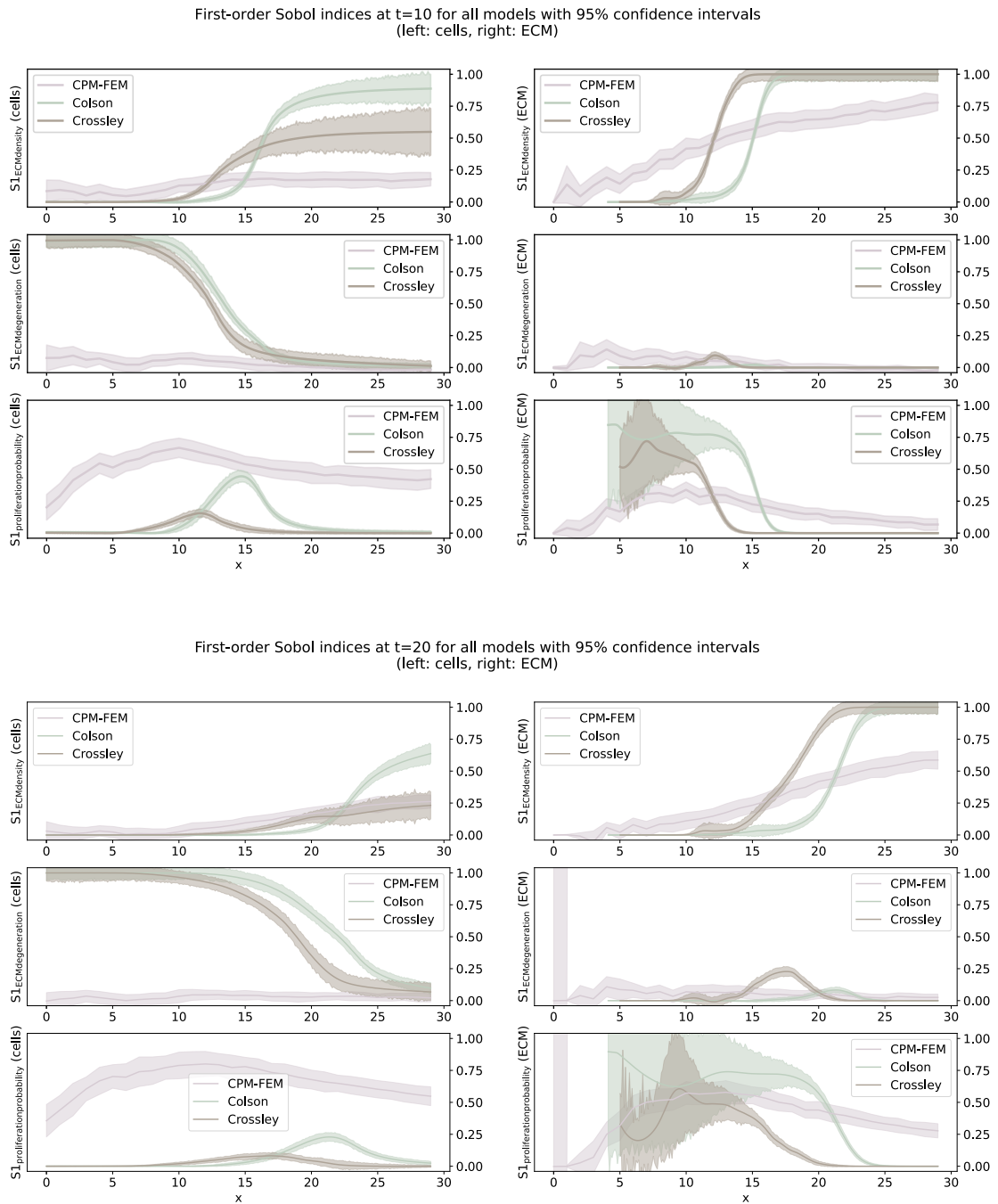
47

Figure 4.3.: Estimations of the first-order sensitivity indices of all models at $t = 10$ and 20 over the spatial domain, with a 95% confidence interval. See Fig. 4.4 for more details.
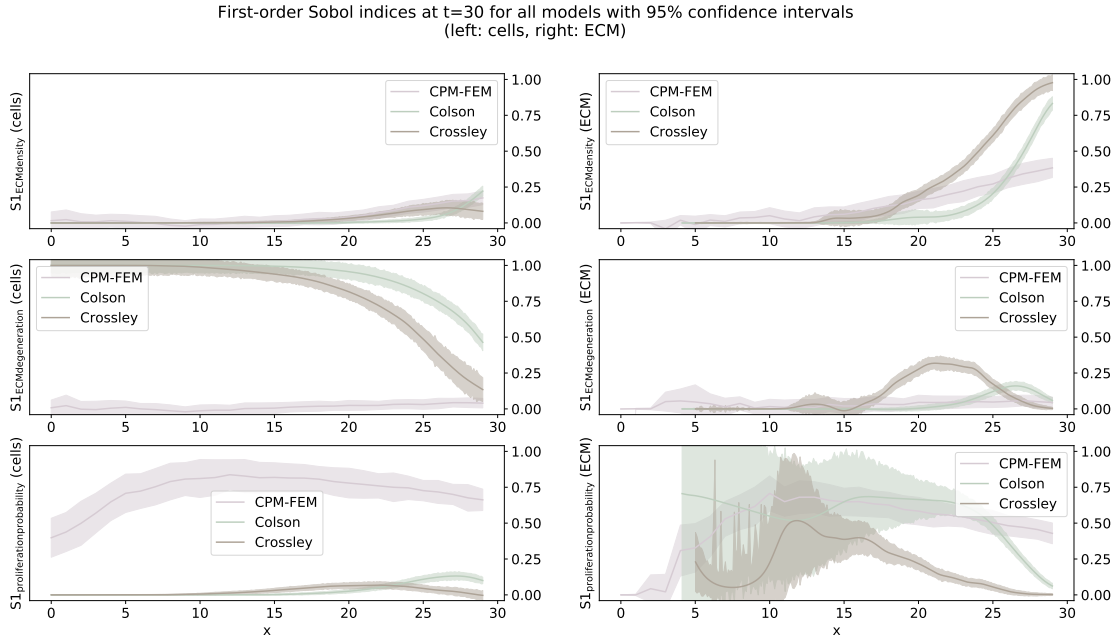
Figure 4.4.: First-order sensitivity indices at $t = 30$. On the left, the effects of the parameter variance on the cell density output variance are shown, on the right it is depicted for the ECM density. The top row shows the influence of the initial ECM density $\hat{m}_0$, the middle row shows the ECM degradation rate λ (or the radius of $\mathcal{A}(\tau(x))$ in case of the CPM-FEM model), and the bottom row shows the indices of the proliferation probability p_{p}, r_{Colson} or r_{Crossley}, depending on the model.

It attracts attention how the indices of the two PDE models have a relatively similar course, with a few exceptions, for instance does the proliferation probability r_{Colson} exhibit a more pronounced influence on the cells' progression than r_{Crossley}, especially in the early time steps. Vice versa, Crossley's degeneration rate λ has a larger effect on the ECM density than Colson's, which grows over time. Furthermore, one can assume the course of the solution curves of the models by observing the first-order index of the ECM degradation rate λ in case of the cell evolution and the index of the initial ECM density $\hat{m}_0$ in case of the ECM evolution. The former looks like the invading cell front, while the latter reminds of the dissolving ECM barrier – a picture similar to the travelling wave solutions in Fig. 3.2. The similarity goes as far as even the faster wave speed of Colson's model is implied by the first-order sensitivity indices, but this of course is just the result of a fortuitous choice in parameter ranges. Based on the observed progressive cell fronts, one can make a forecast: for $t \to \infty$, the ECM degradation will become the only influential parameter in both PDE models. It is also interesting how the proliferation

probability's indices have their peak more or less close to the point of steepest decrease of the ECM degeneration's indices (in the case of the cells), and similarly how the ECM degeneration's peak is located close to the point of steepest increase of the initial ECM density's indices (in the case of the ECM). For the cells, this is especially well visible for smaller values of t, while for the ECM, this becomes more apparent as t grows, and it is more precise for Colson than for Crossley. This observation indicates (self-evident) interaction effects on the cell's progression between proliferation probability and ECM degeneration, and on the ECM's retraction between ECM degeneration and initial ECM density. Another candidate for second-order effects is the combination of initial ECM density and proliferation probability in case of the cell evolution. Here, the peak of the proliferation probability's indices is close to the x-coordinate where the initial ECM density index curve rises. There is a similar effect observable in the case of the ECM evolution, where the decline in the proliferation probability's influence coincides with the increasing initial ECM density. These interactions are investigated further once the second-order indices are evaluated.

The indices of the CPM-FEM model do not exhibit as much variation as the ones of the PDE models, in most cases they appear quite steady and it is not possible to make a direct comparison to the model trajectories – except for perhaps the course of the initial ECM density's index, which gets shifted to the right domain boundary as the ECM gets more and more degraded. Concerning the cells, the effect of the proliferation probability is the most prominent and it grows more influential over time. The same holds true for this parameter regarding the ECM evolution, with the difference that here, also the before-mentioned initial ECM density has a strong influence especially on the second half of the domain in the early time steps.

In the next step, the interaction effects are investigated. The estimates for the total-order indices (see appendix A.7) are substantially larger than the first-order indices of initial ECM density $\hat{m}_0$ and radius of $\mathcal{A}(\tau(x))$ in the case of the CPM-FEM model, at least in the first half of the domain. For the PDE models, the differences between first- and total-order indices are not standing out, but as explained above, pairwise interactions are still expected.

Fig. 4.5 and 4.6 show the estimated second-order indices for all models. The presumptions on the interaction effects between the PDE's parameters are more or less confirmed by the plots: ECM degeneration and initial ECM density exhibit large interaction effects on the ECM evolution, and proliferation and initial ECM density have the most conspicuous second-order effect regarding the cells. Further, proliferation probability and ECM degeneration have at least some joint effect on the cells, but it is comparably low, just as the interaction effect between proliferation and initial ECM density on the ECM.
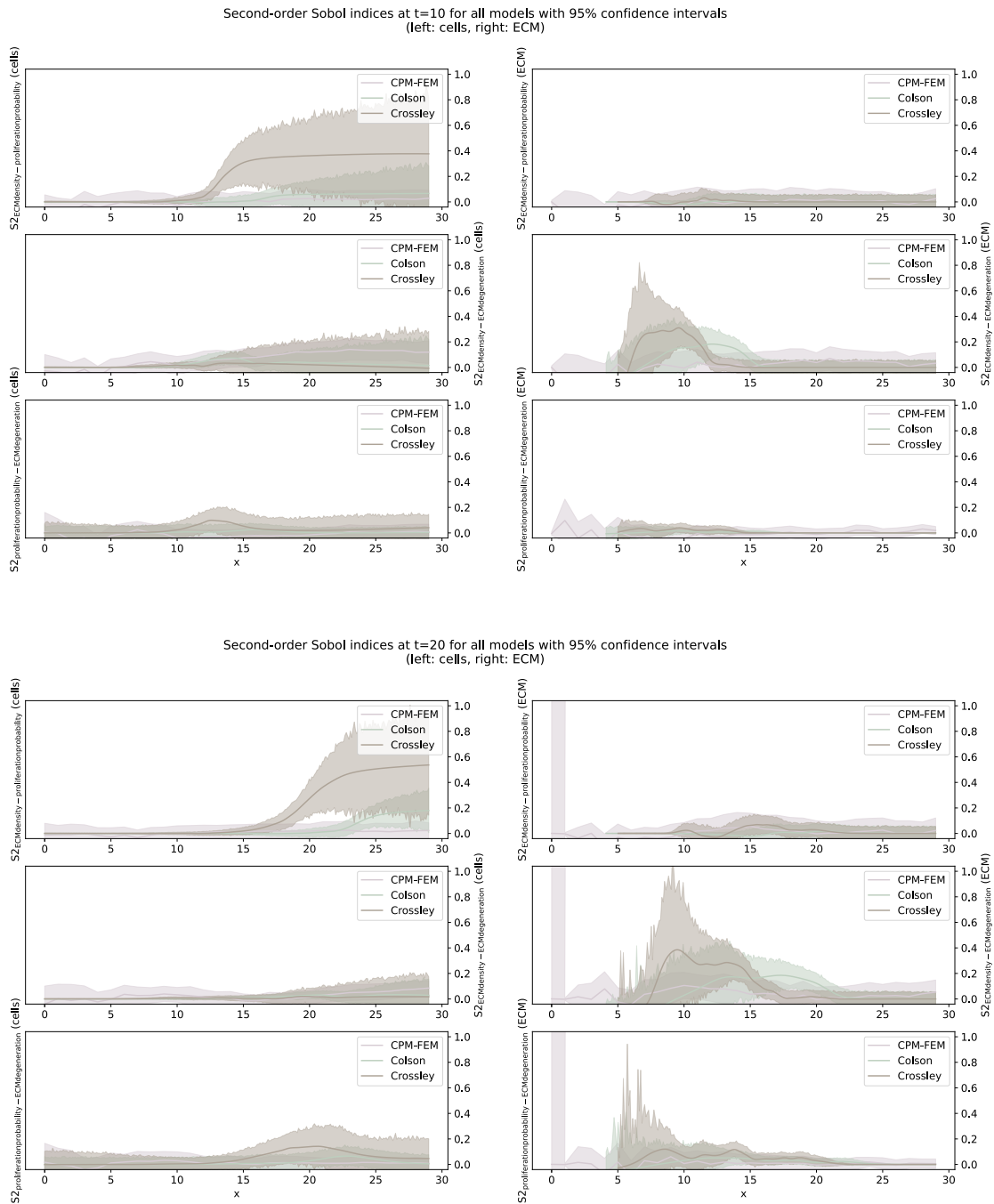
Figure 4.5.: Second-order sensitivity indices at $t = 10$ and 20.

The second order indices show an unexpected rise in interaction effect between proliferation probability and ECM degeneration for the ECM at $t = 30$. But the results need to be
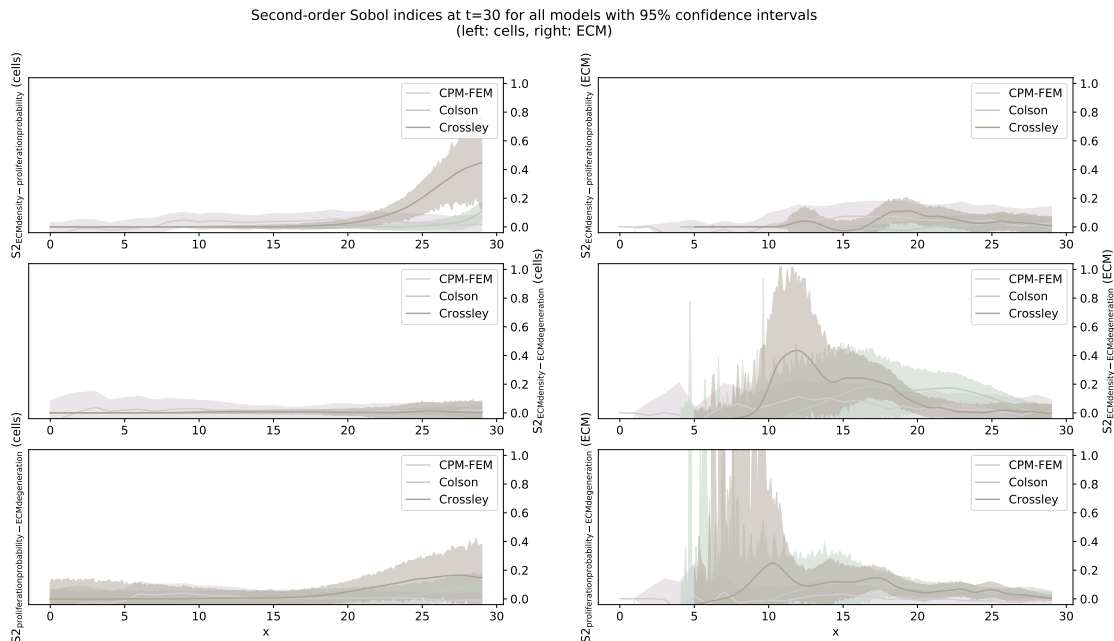
Figure 4.6.: Second-order sensitivity indices at $t = 30$.

put into perspective. In most cases, the total-order indices do not deviate substantially from the first-order indices in case of the PDE models, hinting at relatively low second-order effects. So it is surprising to see such strong interactions in many of the mentioned cases. A possible explanation is the accuracy of the estimated second-order indices, which often is quite low as it can be seen in the wide confidence intervals. Hence there seem to be interactions between the mentioned parameters, but they are not substantial – except for perhaps the combined effect of initial ECM density and proliferation probability in the case of Crossley's cell evolution, since here, even with the large confidence region the difference between first- and total order indices is visible.

Another surprise is the absence of second-order effects in case of the CPM-FEM model. An explanation could be third-order interactions, i.e. the interaction of all three parameters, which can be computed from the estimates of the first-, second- and total-order indices (using a simplified notation with model parameters X_1, X_2, X_3):

$$\hat{S}_{123} = \hat{S}_{T1} - \hat{S}_1 - \hat{S}_{12} - \hat{S}_{13}.$$

Of course, one can also use the respective lower and total-order indices of the other parameters, in theory the resulting $\hat{S}_{123}$ should always be the same. To compensate for inaccuracies, $\hat{S}_{123}$ is calculated for all three combinations and their mean and standard

deviation are used, which should be a sufficient error estimate since the purpose of these computations is only to test the assumption that the CPM-FEM model exhibits strong third-order effects. The results of this estimation are visualized in Fig. 4.7. While there is one value above 1 at $t = 10$ concerning the influence on the ECM density which most likely is the result of accuracy issues, the rest of the estimates lies within the valid range and it is clearly visible that indeed there are strong effects coming from the interaction of all three parameters in the first half of the domain.
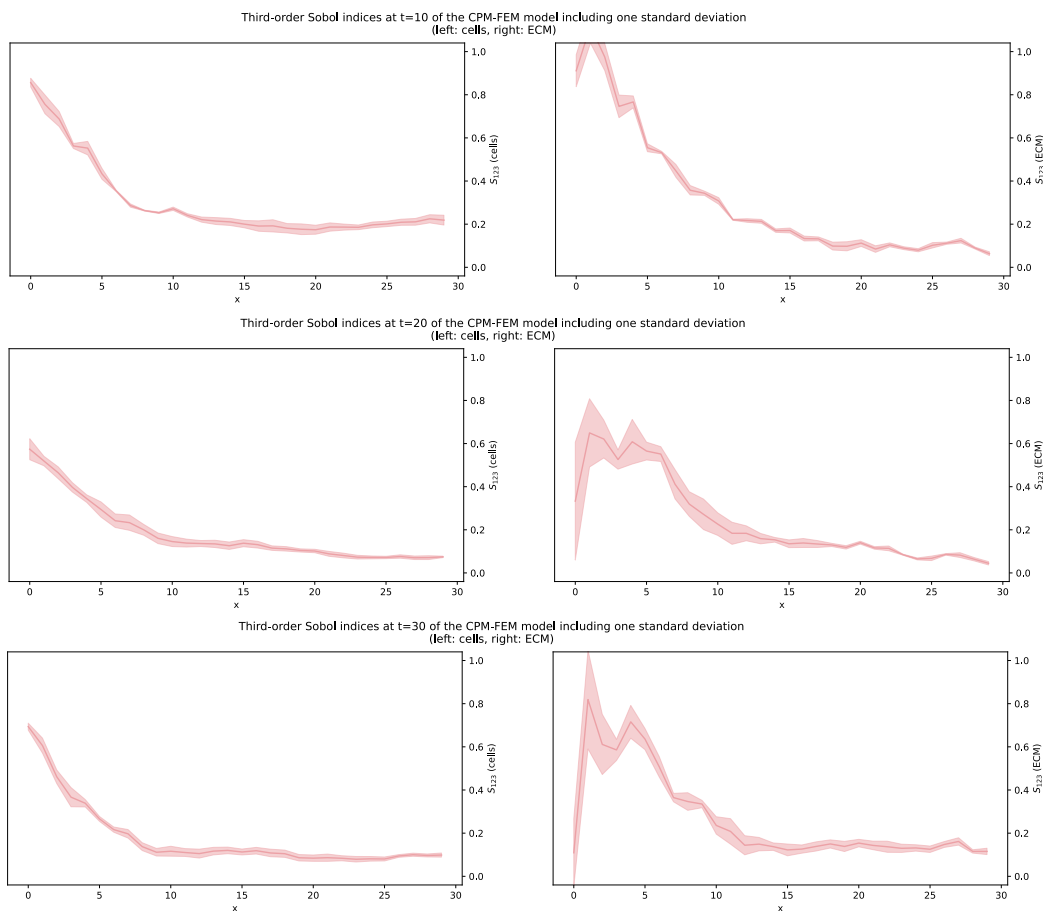
Figure 4.7.: Estimates of the third-order sensitivity indices of the CPM-FEM model at three time steps $t = 10, 20, 30$.

In summary, the sensitivity analysis revealed some similarities, but also major differences in the parameter interplay of the models. The PDE models – which behave very similar in this analysis except for the cases discussed above – exhibit mainly first-order effects, where in the beginning the leading parameter is the initial ECM density $\hat{m}_0$, and as t grows, $\hat{m}_0$ gets replaced by the ECM degradation rate λ. In the CPM-FEM model,

relevant first-order effects are found only for the proliferation probability p_p, and – if t is small – for the initial ECM density $\hat{m}_0$. A large partition of the model output variance in the CPM-FEM model is caused not by a single parameter, but by the interplay of all three parameters $\hat{m}_0$, $\mathcal{A}(\tau(x))$ and p_p.

5. Qualitative Comparisons

In the previous chapter, major differences between the PDE models and the CPM-FEM model regarding parameter sensitivity have been detected. One question that arises from these findings is how the outputs of these models differ in a qualitative manner, given their unequal inherent dynamics. Can parameters be found that lead to similar results for all three models?

In this chapter, the trajectories of the parameterized PDE models are compared to the CPM-FEM model output. Via parameter fits and wave speed comparisons it is studied how adequately the models approximate each other – first using only the parameters from the sensitivity analysis for the fit, then including all PDE model parameters to obtain the whole picture. In this context, it is also investigated whether phenomena such as durotaxis have an influence on model fit adequacy and whether they affect invasion speed in a manner that can be detected in the one-dimensional domain.

5.1. PDE Parameter Fit: Three Parameters

To see how well the PDE models are able to capture the characteristics of the CPM-FEM model output, a parameter fit is performed. For this fit, the same three parameters as in the sensitivity analysis are used (initial ECM concentration, ECM degradation rate and proliferation probability). The fitting procedure can be reduced to a two-parameter-fit, because the initial ECM concentration is a parameter that works like a boundary condition for both models. If fixed to the same value in all models, it is a helpful information for the data fitting function and improves the estimations for the remaining two parameters λ and r (dropping the subscripts for a moment).

The parameter fitting works as follows (see for instance [83]): the CPM-FEM model incorporates randomness, hence it delivers the (perturbed) observations y for the fit. Then, the solutions of the PDE models are used as fitting functions f by minimizing the summed squares of the residuals between the observations and the fitting function. Say the domain of the CPM-FEM has n layers, meaning that there are n points $x_1, ..., x_n$ in the one-dimensional domain for which data is available, then the goal is to find parameters λ, r such that

55

$$\text{SSR} := \sum_{i=1}^{n}(f(x_i, \lambda, r) - y_i)^2 = \min. \tag{5.1}$$

The CPM-FEM model is evaluated 100 times after $t = 60$ iterations to smoothen the observations. $t = 60$ is also the time step that is used for the fitting. For all three models, the same fixed parameter values are used as in the sensitivity analysis.

The initial ECM density is set to $\hat{m}_0 = 0.8$, the radius of the area of influence is chosen as $0.3d$ and the proliferation probability is $p_{\text{p}} = 0.005$.

As initial guesses for the PDE parameters λ and r, for both models $\lambda = 0.2$ and $r = 0.005$ are chosen, without any bounds to find an unrestricted optimum. The fitting is performed using the Python function `scipy.optimize.curve_fit` [84], which solves (5.1) using a Levenberg-Marquardt algorithm.

This procedure is performed for three different experimental settings of the CPM-FEM model: once with all effects, i.e. the full system energy Hamiltonian, once without the durotaxis-factor in the Hamiltonian, and once without attachment cost.

5.1.1. Parameter Fitting Including All Effects

The results of the parameter estimation with the full CPM-FEM model can be found in Tab. 5.1.

Parameter	Initial guess	Found optimum	SSR
Crossley's degradation rate λ	0.2	3.7989	0.8295
Crossley's proliferation probability r_{Crossley}	0.005	0.1561	
Colson's degradation rate λ	0.2	0.2268	0.1988
Colson's proliferation probability r_{Colson}	0.005	0.1110	

Table 5.1.: Optimal parameter values and sum of squared residuals for the PDE models.

Fig. 5.1 shows the PDE solutions that result from the optimal parameter choice together with the data from the CPM-FEM model, not only at $t = 60$ (which is the time point that was used for the fitting), but also for $t = 0, 20$ and 40.

Even though the shapes of the fitted curves do not represent the CPM-FEM results very well, they are able to reproduce the wave speed of the travelling waves that are exhibited by the CPM-FEM model. This is visible in the intersection point of each model's cell and ECM density curve: even though only $t = 60$ was used for the parameter fit, the x-values of these intersection points are relatively close to each other at every depicted time step. In the field of PDE modeling of invasion processes in nature, wave speed is usually defined as the velocity of the propagating wave front [52, 54, 85, 86]. It is difficult to define a front
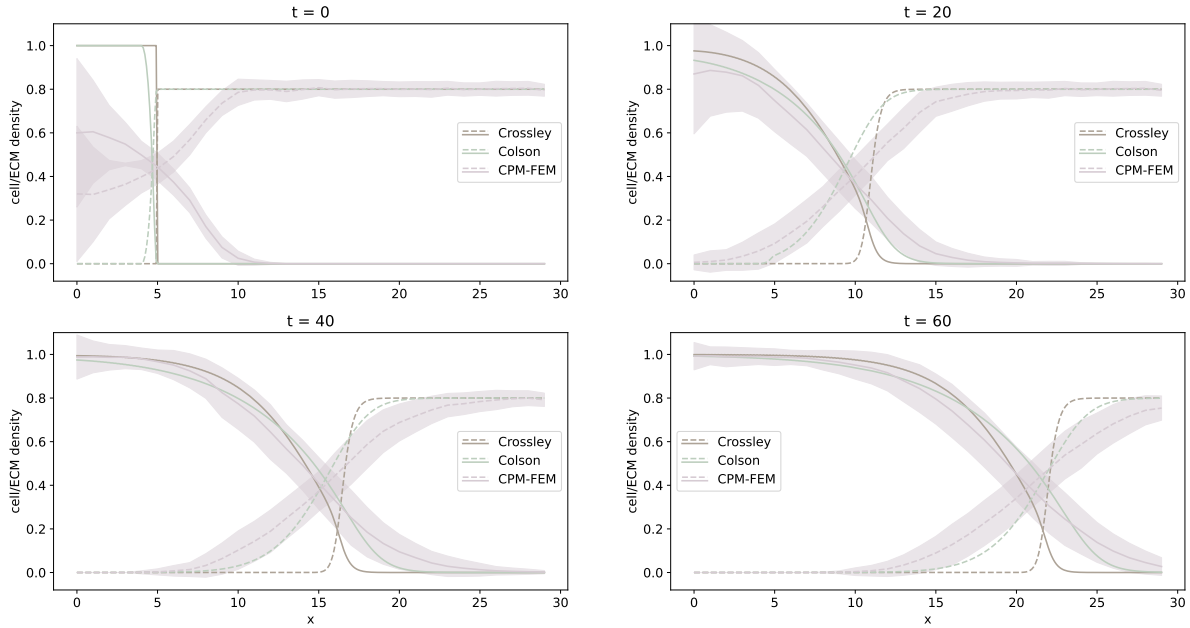
Figure 5.1.: Fitted PDE models and the data generated by the average over 100 CPM-FEM model evaluations, depicted for the time steps $t = 0, 20, 40, 60$. The CPM-FEM model average is depicted with a confidence region of one standard deviation. The solid lines are the cell densities, the dashed lines are the respective ECM densities.

in the case of the FEM-CPM results because at the early time steps, the cell concentration at $x = 0$ is less than 1 due to the small number of initial cells – they are placed in the middle of the grid, i.e. close to $x = 0$, but this placement still incorporates randomness, which on average leads to the low initial concentration that can be seen in Fig. 5.1, $t = 0$, and even $t = 20$. To still be able to perform wave speed measurements, in this work, wave speed is defined as horizontal shift velocity of the intersection point of the cell density curve with a horizontal line at $y = 0.5$. Fig. 5.2 shows the wave speeds for the three models that result from this definition. The velocities are estimated using the intersection point shift between $t = 0$ and $t = 5$, $t = 5$ and $t = 10$, and so on. Compared to the PDE models which both approximate asymptotes quickly, the CPM-FEM results keep oscillating for the entire observed timeframe. It is the only model that incorporates randomness, so this is not surprising. Ignoring the oscillations, its qualitative behavior is comparable to the ones of the PDEs, while quantitatively it appears to be slightly better represented by Crossley's model. The wave speed plot also depicts how Colson's waves

are the fastest, as it can also be seen in Fig. 5.1 how over time, Colson's waves pass Crossley's and the CPM-FEM's.

It has to be noted that the uncertainty for the CPM-FEM wave speed is very high. The plot in Fig. 5.2 intentionally does not show its standard deviation as it is so large that it would require a vertical axis of the four-fold length.
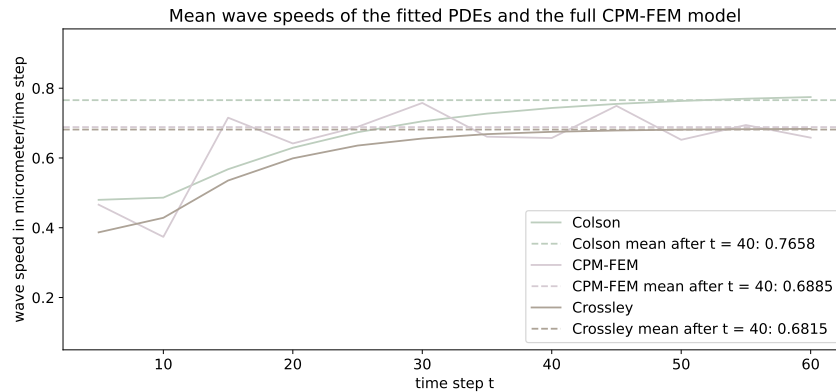
Figure 5.2.: Wave speed estimates of the fitted PDE models and the average of 100 CPM-FEM model evaluations. The dashed lines are the respective mean values of these wave speeds, where the mean is calculated for the time steps after $t = 40$ to approximate the PDE asymptotes. Note that for all wave speed plots (i.e. also Fig. 5.4 and 5.7), the same vertical axis limits are used to make visual comparisons possible.

Returning to the discussion of the parameter fit quality, the SSR values in Tab. 5.1 give an orientation of the residual magnitude. While they indicate that the fit for Colson's model was more successful, it is clear from the graphic that shape-wise, both fits are not perfectly suited to describe the course of the CPM-FEM model evaluations.

Both, the cell- and the ECM densities generated by the CPM-FEM model proceed in a much more gradual manner than the solutions of the PDE models – an observation that is consistent with the findings on the first-order sensitivity indices, where the PDE models exhibit more accentuated effects. The perhaps most striking difference is the steepness of the ECM front that is about to be degraded. Even though the area of influence was chosen quite small to provoke a concise interface between cell front and ECM, for the CPM-FEM model the ECM front is a gentle ramp, while for both PDE models it has more similarities with a scarp. There are two possible reasons for both, the steepness of the PDE curves, and the comparably shallow CPM-FEM simulations.

Regarding Crossley's model, which is the PDE model with the steepest ramps, it is noticeable how there is almost no spatial overlap of cells and ECM. In the agent-based

model formulation it is stated that cells can degrade ECM elements only if the ECM element is located in the same lattice site. Hence a critical density of cells needs to reach the ECM-"wall" before it can effectively degrade it. Independently of the shape of the invasive front, this is a remarkable limitation to the model, as it shows that if there is a point in the domain where the ECM density is 1, this point imposes an unbreachable wall to cell invasion [54] – a phenomenon that is not observed in experiments, where malignant cells degrade the ECM using diffusing proteases. Colson's model does not feature volume-filling by the ECM, which allows a larger cell-ECM-overlap and hence more gentle slopes.

Regarding the FEM-CPM model, the cause of the shallow curves and the large area of overlap has to be the frayedness of the simulated cell front, as it can be seen in the examples of section 3.4.5. But is the rough spheroid boundary solely a result of the randomized choice of proliferating cells, or could durotaxis provoke a less even spread of the cells? How do the adhesive forces between cells and ECM influence this pattern? Does the consideration of these phenomena influence travelling wave speed? These are the questions that are investigated in the next sections.

5.1.2. Parameter Fitting without Durotaxis

Maintaining all parameter values from the previous section, the CPM-FEM model is now evaluated while ignoring the durotaxis-component of the system energy Hamiltonian. It is not too far-fetched to assume that durotaxis leads to a more uneven cell spheroid boundary, given the appearance of the exemplary plots in section 3.4.5 with the "durotactic fingers". Of course it is disputable whether such effects are perceivable in the relatively small spatial and temporal domains that are considered here.

The same method for data fitting is used as before. Its results can be found in Tab. 5.2. Interestingly, both parameters λ and r in both models have been estimated lower than in the case with durotaxis, hinting at a likely lower invasion speed. Also, the SSR values are slightly lower but the improvement is marginal.

Parameter	Initial guess	Found optimum	SSR
Crossley's degradation rate λ	0.2	3.2647	0.8019
Crossley's proliferation probability r_{Crossley}	0.005	0.1346	
Colson's degradation rate λ	0.2	0.1880	0.1838
Colson's proliferation probability r_{Colson}	0.005	0.0980	

Table 5.2.: Optimal parameter values and sum of squared residuals for the PDE models, without durotaxis.

Fig. 5.3 and 5.4 show the fitted models and their averaged wave speeds. Qualitatively, the plots reveal the same situation as in the case with durotaxis, and especially the frayedness of the cell-ECM-interface does not seem to have changed in a remarkable way. It might have been reduced slightly, explaining the lower SSE values, but as mentioned before, this reduction is not substantial.
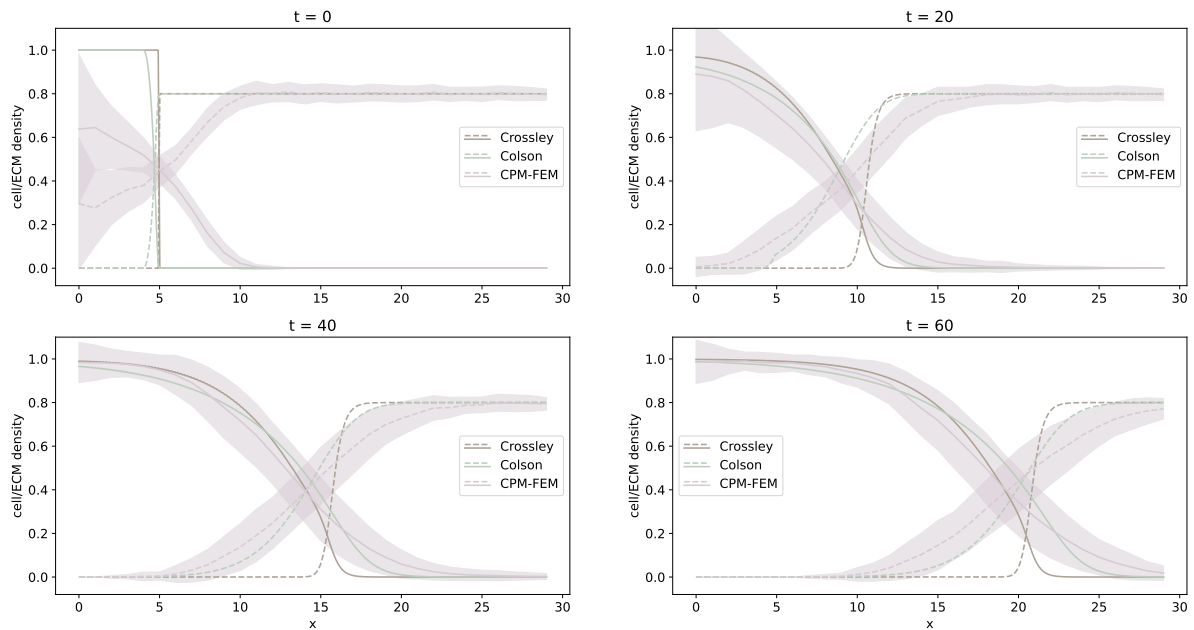
Figure 5.3.: Fitted PDE models and the data generated the durotaxis-free CPM-FEM model. For more details see Fig. 5.1.

However, both the PDE parameter estimates and the mean wave speed indicate that the CPM-FEM wave speed might have decreased in the scenario without durotaxis. This seems like a reasonable hypothesis given the durotactic mechanism in which the substrate strains caused by one cell leaving the spheroid motivates fellow cells to follow it, as it is visible in Fig. 3.10. The mechanism might be strong enough to influence the invasion speed in the one-dimensional representation. It should be emphasized that "durotaxis increases wave speed" is not the same as "durotaxis increases migratory speed", as this would contradict experimental findings [30]. The invading cell front is hypothesized to be faster in the case of durotaxis because the cells move away from the spheroid in a more directed manner, not because their migration speed increases.

Mean wave speeds of the fitted PDEs and the CPM-FEM model without durotaxis

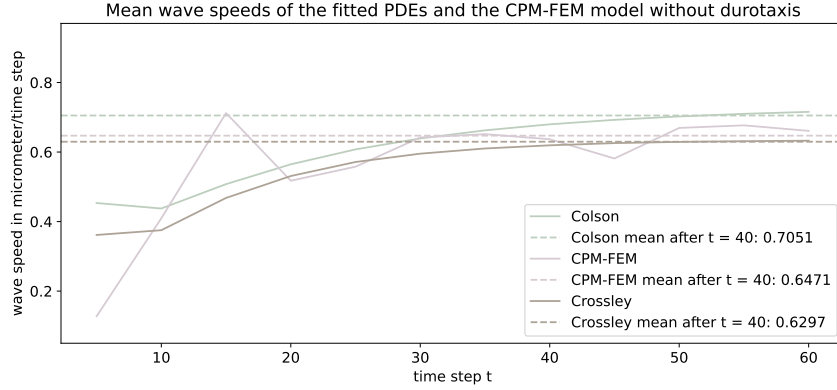Figure 5.4.: Wave speed estimates of the fitted PDE models and the durotaxis-free CPM-FEM model. For more details see Fig. 5.2.

The assumption of different wave speeds can be investigated more thoroughly using a statistical test. Let the first sample $X := \{X_1, ..., X_{100}\}$ be constituted by the 100 model evaluations with durotaxis, the second sample $Y := \{Y_1, ..., Y_{100}\}$ consists of observations from the evaluations without durotaxis. The respective observations are then given by the x-coordinates of the intersection points of the cell density curves with $y = 0.5$ for all CPM-FEM model evaluations, tested at various time steps from $t = 0$ to 60 (from now on, the notations X_t and Y_t for X and Y at t is used). Normal distribution is tested for X and Y at each t using the D'Agostino-Pearson Omnibus test with a significance level of $\alpha = 0.05$ which evaluates the symmetry and variance of the samples [87]. If according to this test both samples can be assumed to be normally distributed, it should be safe to use the two-sample independent t-test. If at least one of the samples is not normally distributed (i.e. the p-value of the test is below α), the Mann-Whitney-U-test is used at the respective time point, which can be seen as a two-sample t-test for arbitrarily distributed samples [88]. In most cases however, the samples are found to be normally distributed at the given significance level. The sample sizes are not always equal as in some instances there is no intersection point (especially at t close to 0 or 60) – but this is not an issue for this kind of test, as long as the samples contain at least 30 observations [89]. At each tested time point t, the tested hypotheses are

$$
\begin{aligned}
\mathcal{H}_0 : \ & \mathbb{E}(X_t) = \mathbb{E}(Y_t), \\
\mathcal{H}_1 : \ & \mathbb{E}(X_t) > \mathbb{E}(Y_t),
\end{aligned}
\tag{5.2}
$$

representing the speculation that the intersection is shifted faster in the case of durotaxis. The p-values of the tests over time are visible in Fig. 5.5. After only 25 time steps,

the p-values are consistently below 0.001, strongly supporting a decision in favor of the alternative hypothesis $\mathcal{H}_1$. The suspicion that durotaxis increases wave speed – according to the definition of wave speed that was introduced here – is confirmed.
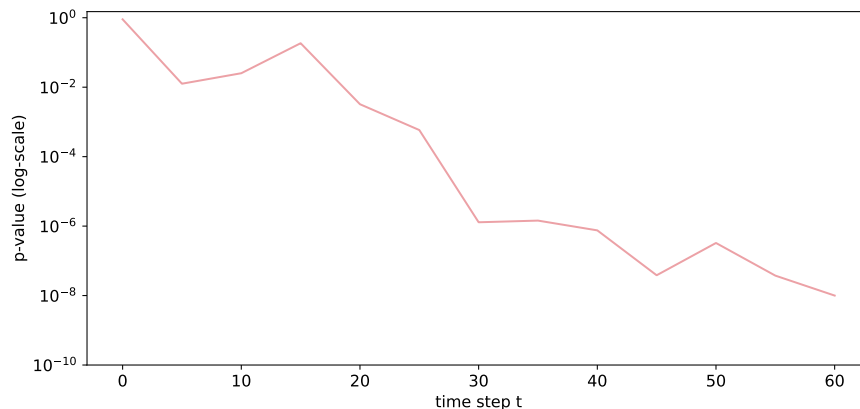
Figure 5.5.: p-values over time for the hypotheses (5.2). Note that the vertical axis is log-scaled.

5.1.3. Parameter Fitting without Attachment Cost

In this last experiment, durotaxis is again included in the CPM-FEM model, but the contact energy contribution to the Hamiltonian is ignored. This means that now it does not matter for the system energy if the cells are closely surrounded by lots of other cells or spread on the grid with some space between them. The data fitting results are summarized in Tab. 5.3.

Parameter	Initial guess	Found optimum	SSR
Crossley's degradation rate λ	0.2	5.6166	0.7882
Crossley's proliferation probability r_{Crossley}	0.005	0.1637	
Colson's degradation rate λ	0.2	0.2781	0.2037
Colson's proliferation probability r_{Colson}	0.005	0.1219	

Table 5.3.: Optimal parameter values and sum of squared residuals for the PDE models, without attachment cost.

The values of the fitted parameters are the largest obtained in all three experiments, suggesting that the invasion speed could be faster than in the cases of the full Hamiltonian and the one without durotaxis, and the SSR values – which are the lowest of all three scenarios for Crossley, but the largest for Colson – provide mixed signals regarding the

frayedness of the cell spheroid boundary. Looking at the comparative plots in Fig. 5.6 and 5.7, the known pattern arises of Colson's faster invasive front passing Crossley's after enough time steps. For the first time, the CPM-FEM wave speed does not appear to approximate an asymptote, it rises quickly and then gradually drops. More samples and a larger time frame would be necessary to see whether this is a product of randomness.
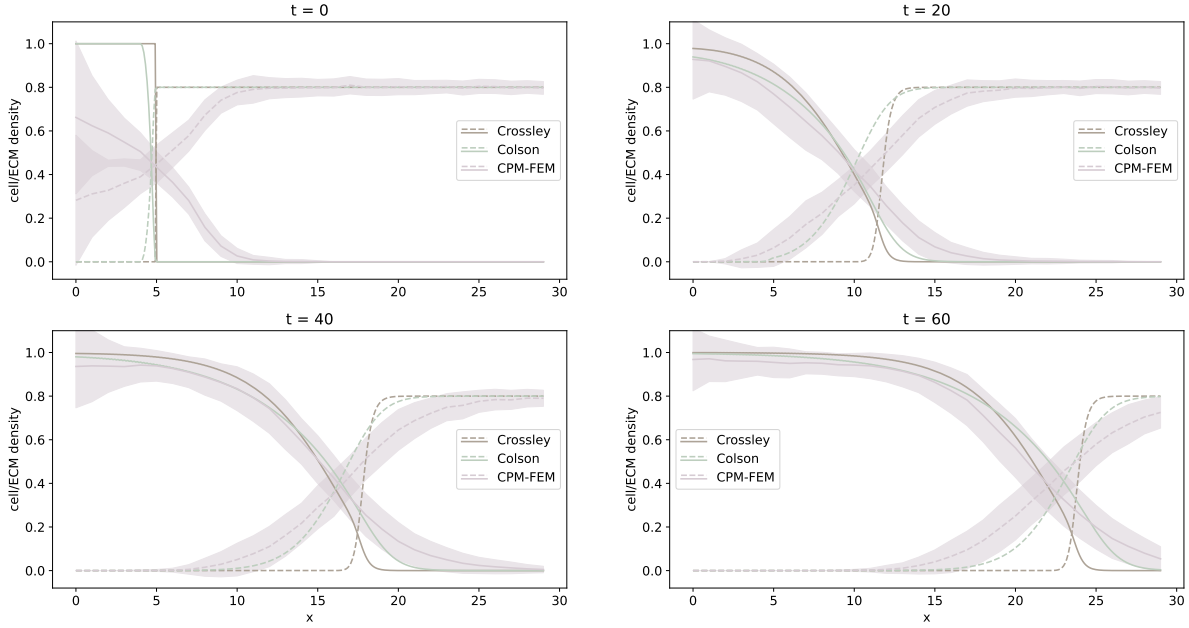
Figure 5.6.: Fitted PDE models and the data generated the contact cost-free CPM-FEM model. For more details see Fig. 5.1.

To test the assumption that the invasive cell front indeed moves faster than in the full CPM-FEM model, another statistical test is conducted analogously to the one in the previous section. The hypotheses – for X_t being the intersection x-coordinates of the CPM-FEM model without contact cost at a specific t, Y_t the intersections of the model with the full Hamiltonian – are given by

$$
\begin{aligned}
\mathcal{H}_0 &: \ \mathbb{E}(X_t) = \mathbb{E}(Y_t), \\
\mathcal{H}_1 &: \ \mathbb{E}(X_t) > \mathbb{E}(Y_t).
\end{aligned}
\tag{5.3}
$$

Fig. 5.8 depicts the p-values of this test and again it takes about 25 time steps to arrive at a significance level below 0.001. However, the p-values do not keep decreasing as it was the case in Fig. 5.5 of the previous section. Their rise after 40 to 50 time steps is
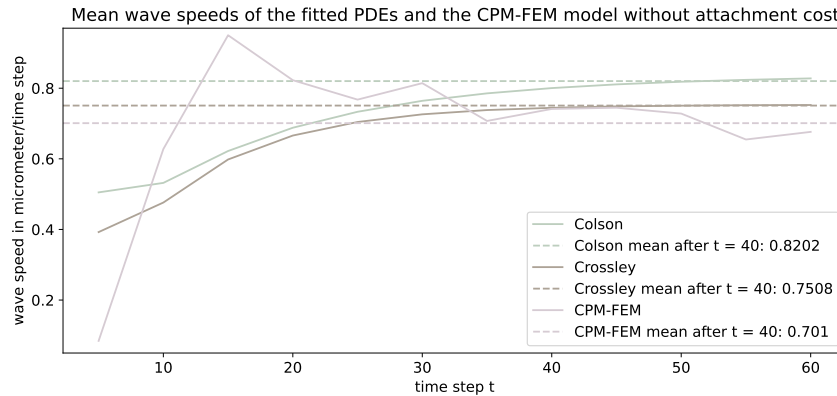
Figure 5.7.: Wave speed estimates of the fitted PDE models and the contact cost-free CPM-FEM model. For more details see Fig. 5.2.

most likely a result of the drop in wave speed that was discussed earlier. The purpose of hypothesis testing is to make sample-independent statements, hence at least for the given time frame up to $t = 60$, $\mathcal{H}_1$ is accepted, although the decision could be different for a larger temporal domain. The present decision however leads to the conclusion that different costs of the attachment between two cells or between cells and ECM elements (here, the adhesion of cell to cell is set to be more energetically expensive than cell to ECM) result in slower invasion speeds.
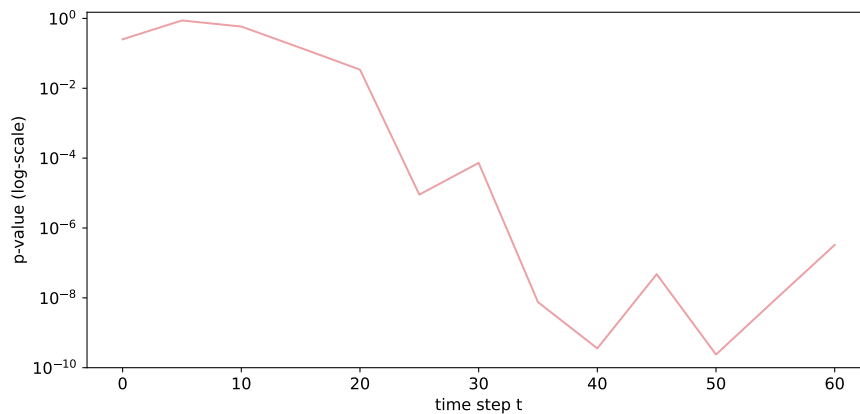
Figure 5.8.: p-values over time for the hypotheses (5.3). Note that the vertical axis is log-scaled.

This result is surprising, given that the cell-to-cell attachment energy J_{CC} is set twice as large as the cell-to-ECM attachment energy J_{CE}. Such a choice of parameters suggests that the cells are moving away more quickly from the initial spheroid, leading to a faster

invasion than in the case $J_{CC} = J_{CE}$. Studying the two-dimensional model configuration might deliver an explanation. Fig. 5.9 depicts some exemplary configurations from the model evaluations with the full Hamiltonian (top row, subfigures a) and c)) and without attachment costs (bottom row, subfigures b) and d)).

a)

c)

b)

d)
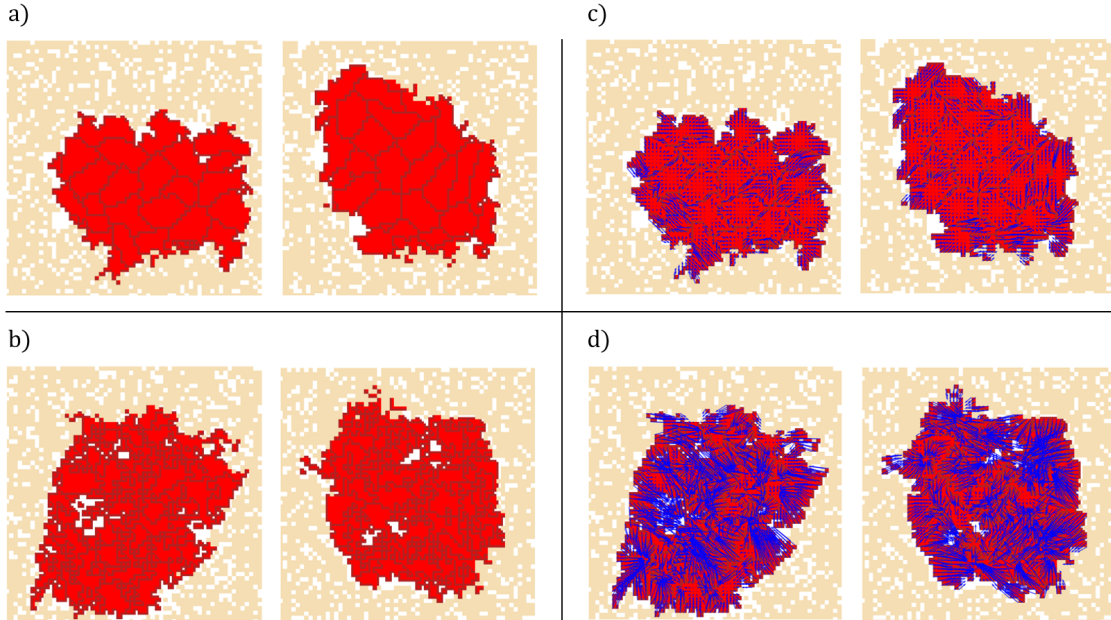
Figure 5.9.: Shape and traction force differences of the full CPM-FEM model (top row) and the contact cost-free model (bottom row) at $t = 60$. a) and b): exemplary configurations of two evaluations of the full model a) and the contact cost-free one b). c) and d): the same configurations as in a) and b), respectively, including traction forces as blue arrows.

In Fig. 5.9a) and b), it is visible how random the shapes of the cells in the scenario without attachment cost b) are, even after 60 time steps, compared to the cells from the full CPM-FEM model a). Note that in a) there are some convoluted, strangely shaped cells at the rim of the spheroid as well, but this is because these cells are daughter cells that were placed in the grid recently. As the surrounding ECM is degraded within the next time steps, there will soon be enough space for them to assume more rounded shapes.

Given the persisting randomness of cell shapes, it is not astonishing that the forces within the cells are much larger for the attachment-cost-less model: the traction forces of the discussed configurations are displayed in Fig. 5.9c) and d). Why do the cells not give in to these forces and assume more balanced shapes?

One possible reason for this could be that if the cells do not have any incentive to reduce their contact points with neighboring cells (by decreasing the length of their outline), they

more or less remain in the random shapes assigned to them at birth, regardless of the traction forces. While the contact energy is a factor that is included in the system energy Hamiltonian H, the traction forces only have an indirect influence on H: the forces lead to substrate deformation, and the substrate deformation influences the durotaxis-term in H. This influence appears to be too small to have an effect, at least in the relatively small time scale of 60 steps.

But how is the cell shape linked to the invasion speed? In Fig. 5.9a) and b), it looks like the rounded cells lead to a more "compact" cell spheroid, with a relatively even rim and almost no holes, whereas the randomly shaped cells produce a more perforated spheroid with a very frayed rim. Both of these observations – more holes and a more frayed spheroid boundary – could lead to increased space consumption, which then makes the wave speed appear faster.

Concluding the section on the three-parameter-fit, while the trajectories of the PDE models do not capture the exact shapes of the invasive cell front or the degraded ECM, it is possible to find parameters which describe the invasion speed relatively well. The overall quality of the PDE fit is only marginally impacted by the choice of considered phenomena in the CPM-FEM model. In the observed time frame, the invasion in the CPM-FEM model is the fastest in the case where attachment cost is ignored, followed by the full model, and the slowest invasion is observed in the case without durotaxis.

5.2. PDE Parameter Fit: All Parameters

The same procedure as in the previous section is now repeated fitting all PDE model parameters. In the case of Crossley, there are six parameters to fit, and for Colson there are eight. The initial ECM density is again given as an *a priori* information, leading to five and seven degrees of freedom left, respectively.

Since the number of parameters now differs between the models, it could be useful to consider another metric for the fit quality next to the SSR. The Aikaike Information Criterion (AIC) is such a metric which was specifically developed for the task of deciding which model of various models provides the best approximation. In the case of using least squares as a data fitting method – as it is done here –, the AIC is given as

$$\text{AIC} = n \ln \left(\frac{\text{SSR}}{n} \right) + 2(k+1),$$

where n is the number of data points from the CPM-FEM model, and k is the number of parameters of the respective PDE model [90]. The lower the AIC, the better suited is the model to describe the given data set.

The algorithm used to minimize the SSR in the case of Colson's model is now changed to a Trust Region Reflective algorithm, since bounds for the parameters need to be provided to maintain the relationship $0 < \omega < \sigma$ of the initial conditions. This is done by introducing an auxiliary parameter $c \in (0,1)$, allowing to compute ω as $\omega = c\,\sigma$. The bounds for Colson's parameter are given as follows: $\lambda \in [0, 100]$, $r \in [0, 1]$, $D \in [0, 5]$, $m_{\max} \in [0, 5]$, $K \in [0, 5]$, $\sigma \in [0, 30]$, and $c \in [0, 1]$.

Fitting all parameters to the data generated by 100 evaluations of the full CPM-FEM model at $t = 60$ leads to the SSR and AIC values given in Tab. 5.4. The found optimal parameter values are listed in appendix A.8, and the resulting trajectories together with the CPM-FEM model output are visualized in Fig. 5.10.

Model	SSR	AIC
Crossley	0.7819	-248
Colson	0.0203	-463

Table 5.4.: SSR and AIC values of the fitted PDE models.
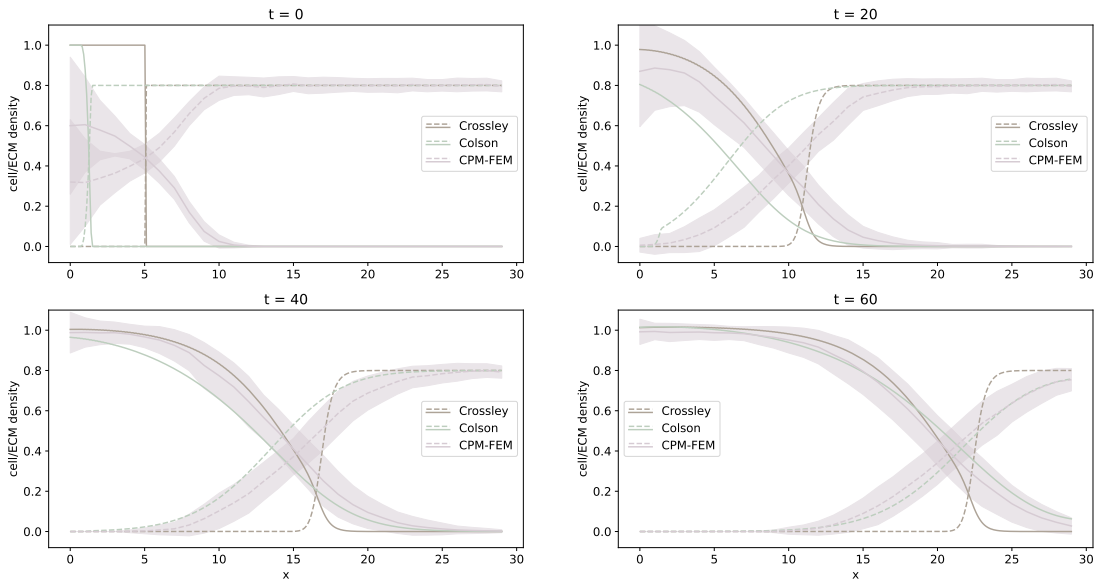
Figure 5.10.: All-parameter fitted PDE models and the data generated by the average over 100 CPM-FEM model evaluations, depicted for the time steps $t = 0$, 20, 40, 60. The CPM-FEM model average is depicted with a confidence region of one standard deviation. The solid lines are the cell densities, the dashed lines are the respective ECM densities.

Compared to the SSR values of the previous fit to the full model in section 5.1.1, where only the degradation rate and the proliferation probability were estimated, the improvement is only minor in the case of Crossley's model, but substantial for Colson's. Unsurprisingly, the AIC scores clearly favor Colson's model.

Two interesting differences to the results of section 5.1 regarding Colson's trajectories are visible in the plots of Fig. 5.10. First, the gentle slopes of both quantities, the invading cell front and the retracting ECM, are mimicked better than before, explaining the low SSR value at $t = 60$. Secondly, the wave speed is not captured well: at the time point $t = 60$ used for the fit, the CPM-FEM wave and Colson's front coincide almost perfectly, but going back in time, their intersections with the imaginary line at $y = 0.5$ disagrees more and more. Crossley however appears to find a compromise between a mediocre shape fit and a relatively good invasion speed fit. To confirm the observations on wave speed, consider Fig. 5.11. It depicts the model-wise averaged wave speeds, calculated as described in section 5.1.1.
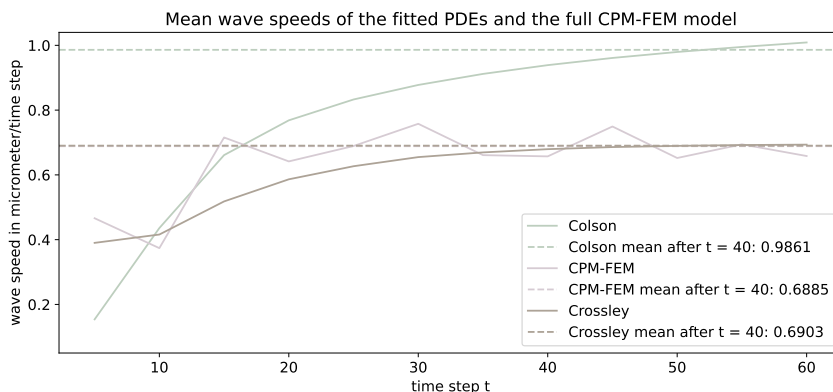
Figure 5.11.: Wave speed estimates of the all-parameter fitted PDE models and the average of 100 CPM-FEM model evaluations. The dashed lines are the respective mean values of these wave speeds, where the mean is calculated for the time steps after $t = 40$.

As expected, Crossley's model fits the invasion velocity of the CPM-FEM model much more accurately than Colson's, to the point where after $t = 40$, their means almost coincide.

The same situation is observed for the cases without durotaxis and without attachment cost. The respective SSR and AIC values, as well as the fitting plots and the wave speed comparisons can be found in appendix A.9.

Concluding the chapter on the qualitative model comparisons, it has been found that both PDE models have distinctive qualities: if the degradation rate and the prolifera-

tion probability are the only degrees of freedom, then both models make similarly good approximations of wave speed and shape. If all parameters are to be estimated, then Colson's model convinces regarding shape fit, whereas Crossley's strength is the robust wave speed approximation.

6. Lessons Learned

It is time to evaluate the methods and results presented in this work. Have answers been found to all questions? What remains to be investigated?

6.1. Limitations

This work studied the behavior of three cell invasion models of different complexity: two PDE models which are on a relatively similar level of complexity, and one hybrid model which by far outshines the PDE models regarding the number of considered parameters and processes. However, even in the development of the CPM-FEM model, it is obvious that many simplifying assumptions have been made. Chapter 2 presented a choice of properties and mechanisms that are involved in cell invasion, and comparing the CPM-FEM model with this selection alone reveals the extent to which this model deviates from reality. To name just a few of the simplifications, the finite element formulation is based on a linear elastic surface without dynamic deformation, even though the ECM is a non-linear elastic material and in nature, deformations of course outlast one Monte-Carlo step. Consequently, phenomena such as microbuckling – which can be considered as a part of dynamic deformation – are ignored, even though they might play an important role for durotaxis in the early stages of tumorigenesis. The CPM part of the model lacks obvious details of the invasive process as well: nutrient distribution is ignored completely, cell death is not considered, and cells do not grow but are born at their target size. The Hamiltonian which links the CPM with the FEM part could also be made more realistic, for instance by allowing cell traction force to have a more direct influence on the system energy.

To see which of these details are worth the additional computation expense, the model should be validated using experimental data from experiments with real cancer cells. The same holds true for the PDE models: only the comparison with real data can make a robust statement about model adequacy.

A limitation that most likely complicated the comparability of the PDE models with the CPM-FEM model is the size of the simulated tumors. While neither Crossley nor Colson make assumptions on the size of the tumors from which their modeled invasive cell front

origins, it can be assumed that they are all either large avascular or vascularized tumors, i.e. tumors of around two millimeters in diameter or larger [91]. Alexander P. Browning and colleagues, who calibrated a PDE model similar to the ones studied here using images from *in vitro* experiments, used cell spheroids of six millimeters initial diameter [51]. Six millimiters – or two, even – is one whole magnitude larger than the simulation results. The CPM-FEM modelled tumors that were investigated in the sensitivity analysis and the qualitative comparisons were "grown" on a 60 by 60 element domain, hence even if they use the space to full capacity, they cannot exceed diameters of 150 µm, i.e. 0.15 mm. But not only the size of the spatial domain is relatively small, also the experimental time frames are quite short. All investigations in this work were performed with either 30 or 60 time steps, even though it has been shown in section 3.4.5 that hundreds of time steps are necessary to obtain a clearer impression of the effects in the CPM-FEM model. In the mentioned calibration by Browning et al., the cells are observed over a time span of four days. "Time steps" is a unitless quantity, hence it is not straightforward to find the equivalent duration, but once again, comparisons with data from experiments with real cells could help with this.

The hypothesis that the combination of a larger tumor spheroid size and more time steps impacts model comparability is illustrated by the following experiment: the CPM-FEM model is evaluated a single time using the same parameters as in the proliferation-example from section 3.4.5, except for the spatial domain size which is augmented to 200 by 200 elements, and the number of time steps, which is stopped at 600. A fit of the PDE models' degradation rates λ and proliferation probabilities r leads to the fit shown in Fig. 6.1.
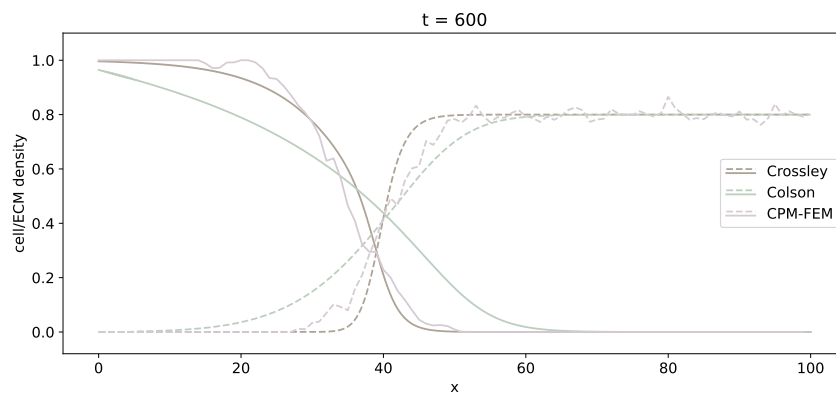

Figure 6.1.: Result of a parameter fit for λ and r using a 200×200-grid and 600 time steps. Solid lines: cell density, dashed lines: ECM density.

It is striking how the roles have changed: now, Crossley's model does not appear to exhibit inadequately steep slopes, but Colson's slopes are too shallow, especially in case of the cells. Of course this is just an example and requires further investigation, but it provides an indication that domain sizes are essential concerning model comparability.

The reason behind the very low choices of spatial and temporal domain sizes were the computational expenses. Even with parallel computing methods that were used for the sensitivity analysis model evaluations, the runtimes were too long to use a larger domain. In potential future works with this model, one of the first aspects to be considered should be runtime optimization and the use of computers with more processors.

6.2. Conclusions

The result that might be overall most surprising is the difference in parameter sensitivity, given that in one dimension, all models produce travelling wave solutions. While in the PDE models, the investigated parameters mostly produce direct, first-order effects that can therefore be easily retraced to the model-inherent processes, the interplay of parameters in the CPM-FEM model is more complex and except for the proliferation probability, no single parameter has a distinct first-order influence on output variance. Considering the model complexity, a category which is dominated by the CPM-FEM model, it might on the other hand be evident that this model's output depends on more nested parameter interactions than the ones of the PDE models.

Another surprising finding, for which an explanation is not as obvious, is the low quality of Colson's wave speed estimate using all parameters for the fit. It is an astonishing result considering that the shapes of the curves coincide well when using the optimized parameters. But perhaps the problem is precisely the optimization? For Colson's model, a different minimizer had to be used given the condition on the parameters for the initially occupied domain. As a test, Crossley's model was fitted to data from the full CPM-FEM model using the same algorithm as was used for Colson. Both, the shape fit and the wave speed approximation were comparably good as in section 5.2, see the summary in appendix A.10. Therefore, the poor wave speed estimations necessarily are a model-inherent issue, and it is likely that the neglect of ECM volume-filling effects are a cause of it, given that Crossley's model performed well in this category under all experimental conditions.

In the cases where only λ and r were fitted, the shapes of the invading and retracting fronts provided by the CPM-FEM model were reproduced rather poorly by both PDE models. However, the wave speeds were approximated relatively well, and it is very likely

that the use of more than a single time step (or two, to be precise, as the time of origin is known as well) for the fit will increase its quality even further.

As a last point regarding the parameter fits it should be mentioned that while the PDE models can be adapted to the CPM-FEM model results, yielding more or less adequate approximations, by model definition it is not possible to regulate the influence of more complex processes such as durotaxis or contact energy expenses using these models. However, these phenomena have significant effects on invasion speed as it has been found here, at least under the limitations listed above.

Regarding these effects, one finding that was not really looked for but is worth mentioning, is how in the early time steps after initialization, the contact cost is a major factor influencing cell shape. Van Oers et al. made a similar observation regarding substrate stiffness which guides cell shape, but they do not mention contact cost as a determining parameter [50]. There is some experimental evidence that the shape of epithelial cells is influenced by its cell-to-cell adhesiveness [92], so perhaps this mechanism is also present in cancer cells – of course not to the extent that was observed in section 5.1.3, but it is interesting that the CPM-FEM model appears to exhibit a similar phenomenon.

One of the motivating questions of this work is which areas of application might be best suited for each of the considered models. Of course, all three models have their right to exist, from the mathematical point of view alone. The strengths of the PDE models are their transparency regarding parameter sensitivity and their ability to estimate the invasion speed with only two degrees of freedom. Especially Crossley's model reliably approximated the wave speeds in the scenarios tested here, hence it could be used for situations as described in the motivation: predicting the tumor growth by estimating the invasion velocity. The CPM-FEM model incorporates too many parameters to be adequate for such purposes, but with its versatility it could be used to replace expensive lab experiments. Yet, before the models can be considered for these tasks, naturally they would need to be tested and calibrated extensively with experimental data – in the best case not only with datasets from *in vitro*, but from *in vivo* studies.

6.3. Next Steps

Any further work involving a model similar to the CPM-FEM model should start by improving its computational performance, for instance by grouping the different tasks in the algorithm more efficiently. Then, the results from this work could be validated or rejected by repeating the presented experiments on a larger temporal and spatial domain. This would not only be interesting in regard to model comparability, but also concerning

the invasion speed differences caused by the in- or exclusion of the diverse Hamiltonian components.

Another possible working point is the reduction of simplifying assumptions on the CPM-FEM model. Especially the modeling of durotaxis on a substrate that behaves in a nonlinear-elastic manner could provide new insights, as well as the implementation of dynamic deformation. Also, extending the model to three dimensions and featuring the degradation-dependent switch between amoeboid and mesenchymal migration could lead to interesting findings.

Finally, regarding applications, the most important step is to test the ability of all models that were discussed in this work to reproduce experimental data. A good starting point could be the *in vitro* data that was used by Browning et al. for their calibrations [51], which can be found in [93].

Bibliography

[1] Y. Liu, R. Wu, and A. Yang, "Research on medical problems based on mathematical models," *Mathematics*, vol. 11, no. 13, p. 2842, 2023.

[2] C. Frantz, K. M. Stewart, and V. M. Weaver, "The extracellular matrix at a glance," *Journal of Cell Science*, vol. 123, no. Pt 24, pp. 4195–4200, 2010.

[3] B. Yue, "Biology of the extracellular matrix: an overview," *Journal of glaucoma*, vol. 23, no. 8 Suppl 1, pp. 20–23, 2014.

[4] O. Lieleg, "Cartilage and osteoarthritis: Cartilage structure and function, self-lubrication, mimicking cartilage lubrication." Lecture slides Introduction to Bio-engineering – Bio-Inspired Material Design, School of Engineering and Design, Technische Universität München, Summer Term 2021.

[5] Q. Wen and P. A. Janmey, "Effects of non-linearity on cell-ECM interactions," *Experimental cell research*, vol. 319, no. 16, pp. 2481–2489, 2013.

[6] R. S. Sopher, H. Tokash, S. Natan, M. Sharabi, O. Shelah, O. Tchaicheeyan, and A. Lesman, "Nonlinear elasticity of the ECM fibers facilitates efficient intercellular communication," *Biophysical Journal*, vol. 115, no. 7, pp. 1357–1370, 2018.

[7] S. Vanni, B. C. Lagerholm, C. Otey, D. L. Taylor, and F. Lanni, "Internet-based image analysis quantifies contractile behavior of individual fibroblasts inside model tissue," *Biophysical Journal*, vol. 84, no. 4, pp. 2715–2727, 2003.

[8] J. P. Winer, S. Oake, and P. A. Janmey, "Non-linear elasticity of extracellular matrices enables contractile cells to communicate local position and orientation," *PLoS ONE*, vol. 4, no. 7, p. e6382, 2009.

[9] J. Notbohm, A. Lesman, D. A. Tirrell, and G. Ravichandran, "Quantifying cell-induced matrix deformation in three dimensions based on imaging matrix fibers," *Integrative biology : quantitative biosciences from nano to macro*, vol. 7, no. 10, pp. 1186–1195, 2015.

[10] S. Münster, L. M. Jawerth, B. A. Leslie, J. I. Weitz, B. Fabry, and D. A. Weitz, "Strain history dependence of the nonlinear stress response of fibrin and collagen networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 30, pp. 12197–12202, 2013.

[11] K. R. Levental, H. Yu, L. Kass, J. N. Lakins, M. Egeblad, J. T. Erler, S. F. T. Fong, K. Csiszar, A. Giaccia, W. Weninger, M. Yamauchi, D. L. Gasser, and V. M. Weaver, "Matrix crosslinking forces tumor progression by enhancing integrin signaling," *Cell*, vol. 139, no. 5, pp. 891–906, 2009.

[12] W. Zhang, Y. Liu, and H. Zhang, "Extracellular matrix: an important regulator of cell functions and skeletal muscle development," *Cell & Bioscience*, vol. 11, no. 1, pp. 1–13, 2021.

[13] A. M. Mohamed, "An overview of bone cells and their regulating factors of differentiation," *The Malaysian Journal of Medical Sciences : MJMS*, vol. 15, no. 1, pp. 4–12, 2008.

[14] N. K. Karamanos, A. D. Theocharis, Z. Piperigkou, D. Manou, A. Passi, S. S. Skandalis, D. H. Vynios, V. Orian-Rousseau, S. Ricard-Blum, C. E. H. Schmelzer, L. Duca, M. Durbeej, N. A. Afratis, L. Troeberg, M. Franchi, V. Masola, and M. Onisto, "A guide to the composition and functions of the extracellular matrix," *The FEBS Journal*, vol. 288, no. 24, pp. 6850–6912, 2021.

[15] J. D. Hood and D. A. Cheresh, "Role of integrins in cell invasion and migration," *Nature reviews. Cancer*, vol. 2, no. 2, pp. 91–100, 2002.

[16] H. Hamidi and J. Ivaska, "Every step of the way: integrins in cancer progression and metastasis," *Nature Reviews Cancer*, vol. 18, no. 9, pp. 533–548, 2018.

[17] C. Bonnans, J. Chou, and Z. Werb, "Remodelling the extracellular matrix in development and disease," *Nature reviews. Molecular cell biology*, vol. 15, no. 12, pp. 786–801, 2014.

[18] M. V. Plikus, X. Wang, S. Sinha, E. Forte, S. M. Thompson, E. L. Herzog, R. R. Driskell, N. Rosenthal, J. Biernaskie, and V. Horsley, "Fibroblasts: Origins, definitions, and functions in health and disease," *Cell*, vol. 184, no. 15, pp. 3852–3872, 2021.

[19] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000.

[20] N. Sfakianakis and M. A. J. Chaplain, "Mathematical modelling of cancer invasion: A review," pp. 153–172, Springer, Singapore, 2021. Conference Proceedings from the International Conference by the Center for Mathematical Modeling and Data Science, Osaka University.

[21] J. Maloney, "What is the difference between invasion and cell migration assays?." `https://www.platypustech.com/difference-invasion-cell-migration-assays`, 12.05.2022. Accessed: 04.04.2023.

[22] P. Friedl and K. Wolf, "Tumour-cell invasion and migration: diversity and escape mechanisms," *Nature Reviews Cancer*, vol. 3, no. 5, pp. 362–374, 2003.

[23] L. Wullkopf, A.-K. V. West, N. Leijnse, T. R. Cox, C. D. Madsen, L. B. Oddershede, and J. T. Erler, "Cancer cells' ability to mechanically adjust to extracellular matrix stiffness correlates with their invasive potential," *Molecular Biology of the Cell*, vol. 29, no. 20, pp. 2378–2385, 2018.

[24] P. Friedl and S. Alexander, "Cancer invasion and the microenvironment: plasticity and reciprocity," *Cell*, vol. 147, no. 5, pp. 992–1009, 2011.

[25] S. George, J. A. J. Martin, V. Graziani, and V. Sanz-Moreno, "Amoeboid migration in health and disease: Immune responses versus cancer dissemination," *Frontiers in Cell and Developmental Biology*, vol. 10, p. 1091801, 2022.

[26] P. Friedl, E. Sahai, S. Weiss, and K. M. Yamada, "New dimensions in cell migration," *Nature Reviews Molecular Cell Biology*, vol. 13, no. 11, pp. 743–747, 2012.

[27] J. D. Murray, *Mathematical biology: I. An introduction*, vol. 17 of *Interdisciplinary Applied Mathematics*. New York: Springer, 3 ed., 2002.

[28] S. B. Carter, "Haptotaxis and the mechanism of cell motility," *Nature*, vol. 213, no. 5073, pp. 256–260, 1967.

[29] P. J. Delves, ed., *Encyclopedia of immunology*. San Diego: Academic Press, 2 ed., 1998.

[30] B. J. DuChez, A. D. Doyle, E. K. Dimitriadis, and K. M. Yamada, "Durotaxis by human cancer cells," *Biophysical Journal*, vol. 116, no. 4, pp. 670–683, 2019.

[31] C. M. Lo, H. B. Wang, M. Dembo, and Y. L. Wang, "Cell movement is guided by the rigidity of the substrate," *Biophysical Journal*, vol. 79, no. 1, pp. 144–152, 2000.

[32] L. G. Vincent, Y. S. Choi, B. Alonso-Latorre, J. C. Del Álamo, and A. J. Engler, "Mesenchymal stem cell durotaxis depends on substrate stiffness gradient strength," *Biotechnology journal*, vol. 8, no. 4, pp. 472–484, 2013.

[33] D. A. Lauffenburger and A. F. Horwitz, "Cell migration: a physically integrated molecular process," *Cell*, vol. 84, no. 3, pp. 359–369, 1996.

[34] A. D. Bershadsky, N. Q. Balaban, and B. Geiger, "Adhesion-dependent cell mechanosensitivity," *Annual review of cell and developmental biology*, vol. 19, pp. 677–695, 2003.

[35] E. G. Rens and R. M. H. Merks, "Cell shape and durotaxis follow from mechanical cell-substrate reciprocity and focal adhesion dynamics: A unifying mathematical model."

[36] S. Jodele, L. Blavier, J. M. Yoon, and Y. A. DeClerck, "Modifying the soil to affect the seed: role of stromal-derived matrix metalloproteinases in cancer progression," *Cancer metastasis reviews*, vol. 25, no. 1, pp. 35–43, 2006.

[37] Y. Jiang, H. Zhang, J. Wang, Y. Liu, T. Luo, and H. Hua, "Targeting extracellular matrix stiffness and mechanotransducers to improve cancer therapy," *Journal of Hematology & Oncology*, vol. 15, no. 1, p. 34, 2022.

[38] M. Egeblad and Z. Werb, "New functions for the matrix metalloproteinases in cancer progression," *Nature reviews. Cancer*, vol. 2, no. 3, pp. 161–174, 2002.

[39] M. A. J. Chaplain and G. Lolas, "Mathematical modelling of cancer cell invasion of tissue: The role of the urokinase plasminogen activation system," *Mathematical Models and Methods in Applied Sciences*, vol. 15, no. 11, pp. 1685–1734, 2005.

[40] T. R. Cox and J. T. Erler, "Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer," *Disease Models & Mechanisms*, vol. 4, no. 2, pp. 165–178, 2011.

[41] PubMed database. `https://pubmed.ncbi.nlm.nih.gov/`. Accessed: 05.01.2024.

[42] R. P. Araujo and D. L. S. McElwain, "A history of the study of solid tumour growth: the contribution of mathematical modelling," *Bulletin of Mathematical Biology*, vol. 66, no. 5, pp. 1039–1091, 2004.

[43] R. A. Fisher, "The wave of advance of advantageous genes," *Annals of Eugenics*, vol. 7, pp. 355–369, 1937.

[44] A. N. Kolmogorov, I. Petrovskii, and N. S. Piskunov, "A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem," *Moscow University Biological Sciences Bulletin*, vol. 1, no. 6, pp. 1–25, 1937.

[45] R. A. Gatenby, "Models of tumor-host interaction as competing populations: implications for tumor biology and treatment," *Journal of Theoretical Biology*, vol. 176, no. 4, pp. 447–455, 1995.

[46] R. A. Gatenby and E. T. Gawlinski, "A reaction-diffusion model of cancer invasion," *Cancer research*, vol. 56, no. 24, pp. 5745–5753, 1996.

[47] A. Szabó and R. M. H. Merks, "Cellular Potts modeling of tumor growth, tumor invasion, and tumor evolution," *Frontiers in Oncology*, vol. 3, p. 87, 2013.

[48] E. L. Stott, N. F. Britton, J. A. Glazier, and M. Zajac, "Stochastic simulation of benign avascular tumour growth using the potts model," *Mathematical and Computer Modelling*, vol. 30, no. 5-6, pp. 183–198, 1999.

[49] T. H. Adair and J.-P. Montani, *Angiogenesis*. San Rafael (CA): Morgan & Claypool Life Sciences, 2010.

[50] R. F. M. van Oers, E. G. Rens, D. J. LaValley, C. A. Reinhart-King, and R. M. H. Merks, "Mechanical cell-matrix feedback explains pairwise and collective endothelial cell behavior in vitro," *PLOS Computational Biology*, vol. 10, no. 8, p. e1003774, 2014.

[51] A. P. Browning, P. Haridas, and M. J. Simpson, "A bayesian sequential learning framework to parameterise continuum models of melanoma invasion into human skin," *Bulletin of Mathematical Biology*, vol. 81, no. 3, pp. 676–698, 2019.

[52] C. Colson, F. Sánchez-Garduño, H. M. Byrne, P. K. Maini, and T. Lorenzi, "Travelling-wave analysis of a model of tumour invasion with degenerate, cross-dependent diffusion," *Proceedings. Mathematical, physical, and engineering sciences*, vol. 477, no. 2256, p. 20210593, 2021.

[53] P. R. Taylor, R. E. Baker, M. J. Simpson, and C. A. Yates, "Coupling volume-excluding compartment-based models of diffusion at different scales: Voronoi and pseudo-compartment approaches," *Journal of the Royal Society, Interface*, vol. 13, no. 120, 2016.

[54] R. M. Crossley, P. K. Maini, T. Lorenzi, and R. E. Baker, "Travelling waves in a coarse-grained model of volume-filling cell invasion: Simulations and comparisons," *Studies in Applied Mathematics*, vol. 151, pp. 1471—-1497, 2023.

[55] D. A. Luke and K. A. Stamatakis, "Systems science methods in public health: dynamics, networks, and agents," *Annual review of public health*, vol. 33, pp. 357–376, 2012.

[56] M. G. Saunders and G. A. Voth, "Coarse-graining methods for computational biology," *Annual review of biophysics*, vol. 42, pp. 73–93, 2013.

[57] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended potts model," *Physical Review Letters*, vol. 69, no. 13, pp. 2013–2016, 1992.

[58] W. Lenz, "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern," *Zeitschrift für Physik*, vol. 21, pp. 613–615, 1920.

[59] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift für Physik*, vol. 31, no. 1, pp. 253–258, 1925.

[60] S. G. Brush, "History of the Lenz-Ising model," *Reviews of Modern Physics*, vol. 39, no. 4, pp. 883–893, 1967.

[61] R. B. Potts, "Some generalized order-disorder transformations," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, no. 1, pp. 106–109, 1952.

[62] S. Turner and J. A. Sherratt, "Intercellular adhesion and cancer invasion: a discrete simulation using the extended potts model," *Journal of Theoretical Biology*, vol. 216, no. 1, pp. 85–100, 2002.

[63] C. Kuttler, "Stochastic models." Lecture script Mathematical Models in Biology, School of Computation, Information and Technology, Technische Universität München, Winter Term 2023/24.

[64] "Monte Carlo, Metropolis and the Ising model." Lecture Notes Introduction to Mathematical and Computational Methods in Physics, Department of Physics, Washington University in St. Louis, Spring Term 2017.

[65] C. P. Robert, "The Metropolis-Hastings algorithm." `http://arxiv.org/pdf/1504.01896v3`, 2015.

[66] W. A. Wall and A. Gebauer, "Vorlesungsmanuskript Finite Elemente." Lecture notes Finite Elemente, Lehrstuhl für Numerische Mechanik, Technische Universität München, Winter Term 2021/22.

[67] C. A. Lemmon and L. H. Romer, "A predictive model of cell traction forces based on cell geometry," *Biophysical Journal*, vol. 99, no. 9, pp. L78–80, 2010.

[68] C. Storm, J. J. Pastore, F. C. MacKintosh, T. C. Lubensky, and P. A. Janmey, "Nonlinear elasticity in biological gels," *Nature*, vol. 435, no. 7039, pp. 191–194, 2005.

[69] J.-M. Dubois and B. Rouzaire-Dubois, "The influence of cell volume changes on tumour cell proliferation," *European Biophysics Journal*, vol. 33, no. 3, pp. 227–232, 2004.

[70] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice: A guide to assessing scientific models.* Hoboken, NJ: John Wiley & Sons, reprint ed., 2004.

[71] M. T. Islam, S. Tang, C. Liverani, S. Saha, E. Tasciotti, and R. Righetti, "Noninvasive imaging of young's modulus and poisson's ratio in cancers in vivo," *Scientific Reports*, vol. 10, no. 1, p. 7266, 2020.

[72] R. A. Harouaka, M. Nisic, and S.-Y. Zheng, "Circulating tumor cell enrichment based on physical properties," *Journal of laboratory automation*, vol. 18, no. 6, pp. 455–468, 2013.

[73] Y. Aratyn-Schaus, P. W. Oakes, and M. L. Gardel, "Dynamic and structural signatures of lamellar actomyosin force generation," *Molecular Biology of the Cell*, vol. 22, no. 8, pp. 1330–1339, 2011.

[74] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis. The Primer.* Hoboken, N. J. and Chichester: John Wiley & Sons, Ltd, 2007.

[75] I. M. Sobol', "Sensitivity analysis for non-linear mathematical models (engl. transl.)," *Mathematical Modeling and Computational Experiment*, no. 1, pp. 407–414, 1993.

[76] H. Rabitz, Ö. F. Ališ, J. Shorter, and K. Shim, "Efficient input—output model representations," *Computer Physics Communications*, vol. 117, no. 1-2, pp. 11–20, 1999.

[77] I. M. Sobol', "On the systematic search in a hypercube," *SIAM Journal on Numerical Analysis*, vol. 16, no. 5, pp. 790–793, 1979.

[78] C. Lemieux, *Monte Carlo and quasi-Monte Carlo sampling.* Springer series in statistics, New York, NY: Springer, 2009.

[79] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, vol. 145, no. 2, pp. 280–297, 2002.

[80] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, 2010.

[81] T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models," *Reliability Engineering & System Safety*, vol. 52, no. 1, pp. 1–17, 1996.

[82] J. Herman and W. Usher, "SALib version 1.4.7: Sensitivity analysis library for Python." `https://github.com/SALib/SALib`, 2023. Software.

[83] P. C. Hansen and G. Scherer, *Least squares data fitting with applications.* Baltimore: Johns Hopkins University Press, 2014.

[84] Scipy, "Scipy version 1.11.2: Open source scientific tools for Python." `https://www.scipy.org/`, 2023. Software.

[85] H. L. Moreland and J. Dockery, "Estimation of the traveling wave speed in pancreatic islets using high period periodic orbits," *Applied Mathematics and Computation*, vol. 439, p. 127482, 2023.

[86] L. Curtin, A. Hawkins-Daarud, K. G. van der Zee, K. R. Swanson, and M. R. Owen, "Speed switch in glioblastoma growth rate due to enhanced hypoxia-induced migration," *Bulletin of Mathematical Biology*, vol. 82, no. 3, p. 43, 2020.

[87] K. Rani Das and A. H. M. Rahmatullah Imon, "A brief review of tests for normality," *American Journal of Theoretical and Applied Statistics*, vol. 5, no. 1, p. 5, 2016.

[88] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[89] L. Fahrmeir, C. Heumann, R. Künstler, I. Pigeot, and G. Tutz, *Statistik: Der Weg zur Datenanalyse.* Berlin: Springer Berlin and Springer Spektrum, 8, rev. ed. 2016 ed., 2016.

[90] H. T. Banks and M. L. Joyner, "AIC under the framework of least squares estimation," *Applied Mathematics Letters*, vol. 74, pp. 33–45, 2017.

[91] M. A. Gimbrone, S. B. Leapman, R. S. Cotran, and J. Folkman, "Tumor dormancy in vivo by prevention of neovascularization," *The Journal of Experimental Medicine*, vol. 136, no. 2, pp. 261–276, 1972.

[92] R. W. Carthew, "Adhesion proteins and the control of cell shape," *Current Opinion in Genetics & Development*, vol. 15, no. 4, pp. 358–363, 2005.

[93] 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," *Journal of Theoretical Biology*, vol. 423, pp. 13–25, 2017.

[94] A. J. Perumpanani, J. A. Sherratt, J. Norbury, and H. M. Byrne, "A two parameter family of travelling waves with a singular barrier arising from the modelling of extracellular matrix mediated cellular invasion," *Physica D: Nonlinear Phenomena*, vol. 126, no. 3-4, pp. 145–159, 1999.

[95] A. R. A. Anderson, M. A. J. Chaplain, E. L. Newman, R. J. C. Steele, and A. M. Thompson, "Mathematical modelling of tumour invasion and metastasis," *Journal of Theoretical Medicine*, vol. 2, no. 2, pp. 129–154, 2000.

[96] A. R. A. Anderson, "A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion," *Mathematical Medicine and Biology: A Journal of the IMA*, vol. 22, no. 2, pp. 163–186, 2005.

[97] L. Peng, D. Trucu, P. Lin, A. Thompson, and M. A. J. Chaplain, "A multiscale mathematical model of tumour invasive growth," *Bulletin of Mathematical Biology*, vol. 79, no. 3, pp. 389–429, 2017.

[98] N. J. Armstrong, K. J. Painter, and J. A. Sherratt, "A continuum approach to modelling cell-cell adhesion," *Journal of Theoretical Biology*, vol. 243, no. 1, pp. 98–113, 2006.

[99] A. Gerisch and M. A. J. Chaplain, "Mathematical modelling of cancer cell invasion of tissue: local and non-local models and the effect of adhesion," *Journal of Theoretical Biology*, vol. 250, no. 4, pp. 684–704, 2008.

[100] M. Scianna, L. Preziosi, and K. Wolf, "A cellular potts model simulating cell migration on and in matrix environments," *Mathematical Biosciences & Engineering*, vol. 10, no. 1, pp. 235–261, 2013.

[101] D. Mitrossilis, J. Fouchard, A. Guiroy, N. Desprat, N. Rodriguez, B. Fabry, and A. Asnacios, "Single-cell response to stiffness exhibits muscle-like behavior," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 43, pp. 18243–18248, 2009.

[102] A. A. Malik, B. Wennberg, and P. Gerlee, "The impact of elastic deformations of the extracellular matrix on cell migration," *Bulletin of Mathematical Biology*, vol. 82, no. 4, p. 49, 2020.

[103] J. Müller, "Many individuals: Diffusion approximation." Lecture script Mathematical Biology, School of Computation, Information and Technology, Technische Universität München, Winter Term 2022/23.

[104] C. Grossmann, H.-G. Roos, and M. Stynes, *Numerical treatment of partial differential equations.* Universitext, Berlin and New York: Springer, 2007.

[105] B. W. Kernighan and D. M. Ritchie, *The C Programming Language.* Prentice Hall, 1988. Software.

[106] Python Software Foundation, "Python programming language, version 3.9.4." `https://www.python.org`. Software.

[107] OpenAI, "GPT-3.5: Generative pre-trained transformer 3.5." OpenAI API Documentation, 2022. `https://beta.openai.com/docs/`. Software.

[108] Y. S. Khan and A. Farhana, "Statpearls: Histology, cell," 2023.

[109] M. M. Doyley, S. Srinivasan, S. A. Pendergrass, Z. Wu, and J. Ophir, "Comparative evaluation of strain-based and model-based modulus elastography," *Ultrasound in Medicine & Biology*, vol. 31, no. 6, pp. 787–802, 2005.

[110] B. Shin, D. Gopaul, S. Fienberg, and H. J. Kwon, "Application of eshelby's solution to elastography for diagnosis of breast cancer," *Ultrasonic Imaging*, vol. 38, no. 2, pp. 115–136, 2016.

[111] M. Akhmanova, E. Osidak, S. Domogatsky, S. Rodin, and A. Domogatskaya, "Physical, spatial, and molecular aspects of extracellular matrix of in vivo niches and artificial scaffolds relevant to stem cells research," *Stem Cells International*, vol. 2015, p. 167025, 2015.

[112] S.-J. Hao, Y. Wan, Y.-Q. Xia, X. Zou, and S.-Y. Zheng, "Size-based separation methods of circulating tumor cells," *Advanced drug delivery reviews*, vol. 125, pp. 3–20, 2018.

[113] M. Scianna and L. Preziosi, "Modeling the influence of nucleus elasticity on cell invasion in fiber networks and microchannels," *Journal of Theoretical Biology*, vol. 317, pp. 394–406, 2013.

A. Appendix

A.1. More Detailed Cell Invasion Modeling Timeline

This timeline is an extension to the one in section 3.1. It is based on the overview papers by Robyn Araujo and Sean McElwain [42], and by Nikolaos Sfakianakis and Mark A. J. Chaplain [20]. To avoid redundancy, the models that have already been explained in section 3.1 are mentioned in an abbreviated format.

- 1937: Ronald A. Fisher's mutation continuum model.

- 1995: Robert A. Gatenby's cell invasion specific continuum model.

- mid- to late nineties: Abbey J. Perumpanani and colleagues followed the recent findings concerning integrin serving as adhesion control and proteolysis regulator in tumor invasion. They published multiple PDE models that link cell invasion with proteolysis and haptotaxis. Among these models, the most famous one is a diffusion-less three-dimensional reaction-advection system featuring the densities of malignant cells, ECM elements and proteases, the enzymes that degrade the ECM. Diffusion, i.e. random cell motility, was omitted as its occurrence is found to be negligible compared to the haptotaxis-induced directional migration of the invasive cells. [20, 42, 94]

- 1999: Emma L. Stott's CPM.

- 2000: using the same three variables as Perumpanani et al., Alexander R. A. Anderson and colleagues started with the development of another deterministic PDE model. While they did include diffusion, they excluded proliferation, and they found that the malignant cells divide in two groups during the process of detaching from the original tumor: the expression of ECM-degrading enzymes (proteases) by the cancer cells builds a haptotactic gradient in the tumor's immediate neighborhood, so that the outer tumor cells start leaving the original tumor site following this gradient. The main part of the tumor cells, however, remains close to the original site and mostly spreads via diffusion, not so much via haptotaxis. One

can observe that the first, haptotaxis-driven group of cells invades the surrounding tissue much faster than the second, diffusion-driven one. Anderson et al. use this first group to explain the development of metastases. With their continuous model as a basis, they derived a discrete, quasi-stochastic cancer cell invasion model that allows tracking the movement of individual cells depending on their likelihood to move in a certain direction (which is influenced by the cancer cell and ECM element density) [95]. In 2005, Anderson combined both models and added the nutrient oxygen as a fourth variable. ECM elements, proteases and oxygen are handled as densities/concentrations, whereas the tumor cells are discrete agents. His model is considered as a milestone on the way towards hybrid invasion models, where "hybrid" means the combination of macroscopic, deterministic terms and microscopic/atomistic, stochastic ones. [20, 96]

- 2001: inspired by Stott et al., Stephen Turner and Jonathan A. Sherratt modified their version of the CPM to capture malignant tumor invasion. While they omitted healthy and necrotic cells, they included the ECM as an obstructive factor of cancer cell movement. Moving to a lattice site which has been visited by many cells in the past requires less energy than moving to a site which has rarely or never been occupied by a cell, because the visitors were degrading the ECM over time. In an extension of their model, they also included proliferation of tumor cells. Both models exhibit the invading front consisting of outer tumor cells detaching from the original tumor site, as it was first observed by Anderson and colleagues. [62]

- 2005: after several years of observing the same three variables (cancer cells, ECM elements and proteases), Marc A. J. Chaplain and Georgios Lolas investigated the ECM degeneration more closely than their predecessors. They did not assume the proteases to be one single substrate, but the whole biological system that generates and activates these enzymes, namely the urokinase plasminogen activation system. Cancer cells express the enzyme uPA (urokinase plasminogen activator) which converts plasminogen, a chemical precursor of plasmin, to the active ECM-degenerating enzyme plasmin. The opponent of uPA is PAI-1 (plasminogen activator inhibitor-1), and in healthy tissues, uPA expression is very limited and strictly controlled through large abundance of PAI-1 to avoid excessive ECM degeneration. A disruption of this balance can be a sign for malignant activity. Even though Chaplain and Lolas included the whole enzymatic cascade in their continuum model, they achieved a representation which is not too complicated and whose numerical simulations exhibit similar properties as the ones from Anderson et al., especially the presence of multiple invading fronts instead of a single wave [20, 39]. Their model is

modified in 2017 by Lu Peng and colleagues to include not only macroscopic measures (i.e. tissue-scale), but also dynamics on the microscopic scale (i.e. cell-scale), namely the processes by which uPA molecules are produced by the cancer cells and removed by binding to cell-surface receptors. The introduction of a second scale is considered necessary since these processes only occur in the "(micro) neighbourhood of the invasive edge of the tumor" [97].

- 2006: so far, the only type of adhesion that has been considered in the mathematical invasion models was cell-matrix adhesion. But there exists another type of adhesion as well: cell-cell adhesion, which happens through the binding of adhesive substrates at the cell surface. It is a crucial process when it comes to tissue development, enabling cells to selectively adhere to other cells and forming patterns which later build organs. A change in tumor cell adhesion is also a sign for upcoming invasion. The first macroscopic continuum model focusing on cell-cell adhesion was developed by Nicola J. Armstrong and colleagues. Assuming that adhesion attracts cells within a so-called "sensing radius" to move towards each other, it shows how disassociated, adhesive cells find cells of the same type and aggregate with them, initiating the growth of secondary tumors at distinct sites of the body [20, 98]. In 2008, their model is combined with the ideas of Anderson et al.'s discrete haptotaxis-diffusion-model [95] by Alf Gerisch and Marc A. J. Chaplain. This new model describes cancer invasion by both cell-cell adhesion and haptotaxis. [20, 99]

- 2013: the CPM gets refined more and more: Marco Scianna, Luigi Preziosi and Katarina Wolf made their cells expend more or less energy depending on mechanical properties of the ECM and introduce details such as nucleus and cytosolic region in the cells [100].

- 2014: René F. M. van Oers' hybrid model.

- 2019: Alexander P. Browning's calibrated continuum model.

- 2020: shortly after the mathematical investigations of durotaxis took place, the influence of ECM deformation on cell migration became interesting. Adam A. Malik and colleagues created a single cell model which they compared to experimental data from durotactic cancer cells. Similarly to the work of van Oers, Rens and Merks [35, 50], they assumed the ECM to consist of a linearly elastic material, but they neglected substrate-induced adhesion completely and focused on mechanical interdependences. The cell in their model has four contact points with the ECM underneath it, and forces within the cell are approximated through spring forces

between the contact points and the cell kernel. It has been found that the traction forces generated by the cells are proportional to the substrate stiffness [101], and as soon as Malik et al. included this property in their model, they observed durotaxis. [102]

- 2021: Colson's minimal continuum model.

- 2023: Rebecca M. Crossley's agent-based model.

A.2. Coarse-Graining Crossley's Agent-Based Model

The probabilities $T_{i\pm}^{\mathrm{m}^j}(t)$, $T_i^{\mathrm{p}^j}(t)$ and $T_i^{\mathrm{d}^j}(t)$ from section 3.2 are coarse-grained in the following to obtain a system of reaction-diffusion equations. The description follows the procedure from [54].

Let J be the number of samples that is taken from the model at time t. Then one can define the average occupancy of lattice site i by

$$\langle u_i(t)\rangle := \frac{1}{J}\sum_{j=1}^{J} = u_i^j(t), \qquad \langle m_i(t)\rangle := \frac{1}{J}\sum_{j=1}^{J} = m_i^j(t).$$

The evolution of the mean cell number in site i during $[t, t+\tau)$ is given by the sum of all possible events multiplied with their probabilities:

$$\langle u_i(t+\tau)\rangle = \langle u_i(t)\rangle$$
$$+ \text{ number of cells from site } i+1 \text{ moving to } i$$
$$+ \text{ number of cells from site } i-1 \text{ moving to } i$$
$$- \text{ number of cells from site } i \text{ moving to } i+1$$
$$- \text{ number of cells from site } i \text{ moving to } i-1$$
$$+ \text{ number of offspring created in site } i,$$

which, using the notation from the agent-based model, looks like this:

$$\langle u_i(t+\tau)\rangle = \langle u_i(t)\rangle$$
$$+ \frac{p_{\mathrm{m}}}{2}\langle u_{i+1}(t)\rangle\left(1 - \frac{\langle u_i(t)\rangle + \langle m_i(t)\rangle}{N}\right)$$
$$+ \frac{p_{\mathrm{m}}}{2}\langle u_{i-1}(t)\rangle\left(1 - \frac{\langle u_i(t)\rangle + \langle m_i(t)\rangle}{N}\right)$$

$$- \frac{p_\mathrm{m}}{2} \langle u_i(t) \rangle \left(1 - \frac{\langle u_{i+1}(t) \rangle + \langle m_{i+1}(t) \rangle}{N} \right)$$

$$- \frac{p_\mathrm{m}}{2} \langle u_i(t) \rangle \left(1 - \frac{\langle u_{i-1}(t) \rangle + \langle m_{i-1}(t) \rangle}{N} \right)$$

$$+ \frac{p_\mathrm{p}}{2} \langle u_i(t) \rangle \left(1 - \frac{\langle u_i(t) \rangle + \langle m_i(t) \rangle}{N} \right).$$

Crossley rearranges the equation above, expands some of the right hand side terms with Δ^2/Δ^2, where Δ is the lattice spacing, and divides by τ:

$$\begin{aligned} \frac{\langle u_i(t+\tau) \rangle - \langle u_i(t) \rangle}{\tau} = {} & \frac{p_\mathrm{m}\Delta^2}{2\tau} \cdot \frac{\langle u_{i-1}(t) \rangle - 2\langle u_i(t) \rangle + \langle u_{i+1}(t) \rangle}{\Delta^2} \\ & + \frac{p_\mathrm{m}\Delta^2}{2\tau N} \cdot \frac{\langle u_i(t) \rangle \cdot (\langle m_{i-1}(t) \rangle - 2\langle m_i(t) \rangle + \langle m_{i+1}(t) \rangle)}{\Delta^2} \\ & - \frac{p_\mathrm{m}\Delta^2}{2\tau N} \cdot \frac{\langle m_i(t) \rangle \cdot (\langle u_{i-1}(t) \rangle - 2\langle u_i(t) \rangle + \langle u_{i+1}(t) \rangle)}{\Delta^2} \\ & + \frac{p_\mathrm{p}}{\tau} \langle u_i(t) \rangle \left(1 - \frac{\langle u_i(t) \rangle + \langle m_i(t) \rangle}{N} \right), \end{aligned}$$

To obtain the PDE that describes the densities of cells and ECM elements at position $x \in \mathbb{R}$ and time $t \in (0, \infty)$, a Kramers-Moyal-Expansion [103] is used. To do this, the equation above is interpreted as a master equation, meaning that the left hand side is considered as the derivative $\langle \dot{u}_i(t) \rangle$. This is possible by taking the limit $\tau \to 0$, which of course also affects the factors on the right hand side. Crossley defines

$$D := \lim_{\tau \to 0} \frac{p_\mathrm{m}\Delta^2}{2\tau}, \qquad r := \lim_{\tau \to 0} = \frac{p_\mathrm{p}}{\tau}.$$

Let $x = hi$ where $h = \Delta$. The basic assumption of the Kramers-Moyal-Expansion is that there exists a smooth function $u(x,t)$ with $hu(x,t) \approx \langle u_i(t) \rangle$ (and similarly $hm(x,t) \approx \langle m_i(t) \rangle$). The functions u and m are the "continuum counterparts" [54] of the number density of cells $\langle u_i(t) \rangle/\Delta$ and the density of ECM $\langle m_i(t) \rangle/\Delta$.[1] These definitions allow to rewrite for instance $\langle u_{i-1}(t) \rangle = hu(h(i-1), t) = hu(x-h, t)$ (there are analogue formulations for $\langle u_{i+1}(t) \rangle$, $\langle m_{i-1}(t) \rangle$ and $\langle m_{i+1}(t) \rangle$), which is plugged into the equation above:

---

[1]Indeed, Crossley et al. use $\langle m_i(t) \rangle/(\mu\Delta)$. The factor μ is a conversion factor between number of cells and unit mass of ECM. μ is not relevant for the purposes described here, so it is omitted for the sake of simplicity.

$$h\partial_t u(x,t) \approx \langle \dot{u}_i(t)\rangle = D\frac{hu(x-h,t) - 2hu(x,t) + hu(x+h,t)}{\Delta^2}$$
$$+ \frac{D}{N}\cdot\frac{hu(x,t)\left(hm(x-h,t) - 2hm(x,t) + hm(x+h,t)\right)}{\Delta^2}$$
$$- \frac{D}{N}\cdot\frac{hm(x,t)\left(hu(x-h,t) - 2hu(x,t) + hu(x+h,t)\right)}{\Delta^2}$$
$$+ rhu(x,t)\left(1 - \frac{hu(x,t) + hm(x,t)}{N}\right).$$

To get rid of $x \pm h$ in the function calls of u and m, a second order Taylor expansion is used of the form

$$u(x+h,t) = u(x,t) + h\partial_x u(x,t) + \frac{h^2}{2}\partial_{xx}u(x,t) + \mathcal{O}(h^3)$$
$$u(x-h,t) = u(x,t) - h\partial_x u(x,t) + \frac{h^2}{2}\partial_{xx}u(x,t) + \mathcal{O}(h^3)$$

(analogously for $m(x+h,t)$, $m(x-h,t)$). Plugging in these expressions – while neglecting the third order terms and simplifying a bit – yields

$$h\partial_t u(x,t) = D\frac{h^3}{\Delta^2}\partial_{xx}u(x,t)$$
$$+ \frac{D}{N}hu(x,t)\cdot\frac{h^3}{\Delta^2}\partial_{xx}m(x,t)$$
$$- \frac{D}{N}hm(x,t)\cdot\frac{h^3}{\Delta^2}\partial_{xx}u(x,t)$$
$$+ rhu(x,t)\left(1 - \frac{hu(x,t) + hm(x,t)}{N}\right).$$

Dividing both sides by Δ and recalling that $h = \Delta$, leads to

$$\partial_t u(x,t) = D\partial_{xx}u(x,t) + \frac{D\Delta}{N}u(x,t)\partial_{xx}m(x,t)$$
$$- \frac{D\Delta}{N}m(x,t)\partial_{xx}u(x,t) + ru(x,t)\left(1 - \frac{\Delta(u(x,t) + m(x,t))}{N}\right).$$

In the last step, another definition is introduced, namely $K := \frac{N}{\Delta}$. Using it, one eventually arrives at the Fokker-Planck equation for the cell density dynamics:

$$\partial_t u(x,t) = D\left(\partial_{xx}u(x,t) + \frac{u(x,t)\partial_{xx}m(x,t)}{K} - \frac{m(x,t)\partial_{xx}u(x,t)}{K}\right)$$
$$+ ru(x,t)\left(1 - \frac{u(x,t) + m(x,t)}{K}\right).$$

Of course the formulation for $m(x,t)$ is still missing. The corresponding PDE for the ECM dynamics can be derived with the same approach, but the calculations are now much simpler as there are not as many factors of influence – the number of ECM elements is only changed by the degradation through cells in the same lattice site. Hence it is

$$\langle m_i(t+\tau)\rangle = \langle m_i(t)\rangle - \text{number of ECM elements degraded in site } i$$
$$= \langle m_i(t)\rangle - p_{\mathrm{d}}\langle u_i(t)\rangle\langle m_i(t)\rangle.$$

After following the same steps as before for the cell density, the ECM density is given by

$$\partial_t m(x,t) = -\lambda u(x,t)m(x,t),$$

where $\lambda := \lim_{\tau \to 0} \frac{p_{\mathrm{d}}\Delta}{\tau}$.

A.3. Numerical Solution of the PDE Models

The models by Crossley and Colson are given as PDEs. Solving them numerically requires both, spatial and temporal discretization. Discretization in time is a task that can reliably be handled by software – here for instance by the Python function `scipy.integrate.solve_ivp` [84] which uses a fourth order Runge-Kutta method and can handle moderate stiffness –, as it only requires solving a system of ordinary differential equations (ODEs) instead of PDEs. The crucial step is the transformation from the PDE system to the ODE system, which is done by discretization in space. In the following, it is demonstrated how the spatial discretization works with the dimensionless model by Crossley (3.3). For the parameterized system as well as Colson's model, it is done analogously. The equation for $\tilde{u}$ can be rewritten (dropping the tildes for simplicity) as

$$\partial_t u = \partial_x((1-m)\,\partial_x u) + \partial_x(u\,\partial_x m) + u(1-u-m). \tag{A.1}$$

A term of the shape $\partial_x(f\,\partial_x g)$, $f = f(x,t)$, $g = g(x,t)$, appears twice in this equation: once with $f = 1 - m$, $g = u$, once with $f = u$, $g = m$. Applying the product rule yields

$$\partial_x(f\,\partial_x g) = \partial_x f\,\partial_x g + f\,\partial_{xx} g. \tag{A.2}$$

As proposed by Crossley et al., the spatial domain $[0,L]$, $L > 0$ is discretized in increments of length $h := 0.1$, yielding a total of $I := \frac{L}{h}$ spatial points (assume L to be a multiple of h). Then, the notation $f_i(t) := f(hi,t)$ is introduced, where $i \in \{1,...,I\}$.

A. Appendix

Following the basic concepts of finite differences (see for instance [104]), the first order derivative can be approximated by a forward difference quotient

$$\partial_x f \Big|_{x=hi} \approx \frac{1}{h}(f_i - f_{i+1}),$$

or a backward difference quotient

$$\partial_x f \Big|_{x=hi} \approx \frac{1}{h}(f_{i-1} - f_i),$$

and the second order derivative by a central difference quotient

$$\partial_{xx} f \Big|_{x=hi} \approx \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}).$$

Note that in (A.2), there is a product of two first order derivatives. For balance, it is approximated using the average of a forward and backward difference quotient, yielding

$$\partial_x f \, \partial_x g \Big|_{x=hi} \approx \frac{1}{2h^2} \left((f_i - f_{i+1})(g_i - g_{i+1}) + (f_{i-1} - f_i)(g_{i-1} - g_i) \right).$$

In total, the spatially discretized version of (A.2) is given by

$$\begin{aligned}
\partial_x (f \, \partial_x g) \Big|_{x=hi} \approx & \frac{1}{2h^2} \left((f_i - f_{i+1})(g_i - g_{i+1}) + (f_{i-1} - f_i)(g_{i-1} - g_i) \right) \\
& + \frac{1}{h^2} f_i (g_{i-1} - 2g_i + g_{i+1}) \\
= & \frac{1}{2h^2} \left(-f_i(g_{i-1} - g_i) + f_i(g_i - g_{i+1}) - f_{i+1}(g_i - g_{i+1}) + f_{i-1}(g_{i-1} - g_i) \right. \\
& \left. + 2f_i(g_{i-1} - 2g_i + g_{i+1}) \right) \\
= & \frac{1}{2h^2} \left(-f_i(g_{i-1} - 2g_i + g_{i+1}) - f_{i+1}(g_i - g_{i+1}) + f_{i-1}(g_{i-1} - g_i) \right. \\
& \left. + 2f_i(g_{i-1} - 2g_i + g_{i+1}) \right) \\
= & \frac{1}{2h^2} \left(f_i(g_{i-1} - 2g_i + g_{i+1}) - f_{i+1}(g_i - g_{i+1}) + f_{i-1}(g_{i-1} - g_i) \right) \\
= & \frac{1}{2h^2} \left((f_i + f_{i-1})g_{i-1} - (f_{i-1} + 2f_i + f_{i+1})g_i + (f_i + f_{i+1})g_{i+1} \right).
\end{aligned}$$

Plugging in $f = 1 - m$, $g = u$, and $f = u$, $g = m$, as it is given in (A.1), leads to

$$\begin{aligned}
\partial_x((1-m)\,\partial_x u) + \partial_x(u\,\partial_x m) \Big|_{x=hi} \approx & \frac{1}{2h^2}((1 - m_i + 1 - m_{i-1})u_{i-1} \\
& - (1 - m_{i-1} + 2(1 - m_i) + 1 - m_{i+1})u_i + (1 - m_i + 1 - m_{i+1})u_{i+1}
\end{aligned}$$

XX

$$+ (u_i + u_{i-1})m_{i-1} - (u_{i-1} + 2u_i + u_{i+1})m_i + (u_i + u_{i+1})m_{i+1})$$

$$= \frac{1}{2h^2}((2 - m_i - m_{i-1})u_{i-1}$$

$$- (4 - m_{i-1} - 2m_i - m_{i+1})u_i + (2 - m_i - m_{i+1})u_{i+1}$$

$$+ (u_i + u_{i-1})m_{i-1} - (u_{i-1} + 2u_i + u_{i+1})m_i + (u_i + u_{i+1})m_{i+1})$$

$$= \frac{1}{2h^2} \cdot 2(u_{i-1}(1 - m_i) + u_i(m_{i-1} + m_{i+1} - 2) + u_{i+1}(1 - m_i))$$

$$= \frac{1}{h^2}(u_{i-1}(1 - m_i) + u_i(m_{i-1} + m_{i+1} - 2) + u_{i+1}(1 - m_i)).$$

Eventually, the discretization of the whole system (3.3) is given by[2]

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{1}{h^2}(u_{i-1}(1 - m_i) + u_i(m_{i-1} + m_{i+1} - 2) + u_{i+1}(1 - m_i)) + u_i(1 - u_i - m_i),$$

$$\frac{\mathrm{d}m_i}{\mathrm{d}t} = -\lambda m_i u_i,$$

for $i \in \{1, ..., I - 1\}$. For the cases $i = 0$ and $i = I$, Crossley uses the no-flux boundary condition which allows to set $u_{-1}(t) = u_0(t)$ and $u_{I+1}(t) = u_I(t)$ because no domain-boundary-crossing changes in cell density are expected (equivalently for m_{-1} and m_{I+1}). Also, one can set $m_0 = m_1 = 0$ as the domain length for the simulation should be chosen large enough such that no ECM elements are expected here. This yields the boundary conditions

$$\frac{\mathrm{d}u_0}{\mathrm{d}t} = \frac{1}{h^2}(u_0(1 - m_0) + u_0(m_0 + m_1 - 2) + u_1(1 - m_0)) + u_0(1 - u_0 - m_0)$$

$$= \frac{1}{h^2}(u_0 - 2u_0 + u_1) + u_0(1 - u_0 - m_0)$$

$$= \frac{1}{h^2}(u_1 - u_0) + u_0(1 - u_0 - m_0),$$

and

$$\frac{\mathrm{d}u_I}{\mathrm{d}t} = \frac{1}{h^2}(u_{I-1}(1 - m_I) + u_I(m_{I-1} + m_I - 2) + u_I(1 - m_I)) + u_I(1 - u_I - m_I)$$

$$= \frac{1}{h^2}(u_{I-1} - u_I - u_{I-1}m_I + u_I m_{I-1}) + u_I(1 - u_I - m_I).$$

Note that no such assumptions are necessary for the equations for the ECM density, as $\frac{\mathrm{d}m_i}{\mathrm{d}t}$ only depends on m_i and u_i, i.e. the same indices.

---

[2]Note that here, the final discretization differs from the one in equations (B2), (B5) and (B6) of [54]. The reason for this is a miscalculation in the publication. The version used in this work should be correct.

The spatially discretized system of equations can now be handed over to an initial value problem solver such as the before-mentioned `scipy.integrate.solve_ivp`. As initial conditions, Crossley uses

$$u(x,0) = \begin{cases} 1 & \text{if } x < \alpha, \\ 0 & \text{if } x \geq \alpha, \end{cases}$$

$$m(x,0) = \begin{cases} 0 & \text{if } x < \alpha, \\ \hat{m}_0 & \text{if } x \geq \alpha, \end{cases}$$

where $\alpha > 0$ represents the length of the initially invaded region at $t = 0$ and $\hat{m}_0 \in [0,1)$ describes the density of non-degraded ECM.

While the same discretization scheme and analogue boundary conditions are used for Colson's model, the initial conditions differ. By introducing a sigmoid shaped function and two smoothing parameters $0 < \omega < \sigma$, Colson's initial conditions are not as sharp:

$$u(x,0) = \begin{cases} 1 & \text{if } x < \sigma - \omega, \\ \exp\left(1 - \frac{1}{1-\left(\frac{x-\sigma+\omega}{\omega}\right)^2}\right) & \text{if } \sigma - \omega \leq x < \sigma, \\ 0 & \text{if } x \geq \sigma, \end{cases}$$

$$m(x,0) = \begin{cases} 0 & \text{if } x < \sigma - \omega, \\ \hat{m}_0(1 - u(x,0)) & \text{if } \sigma - \omega \leq x < \sigma, \\ \hat{m}_0 & \text{if } x \geq \sigma. \end{cases}$$

A.4. Constitution of the Stiffness Matrix K and the Finite-Element Equations

To gather more theoretical background on equation (3.7), the strong and the weak form of the problem as well as the discretization of the latter is derived here, using some terms and explanations from [66]. In two-dimensional linear elasto-statics, the strong form of the problem is given by three governing equations and two boundary conditions. The first of these equations is the so-called **Balance Equation**, it states that externally applied loads $\hat{b}$ (coming from the cells) must be balanced by the resulting stress σ within the material.

It is given by

$$0 = L^{\mathrm{T}}\sigma + \hat{b} = \begin{pmatrix} \partial_{x_1} & 0 & \partial_{x_2} \\ 0 & \partial_{x_2} & \partial_{x_1} \end{pmatrix} \sigma + \hat{b},$$

where ∂_{x_i} means the partial derivative with respect to x_i, i.e. L^{T} is the divergence operator. Note that stress – as well as strain – is a three-dimensional vector: $\sigma = (\sigma_{11}, \sigma_{22}, \sigma_{12})^{\mathrm{T}}$, where σ_{11} and σ_{22} are the normal stresses in x_1- and x_2-direction, respectively, and σ_{12} is the shear stress (in two dimensions, it is assumed that the shear stress σ_{21} parallel to a surface that is oriented in x_1-direction is the same as the shear stress σ_{12} parallel to a surface oriented in x_2-direction, hence only one of them is given in σ).

The next equation is the **Kinematic Equation** which describes the strain ε, i.e. how the material is stretched and compressed, as a result of the displacements u:

$$\varepsilon = Lu,$$

where the strain is similarly defined as the stress, but the third component is doubled: $\varepsilon = (\varepsilon_{11}, \varepsilon_{22}, 2\varepsilon_{12})^{\mathrm{T}}$.

The last of the governing equations is the **Constitutive Equation** which links the stresses and strains under the assumption of given material properties. Van Oers assumes plane stress conditions, meaning there is a constant mechanical stress state across the thickness of the ECM. This leads to a material or constitutive matrix of the following form:

$$C = \frac{E}{1 - \nu^2} \begin{pmatrix} 1 & \nu & 0 \\ & 1 & 0 \\ \text{sym.} & & \frac{1-\nu}{2} \end{pmatrix},$$

where $E > 0$ is Young's modulus and $-1 \leq \nu \leq \frac{1}{2}$ is Poisson's ratio, both material specific parameters. The equation is given by

$$\sigma = C\varepsilon.$$

Van Oers uses Dirichlet boundary conditions for all nodes at the grid boundary Γ: their displacement is always 0. This condition is emphasized by the constraint that the CPM is not allowed to place any cells at boundary lattice sites, leading to flux boundary conditions that suggest themselves: they are given by zero traction along the outward normal vector n.

In summary, the governing differential equation of the problem is given by a rearrangement of the field equations inside the domain Ω

$$0 = L^{\mathrm{T}}CLu + \hat{b} \tag{A.3}$$

and the boundary conditions on Γ

$$u = 0$$

$$\langle \sigma, n \rangle = 0.$$

To obtain the problem's weak form, the displacements u are only required to satisfy the governing differential equation in an integral sense instead of pointwise satisfaction. This means (A.3) becomes

$$\int_{\Omega} \langle L^{\mathrm{T}}CLu + \hat{b}, w \rangle \, \mathrm{d}\Omega = \int_{\Omega} \langle CLu, Lw \rangle \, \mathrm{d}\Omega + \int_{\Omega} \langle \hat{b}, w \rangle \, \mathrm{d}\Omega = 0, \tag{A.4}$$

where a test function w was inserted to ensure that any approximate solutions u do not deviate too much from the analytical solution (for instance, if u solves the equation above, any function $u + u^*$ with $\int_{\Omega} L^{\mathrm{T}}CLu^*\mathrm{d}\Omega = 0$ solves that equation as well, even though u^* could be a periodic function with very large amplitude leading to massive errors). In general, w is an element of the test function space which is given by $\mathcal{W} := \{w : \Omega \to \mathbb{R}^2 : w(x) = 0$ for $x \in \Gamma$ and the above integral exists$\}$, but it is possible to restrict the choice of w (and also the choice of trial functions, i.e. the functions that are used as candidates for approximate solutions u) using the discretization through the grid. After finishing the boundary value problem's weak form in the next step, the specific choice of test and trial functions are discussed.

This form differs from the one in (A.4) as it implicitly features the boundary conditions on Γ. An exact solution u of the problem (A.3) fulfills the Balance Equation, i.e. $L^{\mathrm{T}}\sigma(u)+\hat{b} = 0$ in Ω, where $\sigma(u)$ means that the stress vector is calculated with the solution u (with this approach, the Kinematic and Constitutive Equation are implicitly used, hence it is not necessary to consider them explicitly). However, if u is an approximate solution, it does not fulfill the equation, and a residual $\mathcal{R}_{\mathrm{BE}} := L^{\mathrm{T}}\sigma(u) + \hat{b} \neq 0$ remains. The same holds true for the flux boundary condition, yielding the residual $\mathcal{R}_{\mathrm{FBC}} := -\langle \sigma(u), n \rangle \neq 0$ on Γ (as the sign does not matter, it is chosen to be negative which will be helpful in the next steps). Just as in the derivation of (A.4), only satisfaction of the governing equations and the boundary conditions over the integral is required, which leads to the weighted residual formulation

$$\int_\Omega \langle \mathcal{R}_{\text{BE}}, w \rangle \, \mathrm{d}\Omega + \int_\Gamma \langle \mathcal{R}_{\text{FBC}}, w \rangle \, \mathrm{d}\Gamma = 0,$$

where w are weighting functions that can be chosen just as the test functions (and therefore the same notation is used). This equation can be rearranged using the Divergence Theorem:

$$
\begin{aligned}
0 &= \int_\Omega \langle L^{\text{T}} \sigma(u) + \hat{b}, w \rangle \, \mathrm{d}\Omega + \int_\Gamma \langle -\langle \sigma(u), n \rangle, w \rangle \, \mathrm{d}\Gamma \\
&= \int_\Gamma \langle \langle \sigma(u), w \rangle, n \rangle \, \mathrm{d}\Gamma - \int_\Omega \langle \sigma(u), Lw \rangle \, \mathrm{d}\Omega + \int_\Omega \langle \hat{b}, w \rangle \, \mathrm{d}\Omega - \int_\Gamma \langle \langle \sigma(u), n \rangle, w \rangle \, \mathrm{d}\Gamma.
\end{aligned}
$$

Conveniently, the first and the last term cancel and the weak form of the boundary value problem is obtained by

$$\int_\Omega \langle \sigma(u), Lw \rangle \, \mathrm{d}\Omega = \int_\Omega \langle \hat{b}, w \rangle \, \mathrm{d}\Omega. \tag{A.5}$$

Note that the Dirichlet boundary conditions do not appear in the weak form, they have to be respected by only allowing trial and test functions that fulfill these conditions.

The choice of trial and test functions is resumed: Each quadratic element has four nodes (one at each corner), so the simplest choice are linear shape functions (with local coordinates r and s):

$$
\begin{aligned}
N_1(r, s) &= \frac{1}{4}(1 - r)(1 - s) \\
N_2(r, s) &= \frac{1}{4}(1 + r)(1 - s) \\
N_3(r, s) &= \frac{1}{4}(1 + r)(1 + s) \\
N_4(r, s) &= \frac{1}{4}(1 - r)(1 + s)
\end{aligned}
\tag{A.6}
$$

Hence, for each element, one can use the trial function $u_h^{(e)}$ (the subscript h symbolizes that this is the discrete solution, and the superscript (e) indicates local element level):

$$u_h^{(e)}(r, s) = \sum_{i=1}^4 N_i(r, s) d_i,$$

where $d_i = \left(d_i^1, d_i^2 \right)^{\text{T}}$ are the discrete displacements for both degrees of freedom of node i in element e. With this in mind, it should be noted that equation (3.7) is solved for these discrete displacements d, i.e. one could also write $Kd = f$ instead. The shape functions and the discrete displacements are shown in Fig. A.1.
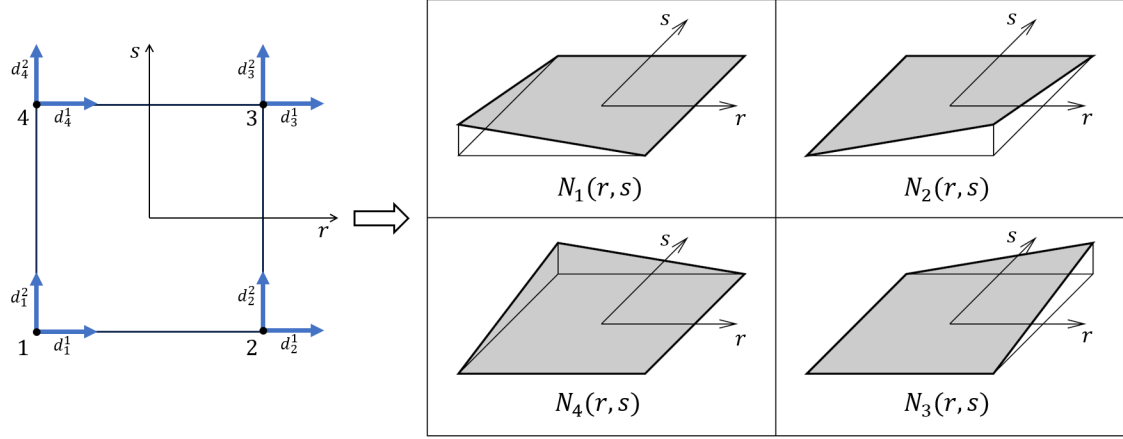
Figure A.1.: Left: definition of the local coordinates r and s and the discrete displacements d_i. Right: depiction of the shape functions (A.6) from [66].

Van Oers uses the Bubnov-Galerkin method meaning that he chooses the same kind of functions for the test functions as well: $w_h^{(e)}(r,s) = \sum_{i=1}^{4} N_i(r,s)g_i$, where the $g_i \in \mathbb{R}^2$ are discrete weights.

Using a matrix-vector notation for the shape functions and the discrete displacements, i.e. a 2×8-matrix for N and an 8×1-vector for d,

$$N = N(r,s) := \begin{pmatrix} N_1(r,s) & 0 & N_2(r,s) & \cdots & 0 \\ 0 & N_1(r,s) & 0 & \cdots & N_4(r,s) \end{pmatrix}, \qquad d^{(e)} := \begin{pmatrix} d_1^1 \\ d_1^2 \\ \vdots \\ d_4^2 \end{pmatrix},$$

one can write $u_h^{(e)} = Nd^{(e)}$. Additionally, let $B := LN$, where $L \in \mathbb{R}^{3\times 2}$ is the transposed divergence operator, then $Lu_h^{(e)} = Bd^{(e)}$.

Before the discretization is combined with the weak form of the problem, the Principal of Virtual Work needs to be introduced. From now on, it is written $w_h = \delta u_h$, and analogously $w = \delta u$, where the δ symbolizes the variation of u_h and u – the operation that allows this conversion. Without going too much into detail, W. A. Wall's instructions are followed on how to insert the virtual displacement $w = \delta u$ into the weak form (A.5) and then the discretization is resumed. The weak form bears the term $\langle \sigma(u), Lw \rangle$, which can be rewritten as $\langle \sigma, L\delta u \rangle$. The symmetry of the stress tensor σ and the Kinematic Equation allow to rearrange this to $\langle \sigma, \delta \varepsilon \rangle$, which in the last step (see [66] for details; the definition of ε and the symmetry of C are used) can be brought back to only using the displacements u as unknowns and yields the form

$$\int_\Omega \langle CLu, L\delta u\rangle \,\mathrm{d}\Omega - \int_\Omega \langle \hat{b}, \delta u\rangle \,\mathrm{d}\Omega = \int_\Omega \delta u^\mathrm{T} L^\mathrm{T} CLu \,\mathrm{d}\Omega - \int_\Omega \delta u^\mathrm{T} \hat{b} \,\mathrm{d}\Omega = 0. \qquad \text{(A.7)}$$

If the discretization of u is inserted in the next step, $\delta u_h^{(e)} = N\delta d^{(e)}$ are obtained as virtual nodal displacements.

With the approximating expressions for the displacements, and the Principle of Virtual Work, the discrete form of (A.7) is given as a sum over all elements

$$\sum_{e=1}^{\text{no. of elements}} \left(\int_{\Omega^{(e)}} \delta d^{(e)\mathrm{T}} B^\mathrm{T} CB d^{(e)} \,\mathrm{d}\Omega - \int_{\Omega^{(e)}} \delta d^{(e)\mathrm{T}} N^\mathrm{T} \hat{b} \,\mathrm{d}\Omega \right) = 0.$$

The only quantity in this equation that depends on spatial coordinates is N, and therefore also B. One can remove $d^{(e)}$ from the integrals and obtain

$$\sum_{e=1}^{\text{no. of elements}} \left(\delta d^{(e)\mathrm{T}} \left(\int_{\Omega^{(e)}} B^\mathrm{T} CB \,\mathrm{d}\Omega \, d^{(e)} - \int_{\Omega^{(e)}} N^\mathrm{T} \hat{b} \,\mathrm{d}\Omega \right) \right) = 0.$$

The integral $\int_{\Omega^{(e)}} B^\mathrm{T} CB \,\mathrm{d}\Omega$ is called element stiffness matrix $k^{(e)}$, and $\int_{\Omega^{(e)}} N^\mathrm{T} \hat{b} \,\mathrm{d}\Omega$ is the element load vector $f^{(e)}$. With van Oers' choice of shape functions, these integrals can be evaluated using Gaussian quadrature (he uses two integration points per axis, i.e. in local geometry $(r, s) = \left(\pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}} \right)$). Via a process called assembly, the local element nodes and their degrees of freedom are assigned to the global nodes, and their respective values in $k^{(e)}$ and $f^{(e)}$ are added up, leading to the global stiffness matrix K and the global force vector f. This of course leads to the system

$$\delta d^\mathrm{T} (Kd - f) = 0,$$

but since the virtual nodal displacements δd is a vector of non-vanishing, arbitrary values, this equation is equivalent to the finite elements equation (3.7).

A.5. Overview of the Authorship in the CPM-FEM Model

The CPM-FEM model is based on the hybrid angiogenesis model by René F. M. van Oers et al. [50]. They developed the model in C [105] and published most of its components as well as a documentation of their code. For this thesis, the public parts of their code were translated to Python [106] and missing modules and functions were added. Some of the translations were performed with the help of ChatGPT [107]. Also, new parts concerning the ECM degradation and the cell proliferation were written. The resulting

code is publicly available at `https://github.com/veronikahofmann/masters-thesis-cellinvasion`.

To give an overview of the extended models' authorship, Tab. A.1 was created.

Module	Authorship
analysis.py	Hofmann
celldivision.py	Hofmann
cellforces.py	van Oers et al.
cellmoves.py	van Oers et al.
colson.py	Hofmann
CPM_dH.py	van Oers et al.
crossley.py	Hofmann
ECM_degradation.py	Hofmann
FE_assembly2.py	van Oers et al., `reduce_K` by Hofmann following the idea of van Oers et al.
FE_local.py	van Oers et al.
FE_nodes2dofs.py	van Oers et al.
FE_solver.py	van Oers et al.
init_conditions.py	`set_restrictions` and `set_forces` by van Oers et al., `init_cells`, `init_spheroid` and `init_ECM` by Hofmann
main.py	van Oers et al., some extensions by Hofmann
parameters.py	van Oers et al., some extensions by Hofmann
plots.py	`force_movie` and `strain_movie` by van Oers et al., `plot_spread_timeseries` by Hofmann
read.py	van Oers et al.
structures.py	van Oers et al.
write.py	van Oers et al., except for `write_increment`, `write_concentrations` and `write_total_cellforces`

Table A.1.: Authorship of the code components.

A.6. Justifications for the Parameter Choices of the Fixed Parameters

In the following, the choices for the fixed parameters in all three models that are used in the sensitivity analysis as well as for the data fitting are explained.

A.6.1. Fixed Parameters of the CPM-FEM Model

The **size of the domain** is one of the main contributors to the computational cost and the resulting runtime. It seems reasonable to choose it as small as possible, yet

large enough to capture the model's dynamics. Crossley and Colson both set their one-dimensional domains to a length of 200 "lattice spacing units", but the travelling waves are visible for much lower values than this. To obtain numerically comparable results in terms of the cell spread, the domain size for the CPM-FEM model is chosen as 60 by 60 elements, leading to a one-dimensional representation of length 30. Accordingly, in the implementations of Crossley's and Colson's models a length of 30 units is set.

Van Oers et al. choose the quadratic elements to have a **side length** of 2.5 µm. Considering that human cells have diameters in the magnitude of 7.5 to 150 µm [108], this seems to be a reasonable scale, allowing each cell to spread over multiple elements.

The **Poisson's ratio** ν is a material constant describing how much a material deforms under loading. When a sample of some material – for instance a collagen fiber, to remain in the biological setting – is pulled at both ends, it not only elongates by a factor $\varepsilon_{\text{longitudinal}}$, but it also becomes thinner by a factor $\varepsilon_{\text{lateral}}$ (note that in this case $\varepsilon_{\text{lateral}} < 0$, as it usually describes a length augmentation). Poisson's ratio is then defined as their negative fraction $\nu = -\varepsilon_{\text{lateral}}/\varepsilon_{\text{longitudinal}}$. As it represents the simplest case, in the finite element model the ECM is assumed to be an isotropic material, a class of materials whose mechanic properties remain the same after rotation. They have Poisson values between 0 and 0.5, where 0 signifies maximal volume change, and 0.5 characterizes an almost incompressible material. Van Oers and colleagues choose $\nu = 0.45$ – while similar values of $\nu = 0.45$ up to $\nu = 0.5$ are used in literature [109, 110], they are all slightly inaccurate as they are deducted from the imagination of the ECM as a linearly elastic and (nearly) incompressible material. The reality is far more complex, not only because these assumptions are simplifying: two samples of the same tissue can expose different Poisson's ratios and they might even change over time, making it very difficult to define an appropriate value for ν. Experiments with breast cancer tumors in mice yielded relatively constant $\nu \approx 0.3$ in the tumor environment over a course of three weeks [71], which is the value chosen for ν in this work.

Similar to Poisson's ratio, the **Young's modulus** E is another material constant. It measures stiffness, meaning how easily a material can be deformed, as long as this deformation is small enough for the stress-strain-ratio within the material to be assumed to be linear. It is defined as $E = \sigma/\varepsilon$, where σ is the stress (meaning force per unit area) and ε is the strain, i.e. the deformation resulting from the applied force. Its unit is $1\text{Pa} = 1\text{N/m}^2$. High values for E therefore represent materials which are hard to deform, while low values are characteristic for easily deformed materials. Different than in the case of Poisson's ratio, Young's modulus can be chosen freely without the need to alter modeling assumptions. Van Oers et al. found that Young's modulus is a key parameter when it comes to cell migration. Most tissues' Young's moduli can be found in the range

of 0.5-30 kPa (see [111] for a summary of publications on different tissues), which is the range van Oers et al. are using. While at the lower and upper ends of this spectrum the cells hardly move, they spread over the whole domain for Young's moduli of 10-14 kPa. Values of around $E = 10$ kPa were also observed in the tumor-surrounding tissue of the before-mentioned mice-experiment [71], hence the choice of $E = 10$ kPa seems suitable for the observation of cancer cells' migratory behavior.

The **cell diameter** d refers to the target area $A(\sigma) = \pi(d/2)^2$ of the cells, i.e. the size at which no energy is dedicated to maintaining it. The sizes of cancer cells vary, alone in the case of prostate cancer cells their diameters have been observed to range from 8 µm in the case of circulating cells up to 21 µm for cells attached to the primary tumor. It appears to be a common phenomenon that the cells forming the tumor are larger than the ones that are moving in the blood stream, even though the latter expose a widespread size distribution, complicating a size comparison [112]. The Micro and Nano Integrated Biosystem (MINIBio) Laboratory of the Pennsylvania State University, who investigate cell flow dynamics and are therefore very interested in the sizes of their cells, have performed a lab inventory in which they found that the average diameter of the cancer cell lines (i.e. tumor-forming cells, in contrast to circulating cells) that are mostly used for their research are between 15–25 µm [72]. Relying on their expertise and keeping in mind that the models discussed in this work concern cancer cells that are initialized in a tumor spheroid, the cell diameter is chosen to be 20 µm. Note that this is the same value as van Oers uses.

The **radius of the initial cell spheroid** should be equal to the length of the initially invaded region of Crossley's and Colson's models. Assuming the one-dimensional domain to have a length of 30 units, a range of 5 units initially occupied by cells has been found suitable to observe the relevant model dynamics given certain intervals of parameters. With the current choices of element size and target cell diameter, the imposition of this 5-elements-radius leads to two cells being placed on the domain at $t = 0$.

The **cell elasticity** λ is a dimensionless parameter describing the impact of the volume constraint in the CPM: the larger λ, the higher the system energy savings or expenses regarding changes in cell size. When comparing to other CPMs in literature, λ is not derived from mechanic properties of the cells, but it is rather estimated in a way such that it fits the other parameter values. In the original CPM, it is $\lambda = 1$, keeping the influence of the volume constraint low [57], in Stott et al.s adaption it is a function of space and nutrient concentration [48], in Turner et al. as well as in Scianna et al.'s publications, λ is a constant chosen such that volume changes are influential, but not the governing factor in the CPM [62, 100, 113]. Van Oers and colleagues set $\lambda = 500$ to compensate that their model uses the target volume as a relative quantity, in contrast to the original, absolute

formulation of the CPM's volume constraint. This parameter value is maintained in this work.[3]

Regarding the **cell-to-cell and cell-to-ECM attachment energy** $J_{CC} := J(\sigma(x), \sigma(x'))$ with $\sigma(x)$, $\sigma(x') > 0$ and $J_{CE} := J(\sigma(x), 0)$ with $\sigma(x) > 0$, respectively, the following considerations are made: for the cells to spread across the domain, they need to be given a motivation to leave the primary tumor site. The simplest way to do this in terms of the CPM is by setting the energy necessary for two cells to hold on to each other higher than the energy of a cell to hold on to the substrate, i.e. setting $J_{CC} > J_{CE}$. In the literature on the CPM that has been discussed so far, the values for both parameters range from 0 to 12 [57, 100, 62, 113]. Here, the same values as in van Oers et al. is used, namely $J_{CC} = 2.5$ and $J_{CE} = 1.25$.

The **intrinsic cell motility** T "represents the extent to which the active cell motility can overcome the reactive forces (e.g. volume constraint or adhesions) in the environment" [50], making it very influential on other parameters such as λ or J_{CC} and J_{CE}. Van Oers et al.'s example is followed to avoid interference, and $T = 1$ is set.

The **cell force constant** μ is the scaling parameter of the traction forces the cells exert on the nodes. The total force generated by a cell is given by around 300 nN, as van Oers cites the results of an experiment on the interplay of the cytoskeletal structure and force production [73]. Hence, μ has to be chosen in a way that on average, each cell generates a total traction force of 300 nN, at least in the long term behavior. It should suffice to require an approximate equality, since cell deformations – which happen inevitably in the model – that deviate from the optimal, energy-saving sphere shape will always lead to larger forces generated by the cells. This can be seen directly in the formula (3.8), where the traction forces are larger with larger distances between the nodes. Over time, when the cells were able to move away from the initial spheroid and their shape is less dictated by the available space, their geometry comes closer to the traction force minimizing circle. However, there is one problem: the cell size, which determines how many nodes are manipulated by the cell and therefore how many summands appear in (3.8), is not fixed and highly depends on the substrate stiffness, which is one of van Oers et al.s main findings [50]. Even though the optimal cell size is given by a parameter, in stiff (32 kPa) environments the cells' area can deviate from it by up to 125 µm^2. Hence, the appropriate μ can only be estimated for fixed Young's moduli E. The above-chosen value of $E = 10$ kPa is used and values for μ are tested. The total traction force for a cell c is calculated with the following formula:

---

[3]See the supplementary material of van Oers et al. for the consequences of changing λ on cell shape.

$$F_{\text{total}}^c = \sum_{\substack{\text{nodes } i \text{ of elements} \\ \text{covered by } c}} \sqrt{(F_i^x)^2 + (F_i^y)^2},$$

where F_i^x is the horizontal component of the μ-dependent force vector F_i, and F_i^y is the vertical one. To obtain most accurate results, various values for μ are tested with only one cell present, without proliferation or surrounding ECM blocking the cell's movement. Van Oers et al. use $\mu = 0.01$ nN/µm, which is the starting value for this test. The mark of 300 nN should not be undercut, but approached closely over time. After 100 Monte Carlo steps, the evolution of the total cell force is visually assessed, see Fig. A.2. μ-values between 0.0095 and 0.0105 yield results closest to what is expected, leading to the choice of $\mu = 0.01$ nN/µm just as van Oers proposes.
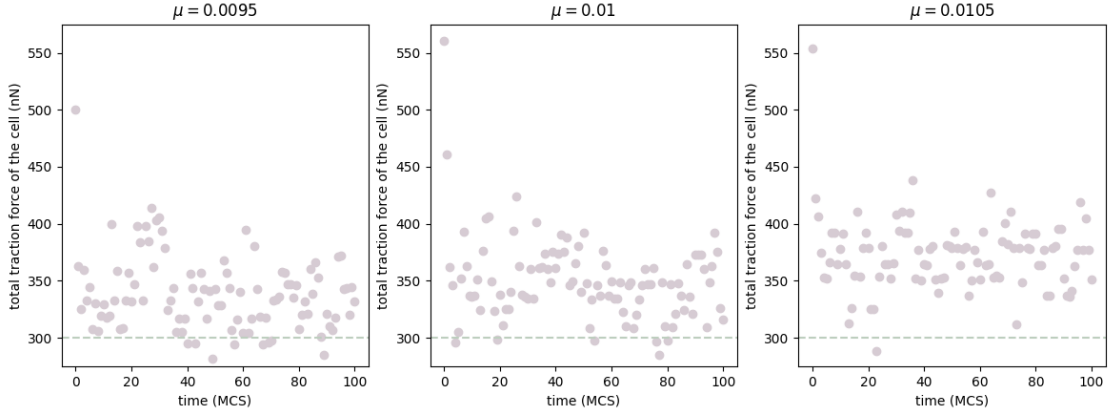

Figure A.2.: Total traction force generated by a single cell under various values for μ. Time in Monte Carlo steps (MCS).

The only parameters left are the parameters of the two functions involved in the durotaxis-Hamiltonian. The **strain-stiffening function** E has two parameters, the material stiffness of the undeformed substrate E_0, and the strain-stiffening speed ε_{st}). It is $E_0 = E$, so the same choice is made as for Young's modulus, $E_0 = 10$. The stiffening parameter ε_{st} is chosen as 0.1 according to van Oers proposal.

The **durotaxis function** h has three parameters, namely its maximum value α, the steepness β, and the shift to the right E_{tr}. h links durotactic attraction to the stiffness of the stretched material. Similarly to the cell elasticity λ, the parameter α (and the other parameters of h to a lesser extent as well) weighs the influence of the durotactic effect on the total system energy. While $\alpha = 10$ should best be chosen as in the template by van Oers not to disrupt the balance of the effect interplay, especially E_{tr} should be examined more closely and adapted to the durotactic properties of cancer cells. Cancer cell lines

responsible for brain, breast and skin cancer have been observed to perform the highest increases in durotactic efficiency on 2-7 kPa stiffness gradients, reaching a plateau for gradients beyond [30]. E_{tr} marks the value of substrate stiffness at which the durotaxis function has its steepest increase. With this information alone it seems reasonable to set $E_{\text{tr}} = 4.5$, assuming the mean of the 2-7 kPa interval is the point of steepest increase. However, the stiffness of the undeformed substrate has been set to $E = 10$ kPa, i.e. $E_{\text{tr}} < E$, which is possible as it just means that the cells are already moving on a surface where durotaxis is a prominent phenomenon and has similar impacts on the cells in any direction, regardless of how much their migration stretches the material. For instance, let a cell that occupies a site x consider an expansion to the neighboring site x'. For simplicity, let the orientation of the maximum stretch at site x' coincide exactly with the direction of the expansion, i.e. $\langle v_1, v_d \rangle = 1$, and let the orientation of the minimum stretch be perpendicular to it, i.e. $\langle v_2, v_d \rangle = 0$. Further assume that the maximum stretch is relatively strong: $\varepsilon_1 = 0.5$, leading to $E(\varepsilon_1) = 60$. The Hamiltonian for the durotaxis in this case is constituted by $\Delta H_{\text{durotaxis}} = -1 \cdot h(60)$. With the choice $E_{\text{tr}} = 4.5$ and the steepness $\beta = 0.5$ as in van Oers' example, $\Delta H_{\text{durotaxis}} \approx -9.99$ is obtained. Comparing this with a scenario in which the stretch is much smaller, i.e. $\varepsilon_1 = 0.01$, it is $E(\varepsilon_1) = 11$ and the Hamiltonian becomes $\Delta H_{\text{durotaxis}} \approx -9.63$. The difference between the two $\Delta H_{\text{durotaxis}}$ is marginal, even though the tested stretches and their accompanying substrate deformations mark virtually the whole range of possible values. It is almost equally attractive for the cell to expand in the direction of high strain as in the direction of hardly any strain. This can be prevented by choosing $E_{\text{tr}} > E$, even though it might not be in perfect accordance with the findings of [30]. $E_{\text{tr}} = E + 4.5 = 14.5$ is chosen, considering E as the "baseline stiffness", which has to be exceeded for significant durotaxis to happen. In the example above, the difference between the Hamiltonians is now $\Delta H_{\text{durotaxis}} \approx -9.99$ in the case of the large strain versus $\Delta H_{\text{durotaxis}} \approx -1.48$ for the small strain.

A summary of the fixed parameters of the CPM-FEM model is listed in Tab. 4.1.

A.6.2. Fixed Parameters of the PDE Models

Including the initial conditions, the parameterized model (3.1)-(3.2) by Crossley et al. contains six parameters, the cell diffusivity D, the carrying capacity K, the proliferation probability r, the degradation rate of ECM λ, the initial ECM concentration $\hat{m}_0$, and the length of the initially invaded region α.

Colson's model (3.4) contains two more, the blocking ECM density $m_{\max}$, and the initial invasion length is split into two parameters σ and ω. While three of these parameters are varied during the sensitivity analysis, most of them are fixed.

In accordance with the domain size of the CPM-FEM model, the spatial PDE model **domain length** L is restricted to $[0, 30]$.

Even though there are estimates for the **cell diffusivity** D in literature (e.g. Browning et al. found $D \approx 0.17$ µm²/s [51]), here it is set to 0.5 since otherwise the number of timesteps that needs to be computed to observe travelling waves requires very large runtimes for the CPM-FEM model.

The **carrying capacity** K is set to 1, since all models consider densities $u, m \in [0, 1]$.

With a spatial domain length of only 30 length units, the **length of the initially invaded region** α must not be chosen too large as otherwise interesting developments could quickly wander out of the domain. On the other hand, setting α too small leads to a very narrow interval in which cells and ECM elements initially interact. In that case, to be able to observe the influence of the other parameters, one needs to solve the PDEs for larger times t, which is not an issue for the models by Crossley and Colson, but requires costly additional runtime for the CPM-FEM model. Also, experiments with the model show that too small α-values lead to unreliable sensitivities or sensitivities that are difficult to interpret. For a relatively stable sensitivity analysis, $\alpha = 5$ is found suitable.

The **unconquerable ECM density** $m_{\max}$ is set to 1 to make the two models more comparable (Crossley's model implicitly contains $m_{\max} = 1$: if one sets $\hat{m}_0 = 1$, the system does not evolve and remains at the initial configuration [54]).

The last parameters are Colson's **initial invasion parameters** σ and ω. From $x = 0$ to $x = \sigma - \omega$, the cell density is constant at $u(x, 0) = 1$. From $x = \sigma - \omega$ to $x = \sigma$, it transitions to $u(x, 0) = 0$. Making a compromise between model comparability and model distinctiveness, $\sigma = 5$ and $\omega = 1$ are chosen.

A summary of the parameter choices for the PDE models can be found in Tab. 4.2.

A.7. Total-Order Sensitivity Indices

Fig. A.3 and A.4 show the total-order sensitivity indices for the three investigated time steps $t = 10, 20, 30$. In Crossley's model, for some instances the indices exceed 1, which again is an accuracy issue, even though these estimations were performed with $n = 2048$. This inaccuracy is accepted: after all, the total-order indices are only used to obtain an indication on whether there could be relevant interaction effects between the parameters which make the calculation of the second-order indices necessary.
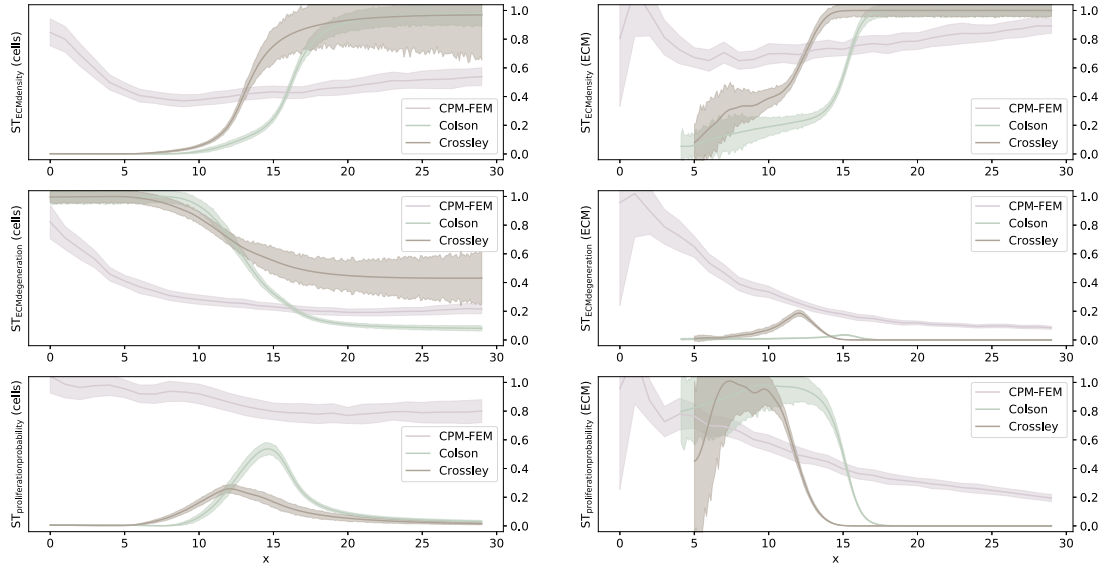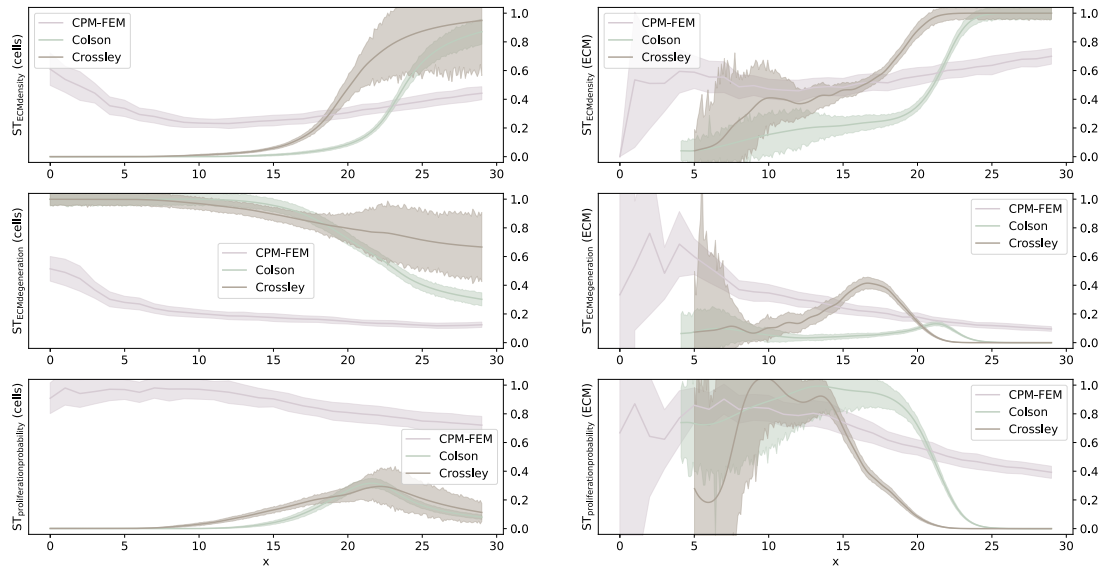
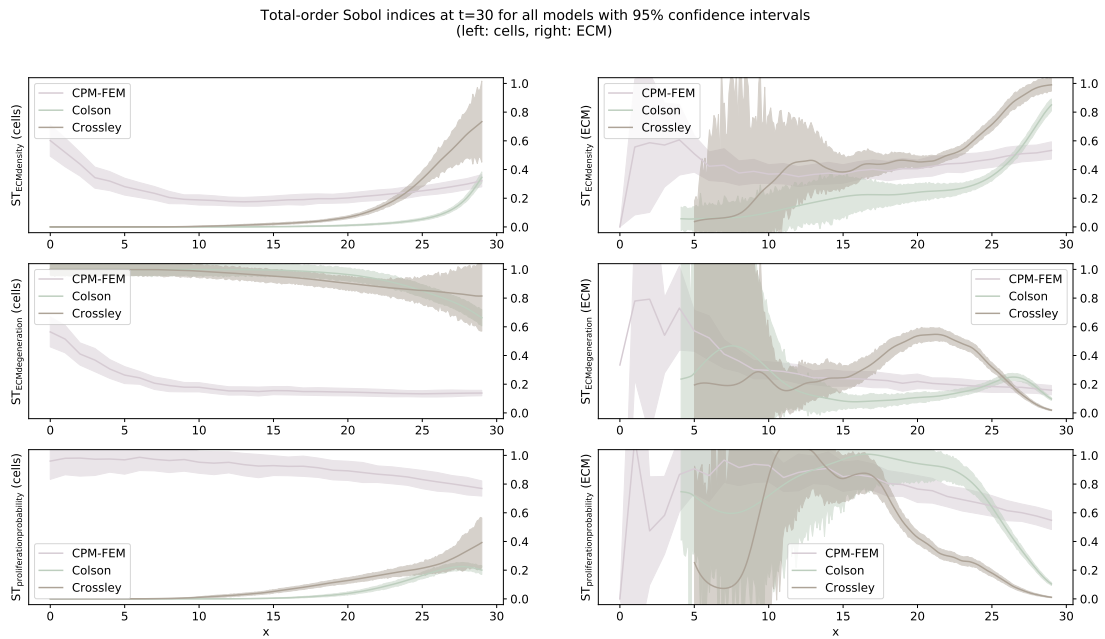Figure A.3.: Total-order sensitivity indices for the time steps $t = 10$ and $t = 20$.

Figure A.4.: Total-order sensitivity indices for $t = 30$.

A.8. Found Optimal Values of the Comprehensive Parameter Fits

Tab. A.2 shows the found parameter values, as well as the SSR and AIC values of all experiments from section 5.2. For plots of the fits without durotaxis and without attachment cost see appendix A.9.

Parameter	Initial guess	Found optimum	SSR	AIC
Full Hamiltonian				
Crossley's degradation rate λ	0.2	3.3769		
Crossley's proliferation r	0.005	0.1351		
Crossley's cell diffusivity D	0.5	0.5620	0.7819	-248
Crossley's capacity K	1	1.0211		
Crossley's initial invasion α	5	5.0000		
Colson's degradation rate λ	0.2	0.1846		
Colson's proliferation r	0.005	0.1037		
Colson's cell diffusivity D	0.5	0.4449		
Colson's unconquerable density m_{max}	1	1.8441	0.0203	-463
Colson's capacity K	1	1.0532		
Colson's initial invasion σ	5	1.5137		
Colson's initial invasion ω	1	0.7778		
No Durotaxis				
Crossley's degradation rate λ	0.2	2.5535		
Crossley's proliferation r	0.005	0.1206		
Crossley's cell diffusivity D	0.5	0.5448	0.7711	-249
Crossley's capacity K	1	1.0327		
Crossley's initial invasion α	5	5.0000		
Colson's degradation rate λ	0.2	0.1652		
Colson's proliferation r	0.005	0.0878		
Colson's cell diffusivity D	0.5	0.4968		
Colson's unconquerable density m_{max}	1	1.6455	0.0277	-444
Colson's capacity K	1	1.0937		
Colson's initial invasion σ	5	1.5561		
Colson's initial invasion ω	1	0.5329		
No Attachment Cost				
Crossley's degradation rate λ	0.2	4.4819		
Crossley's proliferation r	0.005	0.1647		
Crossley's cell diffusivity D	0.5	0.5852	0.7457	-251
Crossley's capacity K	1	0.9762		
Crossley's initial invasion α	5	5.0000		
Colson's degradation rate λ	0.2	0.2238		
Colson's proliferation r	0.005	0.1304		
Colson's cell diffusivity D	0.5	0.4874		
Colson's unconquerable density m_{max}	1	1.5910	0.0278	-444
Colson's capacity K	1	1.0020		
Colson's initial invasion σ	5	0.3094		
Colson's initial invasion ω	1	0.1535		

Table A.2.: Optimal parameter values for the PDE models.

A.9. Fit Quality and Plots from the Comprehensive Fit: No Durotaxis and No Attachment Cost

In the following, the SSR and AIC values as well as the plots of the fitted trajectories and the wave speed comparisons can be found for the CPM-FEM model without durotaxis and without attachment cost, respectively. The found optima for the estimated parameters are listed in appendix A.8.

A.9.1. Parameter Fitting without Durotaxis

See Tab. A.3 for the AIC and SSR of the models, Fig. A.5 and Fig. A.6 for an illustration of the fit and the wave speed comparison.

Model	SSR	AIC
Crossley	0.7711	-249
Colson	0.0277	-444

Table A.3.: SSR and AIC values of the fitted PDE models, without durotaxis.
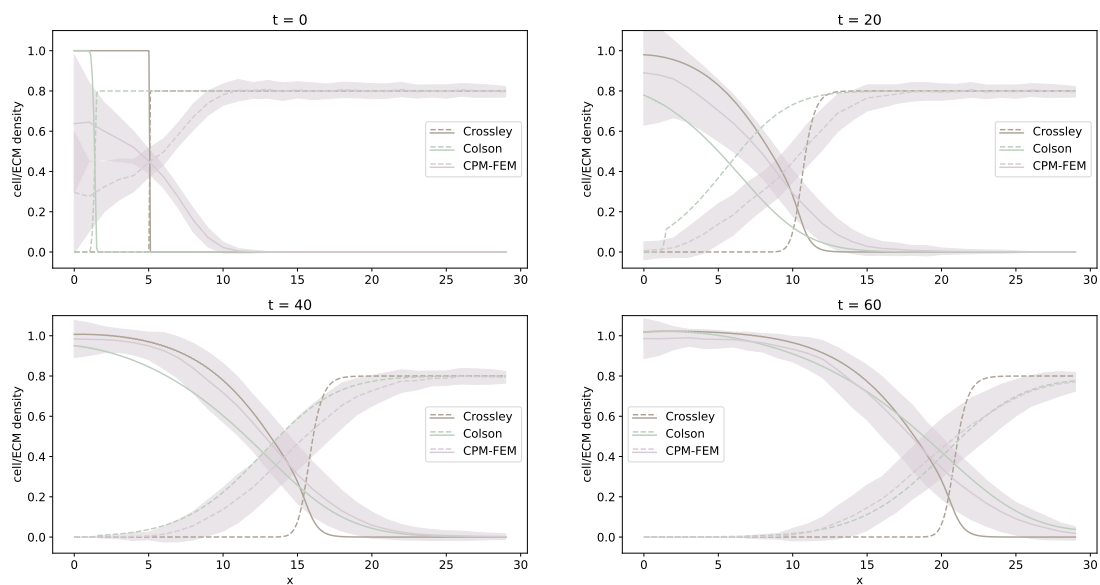
Figure A.5.: All-parameter fitted PDE models and the data generated by the average over 100 CPM-FEM model evaluations ignoring durotaxis. For more information, see Fig. 5.10.
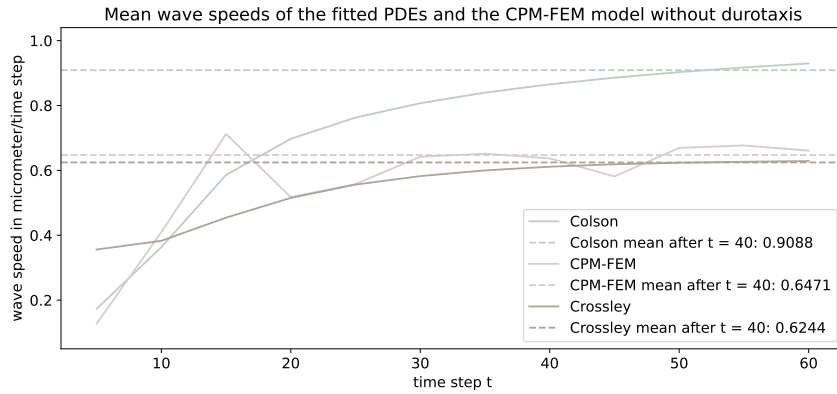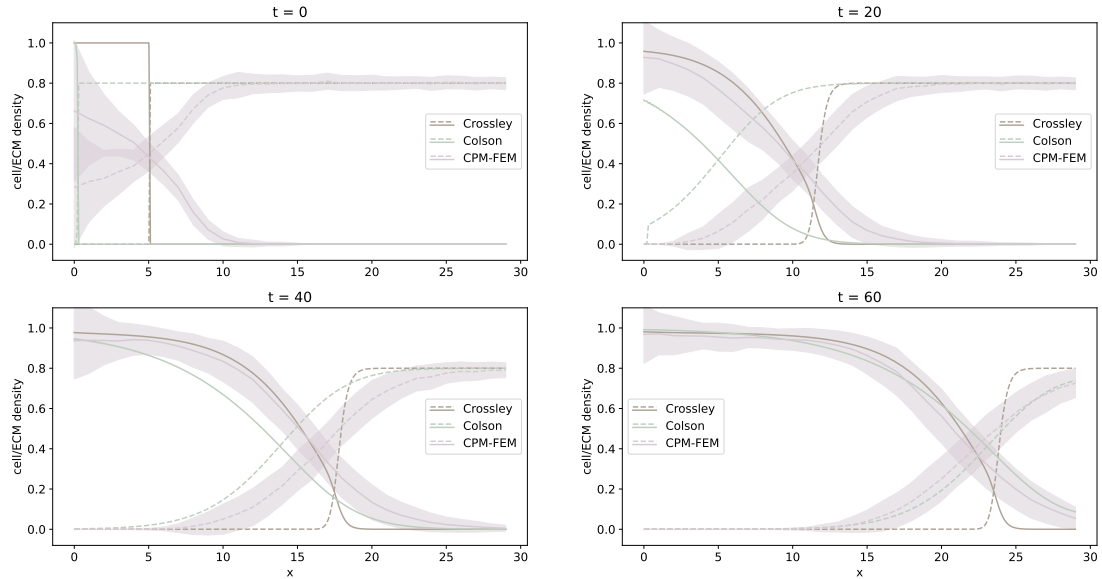
Figure A.6.: Wave speed estimates of the all-parameter fitted PDE models and the average of 100 CPM-FEM model evaluations. The dashed lines are the respective mean values of these wave speeds, where the mean is calculated for the time steps after $t = 40$.

A.9.2. Parameter Fitting without Attachment Cost

See Tab. A.3 for the AIC and SSR of the models, Fig. A.7 and Fig. A.8 for an illustration of the fit and the wave speed comparison.

Model	SSR	AIC
Crossley	0.7457	-251
Colson	0.0278	-444

Table A.4.: SSR and AIC values of the fitted PDE models, without contact cost.

Figure A.7.: All-parameter fitted PDE models and the data generated by the average over 100 CPM-FEM model evaluations ignoring durotaxis. For more information, see Fig. 5.10.
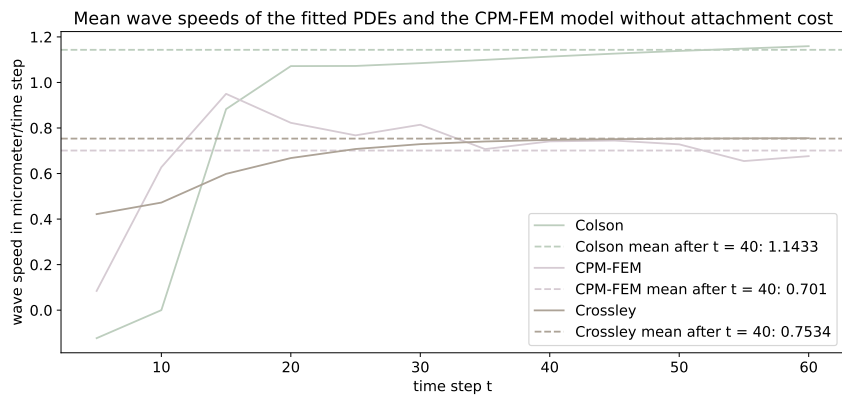

Figure A.8.: Wave speed estimates of the all-parameter fitted PDE models and the average of 100 CPM-FEM model evaluations. The dashed lines are the respective mean values of these wave speeds, where the mean is calculated for the time steps after $t = 40$.

A.10. Comprehensive Fit with Constrained Optimization for Both Models

Using the data generated by the full CPM-FEM model, all parameters of Crossley's model were fitted using a Trust Region Reflective algorithm (which is the same algorithm that is used for the comprehensive fit of Colson's model), to see whether the choice of optimization algorithm impacts the quality of the fit. The parameter bounds are chosen as $\lambda \in [0, 100]$, $r \in [0, 1]$, $D \in [0, 5]$, $K \in [0, 5]$, and $\alpha \in [0, 30]$. A comparison of the found optimal parameter values can be found in Tab. A.5, and Fig. A.9 and A.10 show the fit and the wave speed using the new parameters.

Found Optima		
Parameter	Result LM	Result TRR
Degradation rate λ	3.3769	2.0956
Proliferation r	0.1351	0.1268
Cell diffusivity D	0.5620	0.6846
Capacity K	1.0211	1.0421
Initial invasion α	5.0000	5.0000
Fit Quality Indicators		
SSR	0.7819	0.7128
AIC	-248	-253

Table A.5.: Optimization results for Crossley's model compared by minimization algorithm: Levenberg-Marquardt (LM) versus Trust Region Reflective algorithm (TRR).
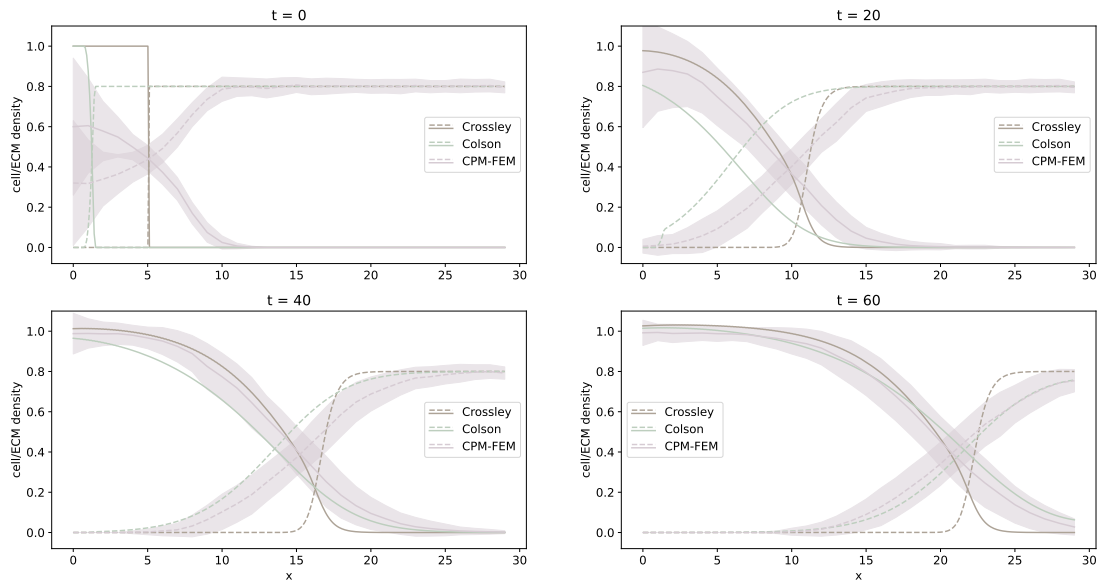
Figure A.9.: All-parameter fitted PDE models and the data generated by the average over 100 CPM-FEM model evaluations using the full model. For Crossley's curve, the parameter values determined by the Trust Region Reflective algorithm were used. For more information, see Fig. 5.10.
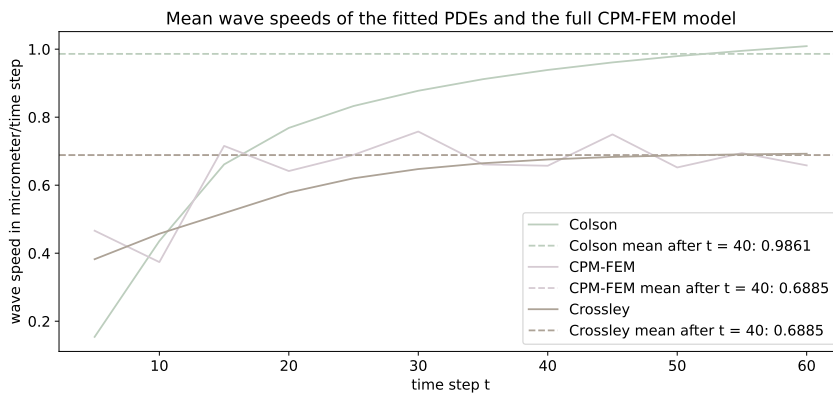

Figure A.10.: Wave speed estimates of the all-parameter fitted PDE models and the average of 100 CPM-FEM model evaluations. For Crossley, the parameter values determined by the Trust Region Reflective algorithm were used. The dashed lines are the respective mean values of these wave speeds, where the mean is calculated for the time steps after $t = 40$.