



Diskret matematisk modell av humant glioblastom och dess invasionsmönster i grå och vit hjärnsubstans

A Discrete Mathematical Model of Human Glioblastoma and Its Invasion Pattern in Gray and White Matter

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Gustav Fixell

Arwa Haamid

Timothy Jarebrant

Vanessa Kehdi

Jonathan Lång

Felix Rylander

Diskret matematisk modell av humant glioblastom och dess invasionsmönster i grå och vit hjärnsubstans

Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers
Vanessa Kehdi Jonathan Lång

Examensarbete för kandidatexamen i matematik inom Matematikprogrammet vid Göteborgs universitet
Arwa Haamid

Kandidatarbete i matematik inom civilingenjörsprogrammet Automation och mekatronik vid Chalmers
Timothy Jarebrant

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk fysik vid Chalmers
Gustav Fixell Felix Rylander

Handledare: Philip Gerlee
Adam Malik

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2025

Förord

Vi vill rikta ett stort tack till våra handledare Adam Malik och Philip Gerlee för deras värdefulla stöd under arbetets gång. Deras expertis och insikter har varit viktiga för hela projektets utveckling.

Ett särskilt tack vill vi rikta till Adam, som genom sitt engagemang i form av regelbundna möten och tillgänglighet via mail har varit ett ovärderligt stöd för gruppen. Hans konstruktiva återkoppling och förmåga att lotsa oss framåt har haft en avgörande betydelse för både vårt arbete och vår arbetsprocess.

Under projektets gång har individuella prestationer dokumenterats i en dagbok och tidsloggats. Veckovis har den gemensamma dagboken sammanställts av en gruppmedlem, ett ansvar som roterat enligt uppsatt schema. Vi är som grupp överens om att samtliga medlemmar bidragit likvärdigt till projektet, vissa har bidragit större delar av modellutveckling än snarare skrivandet. I tabell 1 nedan klargörs den huvudsakliga ansvarsfördelningen för den skriftliga rapportens delar. Men det ska understrykas att samtliga gruppmedlemmar korrekturläst och justerat rapporten i sin helhet, mer än vad som nämns i bidragstabellen.

Tabell 1: Bidragsrapport med huvudsakligt författarskap över kandidatarbets huvudsakliga skriftliga delar.

| Avsnitt | Rubrik | Författare (f), revisor (r) |
|---------|---|---|
| - | Förord | Felix (f) |
| - | Populärvetenskaplig presentation | Felix (fr) & Gustav (fr) |
| - | Sammandrag & Abstract | Gustav (fr) & Arwa (f) |
| 1 | Inledning | Gustav (fr) & Jonathan (fr) & Vanessa (fr) |
| 1.1 | Problemformulering | Gustav (f) & Jonathan (r) & Timothy (r) |
| 1.2 | Syfte | Gustav (f) & Jonathan (fr) & Vanessa (fr) |
| 2 | Metod | Jonathan (fr) & Felix (fr) |
| 2.1 | Modellering | Jonathan (f) & Vanessa (r), Gustav (r) & Arwa (r) |
| 2.1.1 | Modellering av cell-cell interaktioner | Jonathan (fr) & Gustav (r) & Arwa (f) |
| - | LGCA | Jonathan (fr) & Felix (r) |
| - | Celldelning | Jonathan (fr) & Felix (r) & Arwa (r) |
| - | Övergång mellan fenotyper | Jonathan (fr), Arwa (f), Felix (r) & Gustav (r) & Vanessa (r) |
| - | Migration | Arwa (f), Vanessa (r), Felix (r) & Gustav (fr) |
| - | Hjärnskalet | Vanessa (fr), Gustav (r), Felix (r) |
| 2.1.2 | Modellering av cell-omgivning-interaktioner | Jonathan (fr) & Gustav (fr) & Timothy (r) |
| 2.2 | Databehandling | Gustav (f) & Felix (f) |
| 2.2.1 | Filtrering och extraktion av relevant data | Gustav (f) & Arwa (f) |
| 2.2.2 | Val av analysplan och tumörkaraktisering | Gustav (f) & Felix (f) & Jonathan (r) |
| 2.2.3 | Intensitetsgitter, diskretisering och normalisering | Arwa (f) & Vanessa (r) |
| 2.2.4 | Segmentering av vit substans | Arwa (f) & Vanessa (r) |
| 2.3 | Skapandet av den initiala tumören | Jonathan (fr) & Gustav (r) |
| 2.3.1 | Uppskattning av celldensitet | Jonathan (fr) & Gustav (r) |

| | | |
|-------|---|--|
| 2.3.2 | Förenkling av LGCA-modellen | Jonathan (fr) & Gustav (r) |
| 2.3.3 | Initial fördelning av fenotyper | Jonathan (fr) & Gustav (r) |
| 2.4 | Simulering och parameteroptimering | Timothy (f) & Gustav (f) & Vanessa (fr) |
| 2.4.1 | Jaccard-index | Vanessa (fr) |
| 2.4.2 | Approximativt bayesiansk beräkning | Gustav (f) & Timothy (f) & Vanessa (r) |
| 2.4.3 | Estimering av <i>a priori</i> -fördelningar | Felix (r) & Timothy (fr), Jonathan (fr) |
| 3 | Resultat | Felix (f) & Jonathan (r) |
| 3.1 | Patient 42 | Felix (f) & Jonathan (fr) & Arwa (fr) |
| 3.2 | Patient 85 | Felix (f) & Jonathan (fr) & Arwa (fr) |
| 3.3 | ABC-SMC Populationsstatistik | Timothy (f) & Felix (f) |
| 4 | Diskussion | Felix (f) |
| 4.1 | Jämförelse av modellerna | Jonathan (fr) & Timothy (f) & Vanessa (r) |
| 4.2 | Begränsningar | |
| 4.2.1 | Biologiska begränsningar | Vanessa (f), Gustav (r) & Jonathan (fr) |
| 4.2.2 | Begränsningar med datamängden | Vanessa (fr) & Gustav (f) & Arwa(f) & Felix (fr) |
| 4.3 | Förbättring | Felix (f) |
| 4.3.1 | Förbättringsområden för modellen | Jonathan (fr) & Arwa (f) |
| 4.3.2 | Förbättringsområden för databehandling | Arwa (f) & Jonathan (fr) |
| 4.3.3 | Förbättringsområden för modellanpassningen | Timothy (fr) & Jonathan (f) |
| 4.4 | Samhälleliga och etiska aspekter | Vanessa (f), Felix (r) |
| 5 | Slutsats | Jonathan (fr) |
| 6 | Användning av AI-verktyg | Vanessa (f), Arwa (f) & Gustav (r) |

Populärvetenskaplig presentation

Den mänskliga hjärnan är en av de mest invecklade strukturerna vi känner till och är för det mesta en stor fördel för oss. Just denna komplexitet har också en baksida, för när något väl går snett är det sällan enkelt att förstå, än mindre att åtgärda. Hjärnans komplexitet leder nämligen till att en fiende, så som en tumör, kan växa obemärkt fram tills att det är för sent. En av de mest aggressiva hjärntumörerna kallas *glioblastom* och drabbar över 300 000 personer varje år. När tumörer av denna sort väl upptäcks har de ofta redan hunnit sprida sig genom hjärnans vävnad. De är alltså inte längre bara en klump, utan har utvecklat ett helt nätverk av celler som slingrar sig mellan friska strukturer bortom den synliga tumören. Skulle man operera bort själva tumörklumpen finns därför en stor risk att de gömda cellerna redan har börjat föröka sig, vilket gör att tumören växer tillbaka igen. Utöver detta så har varken strålbehandling eller läkemedel tillräcklig effekt för att stoppa sjukdomen till fullo. Allt detta sammantaget gör att denna typ av hjärntumör blir extremt svårbehandlad. Utmaningen ligger dock inte bara i att avlägsna tumören utan också att förstå hur den växer, sprider sig och förändras över tid. I detta arbete undersöks just dessa processer, men inte med skalpeller eller mikroskop, utan med matematik!

Att använda matematik för att beskriva biologiska system har en längre historia än man kan tro. Redan på 1200-talet tecknade Leonardo Fibonacci i sin bok *Liber Abaci* ner en talföljd som han menade beskrev hur kaninpopulationer växer över tid. Denna kända talföljd och de ingående talen kom senare att kallas för Fibonaccital. Senare, på 1700-talet, visade Daniel Bernoulli hur man kunde analysera effekterna av smittkoppor med hjälp av matematik och Thomas Malthus presenterade idén om exponentiell befolkningstillväxt. Den senare kom att influera Darwins evolutionsteori. På 1800-talet vidareutvecklades dessa idéer, bland annat genom Pierre François Verhulsts logistiska modell och Fritz Müllers matematiska argument för naturligt urval. Under 1900-talet började utvecklingen av biomatematik accelerera, exempelvis genom pionjären D'Arcy Thompson, som i sin bok *On Growth and Form* visade hur matematiska lagar styr levande organismers former. Efter 1960-talet kunde man även ta hjälp av datorkraft för matematiska beräkningar vilket gav upphov till mer komplicerade modeller av biologiska system.

Idag har modellerna blivit så avancerade att de kan beskriva mycket komplexa biologiska förlopp, såsom hur en tumör växer, förändras och sprider sig i kroppen. Tillväxten av glioblastom är exempel på ett sådant, men för att skapa en modell behöver man förstås först identifiera vad som karaktäriserar dessa tumörer. Som tur är har ett glioblastom och cellerna det består av flera karaktärsdrag som kan återskapas matematiskt. Det finns exempelvis en välgrundad anledning att tro att cancercellerna i en tumör har två olika "personligheter": dels finns det celler som i huvudsak förökar sig, dels de som sprider sig till nya delar av hjärnan. Dessutom antas individuella celler påverka varandra, till exempel genom att stöta bort varandra om de hamnar för tätt inpå. Till sist har miljön tumörerna befinner sig i påverkan på hur tillväxten ser ut; saker som förekomsten av näringsämnen eller nervtrådar är exempel på detta.

Vår modell bygger på en så kallad slumpvandring som innebär att alla celler tilldelas en startposition, och sedan med viss sannolikhet tillåts ta steg i en bestämd riktning. Stegen sker med jämna tids mellanrum, och innan varje steg rullas en tärning för att avgöra om riktningen ska ändras. Som tidigare nämnts finns också andra faktorer som påverkar cellernas rörelse, och även effekten av dessa kan översättas till matematikens språk genom att tillskrivas en sannolikhet som prövas innan varje steg. Till exempel kan en cell ändra personlighet från invaderande till förökande med viss sannolikhet för varje steg som tas i tiden. Detta sammantaget ger ett kraftfullt sätt att simulera tumörens beteende. Inkluderandet av förekomsten av nervtrådar gör att de simulerade cellernas beteende ännu bättre kan spegla det av deras kusiner i den verkliga världen.

Sammandrag

Detta arbete undersöker huruvida implementering av en mekanism för interaktion mellan celler och deras omgivning kan förbättra en modell för tillväxt av hjärntumörer av typen *glioblastom* (GBM). Modellen som tagits fram är diskret och stokastisk, och är baserad på versionen av *lattice gas cellular automata* som formulerades av Tektonidis m.fl. Tillägget består i en fallenhet för celler att vandra längs vit hjärnsubstans före grå sådan, vilken integrerats genom segmentering av verkliga hjärnor. Modellen bygger på ett antal parametrar knutna till sannolikhet. Dessa optimerades genom tillämpandet av approximativ bayesiansk beräkning (eng. *approximate Bayesian computation*; ABC).

Både den ursprungliga och den vidareutvecklade modellen har jämförts med MR-bilder och segmenteringar av riktiga GBM från två olika patienter. Datan med vilken simuleringarna jämförts kommer från datamängden LUMIERE. Jämförelsen gjordes med hjälp av Jaccard-index, och resultaten tyder på att den utvecklade modellen kan prestera bättre. Dessa resultat bör dock betraktas med försiktighet, eftersom de till följd av begränsningar i datan och de biologiska antaganden som gjorts är förenade med viss osäkerhet. Detta diskuteras i rapporten. Arbetet bidrar till att lägga grunden för vidare utveckling av modeller som tar hänsyn till den påverkan omgivningen har på hur GBM växer.

Abstract

The thesis presents a discrete mathematical model of the growth of *glioblastoma* (GBM), and examines whether the inclusion of cellular interaction with the environment can improve the model. Building on the stochastic *lattice gas automata* model as presented by Tektonidis *et al.*, the study aims to reach improved results by incorporating environment interaction in the form of cellular bias to walk along white matter over gray. The matter was done by segmenting the white matter of real human brains. The model includes several parameters tied to probabilities, which were optimised using *approximate Bayesian computation* (ABC).

To evaluate the performance of the models, MRI data and corresponding segmentations from two patients, taken from the LUMIERE dataset were utilized for comparison. In the comparison, the Jaccard-index was used as a quantifiable measure to compare the overlap between real data and simulated model. The results suggest that the extended model was a better fit to the real data, however due to the limitations of the dataset and the biological assumptions made the result should be interpreted with caution. The study provides a deeper understanding on the importance of tumor growth models and sets a foundation for future model developments of GBM growth.

Innehåll

| | | |
|----------|---|-----------|
| 1 | Inledning | 1 |
| 1.1 | Problemformulering | 1 |
| 1.2 | Syfte | 2 |
| 2 | Metod | 2 |
| 2.1 | Modellering | 2 |
| 2.1.1 | Modellering av cell-cell interaktioner | 3 |
| 2.1.2 | Modellering av cell-omgivning-interaktioner | 5 |
| 2.2 | Databehandling | 6 |
| 2.2.1 | Filtrering och extraktion av relevant data | 7 |
| 2.2.2 | Val av analysplan och tumörkaraktärisering | 7 |
| 2.2.3 | Intensitetsgitter, diskretisering och normering | 8 |
| 2.2.4 | Segmentering av vit substans | 8 |
| 2.3 | Skapandet av den initiala tumören | 9 |
| 2.3.1 | Uppskattning av celldensitet | 9 |
| 2.3.2 | Förenkling av LGCA-modellen | 10 |
| 2.3.3 | Initial fördelning av fenotyper | 10 |
| 2.4 | Simulering och parameteroptimering | 10 |
| 2.4.1 | Jaccard-index | 10 |
| 2.4.2 | Approximativ bayesiansk beräkning | 10 |
| 2.4.3 | Estimering av <i>a priori</i> -fördelningar | 11 |
| 3 | Resultat | 13 |
| 3.1 | Patient 42 | 14 |
| 3.2 | Patient 85 | 15 |
| 3.3 | ABC-SMC populationsstatistik | 15 |
| 4 | Diskussion | 16 |
| 4.1 | Jämförelse av modellerna | 16 |
| 4.2 | Begränsningar | 17 |
| 4.2.1 | Biologiska begränsningar | 17 |
| 4.2.2 | Begränsningar med datamängden | 17 |
| 4.3 | Förbättringsområden | 18 |
| 4.3.1 | Förbättringsområden för modellen | 18 |
| 4.3.2 | Förbättringsområden för databehandlingen | 18 |
| 4.3.3 | Förbättringsområden för modellanpassningen | 18 |
| 4.4 | Samhälleliga och etiska aspekter | 19 |
| 5 | Slutsats | 19 |
| 6 | Användning av AI-verktyg | 19 |
| A | Bilaga 1 - Källkod | 23 |
| B | Bilaga 2 - <i>A posteriori</i>-fördelningar | 23 |
| B.1 | Patient 85 | 23 |
| B.2 | Patient 42 | 25 |

1 Inledning

Över 120 olika typer av tumörer har hittats i den mänskliga hjärnan. Av dessa är *gliomer* den vanligaste sorten. Bland gliomerna är *glioblastom* (GBM) den oftast förekommande och mest aggressiva varianten, som utgör nästan hälften av alla fall av elakartade tumörer som angriper centrala nervsystemet [1]. Medianöverlevnadstiden är 15 månader och mindre än 5% av patienter överlever fem år efter de diagnostiserats [2], [3].

I sin naturliga miljö, *in vivo*, består denna slags tumör av en kärna, och ett yttre lager av cancerceller som invaderar närliggande vävnad. Celler som utgör tumören tros vara av olika fenotyp, det vill säga ha olika egenskaper. De två fenotyperna återfinns i huvudsak i varsin av dessa delar av tumören: i kärnan främst *proliferativa* (förökande) celler, och i det yttre skiktet i första hand *motila* (rörliga) celler. Motila celler kan dock även finnas i kärnan, och vice versa. Tumören kan även växa mer infiltrativt och diffust i vissa fall, men mer kompakt i andra fall.

Behandlingen av GBM består idag av operationer med målet att avlägsna så stor del av tumören som möjligt, följt av strålningsterapi och cytostatika. GBM:s oregelbundna spridningsmönster försvårar behandling avsevärt, i synnerhet kirurgi, då risken för återväxt är hög om inte alla cancerceller avlägsnas [1], [4]. Det har gjorts flera försök att matematiskt modellera GBM och jämföra simuleringarna med faktiska tumörer [5]. En väl anpassad modell kan avslöja spridningsmönster och därmed ge ledtrådar om vilka biokemiska processer som styr tumörens spridning. Dessa processer kan i sin tur utgöra måltavlor för utveckling av nya läkemedel och behandlingsmöjligheter.

Det är dock inte endast behandling, utan även studerandet av dessa tumörer som försvåras till följd av deras oregelbundna natur. Det finns en uppsjö av studier av GBM som gjorts utanför människohjärnan, och de kommer i olika antal dimensioner och i diverse miljöer. Exempelvis går det att använda sig av cellinjer, vilket är kulturer av celler som extraherats från tumörer *in vivo*, för att odlas och tillåtas växa kontrollerat i homogena kollagenbaserade matriser [1]. När GBM växer utanför sin naturliga miljö, *in vitro*, går dock viss information förlorad. Påverkan densiteten av vit hjärns substans har på tillväxten, hur cellerna rör sig nära nervtrådar, hjärnans kemiska gradienter och blodkärls inverkan är bara en bråkdel av effekterna som inte kan tas i beaktning när cellerna odlas artificiellt [6]. GBM som växer i en homogen matris får därför alltid en sfäriskt symmetrisk form, vilket inte överensstämmer med det som observeras *in vivo*. Fördelen med studier som görs *in vitro* är att man mer fokuserat kan studera och modellera cell-cell interaktioner. Dessa finns det nämligen anledning att tro är bevarade och relevanta för spridning även inom mänsklig vävnad [1]. Modeller som enbart tar hänsyn till cell-cell interaktioner kan därför utgöra en grund för vidareutvecklade modeller som även tar hänsyn till tumörens omgivning.

Det finns stor variation också bland de existerande matematiska modeller som gjorts av GBM. Somliga angriper problemet deterministiskt, med hjälp av partiella differentialekvationer [4], [6], medan andra modeller är stokastiska slumpvandringar av olika slag [1], [7]. I detta arbete kommer fokus läggas på det sistnämnda tillvägagångssättet. I denna typ av modell tillåts celler typiskt sett ta steg mellan diskreta punkter på ett rutnät med viss bestämd sannolikhet. Varje gång de motila cellerna tar ett steg i planet, tar modellen ett steg i tiden. Därtill kan både punkter och celler tillskrivas olika egenskaper. För punkterna gäller detta exempelvis hur många celler som samtidigt kan vistas där, medan cellerna tillerkänns parametrar som bestämmer sannolikheten för bland annat att de delar på sig eller byter fenotyp. Dessa villkor ger ett kraftfullt sätt att simulera tillväxten för GBM. Tektonidis m.fl. har utvecklat en diskret slumpvandringsmodell av GBM:s spridningsmönster *in vitro*, vilket utgör den teoretiska grunden för denna studie [1].

1.1 Problemformulering

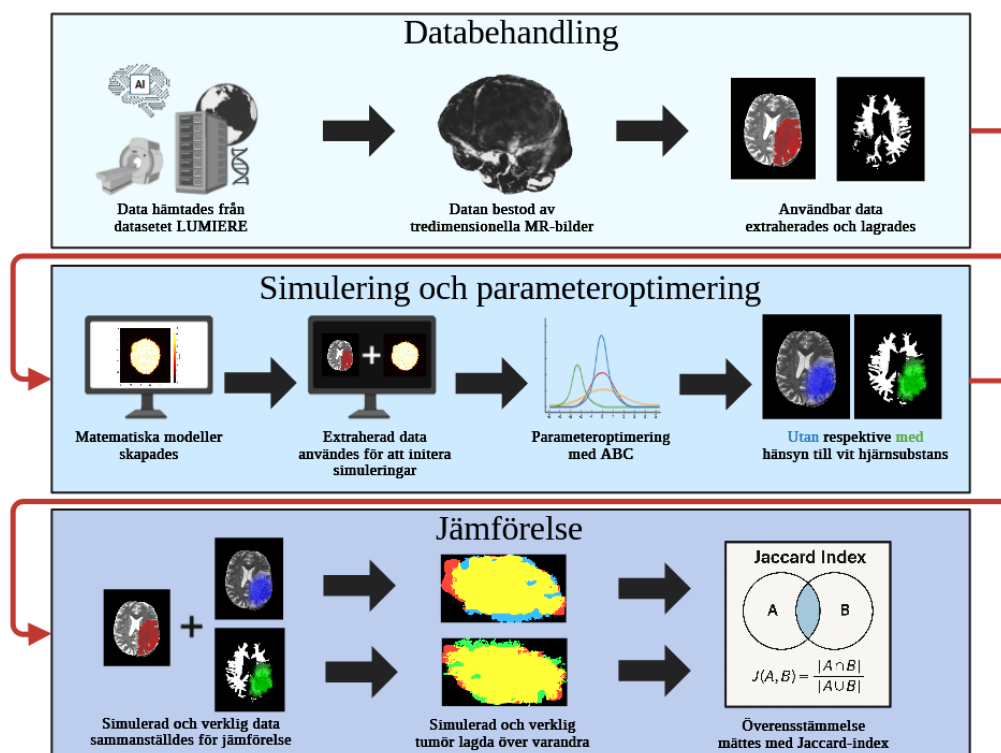
Arbetets huvudfråga är huruvida en modifierad version av spridningsmodellen skapad av Tektonidis m.fl., som tar hänsyn till cellers interaktion med vit hjärns substans, kan uppnå simuleringar vars spridningsmönster efterliknar verkliga tumörer bättre än de från originalmodellen.

1.2 Syfte

Det finns ett behov av bättre modeller av glioblastoms spridningsmönster för att kunna uppskatta sjukdomens utvecklingsförlopp. Därför syftar detta arbete till att utveckla Tektomidis m.fl.:s diskreta slumpvandringmodell för simulering av GBM-tillväxt så att cellernas rörelsemönster påverkas av förekomsten av vit hjärnsubstans, och sedan jämföra den modifierade versionen med originalet genom att mäta likheten mellan simulation och MR-bilder av verkliga GBM. Likheten kvantifieras med Jaccard-index.

2 Metod

I denna del beskrivs den metod som använts för att utveckla, validera och analysera simuleringsmodeller för glioblastomtillväxt. Metoden omfattar fyra huvudsakliga delar: uppbyggnaden av modellerna, behandlingen av patientdata, skapandet av den initiala tumören och till sist utvärderingen av modellernas kvalitet genom parameteroptimering samt jämförelse med verkliga tumörförlopp. Ett övergripande flödesschema illustreras i figur 1.



Figur 1: Flödesschema för utveckling, optimering och jämförelse av de två simuleringsmodellerna.

2.1 Modellering

Tumörer kommer i många olika former. På makroskopisk nivå karakteriseras de av sin storlek och form. Dessa attribut har dock sina rötter i mikrokosmos, där beteenden hos individuella celler leder till de utseenden vi observerar. Dessa beteenden kan i huvudsak delas in i hur cellerna interagerar med varandra (cell-cell-interaktioner) och hur de samspelar med omgivningen (cell-omgivning-interaktioner).

Som nämnts ovan kan tumörens ingående celler delas i två separata fenotyper: proliferativa och motila celler [1], av vilka proliferativa celler kan föröka sig men kan inte röra på sig och motila celler kan röra på sig men inte föröka sig. Denna mekanism av uppdelning kallas för *Go-or-Grow-hypotesen* och har bekräftats experimentellt [4], men som i all biologi framträder verkligheten i mer

nyanserade former. Att göra detta antagandet gällande cellernas fenotyp är dock en förenkling som visat sig användbar för simulering. Viktigt att förstå är att uppdelning i fenotyper är en dikotomi. En cell är alltså antingen proliferativ eller motil, men aldrig något tredje alternativ eller båda samtidigt.

Två modeller kommer att beskrivas i detalj. Den första modellen är baserad på den som skapats av Tektonidis m.fl., och består enbart av cell-cell-interaktionerna. Hädanefter kommer denna att refereras till som *T-modellen*. Den vidareutvecklade modellen inkluderar även interaktionen mellan celler och vit hjärnsubstans som förekommer i omgivningen, och kommer framöver att kallas *VS-modellen*.

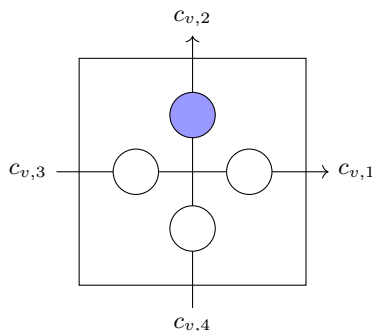
2.1.1 Modellering av cell-cell interaktioner

T-modellen är en så kallad *lattice gas cellular automata* (LGCA) [1]. Detta är en diskret modell, där celler simuleras individuellt och tiden diskretiseras i form av tidssteg. Rent praktiskt innebär ett tidssteg en uppdatering av alla cellers position och fenotyp. Uppdateringarna är baserade på cellernas stundvisa position, fenotyp samt den lokala celldensiteten. I denna modell beaktas tre centrala spridningsbeteenden baserade på cell-cell-interaktioner: celledelning, byte av fenotyp och migration. Dessa mekanismer beskrivs mer ingående i de följande avsnitten.

LGCA

LGCA är en diskret modell där cellerna kan befinna sig i två gitter, här benämnda \mathcal{L}_p och \mathcal{L}_m , som huserar proliferativa respektive motila celler. Båda gittern är tvådimensionella och har samma storlek $L_1 \times L_2 \subseteq \mathbb{Z}^2$ samt varsitt eget koordinatsystem. Det går emellertid att definiera ett totalt gitter \mathcal{L}_{tot} som rymmer både \mathcal{L}_p och \mathcal{L}_m . Punkterna i gittern kallas för noder och de noder som delar samma koordinat i de två gittern kallas systemnoder. I varje nod finns ett visst antal kanaler. En kanal kan antingen vara tom eller innehålla en cell. Nodkapaciteten för \mathcal{L}_p och \mathcal{L}_m kallas b_p respektive b_m . Den totala nodkapaciteten definieras som $\tilde{b} = b_p + b_m$. Kanalstrukturen för \mathcal{L}_m -noderna visas i figur 2. Varje kanal i en nod tillhörande \mathcal{L}_m motsvarar en av de fyra riktningarna som cellen kan röra sig i, vilka beskrivs av

$$c_{v,1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad c_{v,2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad c_{v,3} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad c_{v,4} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \quad (1)$$

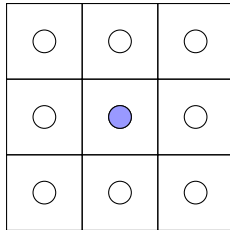


Figur 2: Visuell representation av kanalstrukturen i \mathcal{L}_m -noderna. Varje cirkel representerar en kanal och den ifyllda kanalen, i detta fall $c_{v,2}$, innehåller en cell. Om cellen migrerar gör den det till $c_{v,2}$ kanalen i noden som ligger direkt ovanför den illustrerade noden. Figuren är en omföreställning av figur A1 i Tektonidis m.fl. [1].

Kanalstrukturen i noderna i \mathcal{L}_p är betydligt enklare än de i \mathcal{L}_m eftersom det inte finns några interna riktningar att ta hänsyn till. Varje nod kan innehålla $b_r \in \mathbb{N}_0$ antal proliferativa celler. Böttger m.fl. antog att $b_r = 8$ när de formulerade sin tvådimensionella LGCA-modell och därför görs samma antagande här [8].

Celldelning

Proliferativa celler förökar sig genom att dela på sig i en process som kallas mitos. Detta modelleras genom att en proliferativ cells åtta omgivande noder undersöks för att se om de är lediga. För varje direkt omgivande nod i \mathcal{L}_p som innehåller $b_p - 1$ eller färre celler finns det en konstant sannolikhet P_p att cellen delar sig och skapar en ny dottercell i den fria grannoden, se figur 3. Efter processen finns modercellen kvar i sin ursprungliga nod.



Figur 3: Visuell representation av kanalstrukturen i \mathcal{L}_m noderna. Varje ruta representerar en nod och varje cirkel innehåller åtta kanaler. Den ifyllda cirkeln representerar en cell i en av kanalerna i noden. Cellen kan föröka sig med en sannolikhet P_p till en kanal i en av de lediga grannoderna.

Övergång mellan fenotyper

I varje tidssteg kan celler byta fenotyp. En cell byter från proliferativ till motil med sannolikheten P_{pm} , och från motil till proliferativ med sannolikhet P_{mp} . Båda dessa sannolikheter är beroende av den lokala celldensiteten ρ , vilken definieras som det totala antalet celler i ett par systemnoder. När den lokala celldensiteten ökar, ökar även sannolikheten för att en cell byter fenotyp. Sannolikheterna definieras enligt

$$P_{pm} = \alpha_{pm} + \frac{1}{2} (\beta_{pm} - \alpha_{pm}) (1 + \tanh(k_{pm}(\rho - \theta_{pm}))),$$

$$P_{mp} = \alpha_{mp} + \frac{1}{2} (\beta_{mp} - \alpha_{mp}) (1 + \tanh(k_{mp}(\rho - \theta_{mp}))).$$

Här representerar α_{pm} och α_{mp} de lägsta sannolikheterna för respektive övergång, medan β_{pm} och β_{mp} motsvarar de högsta. Tröskelvärdena θ_{pm} och θ_{mp} anger vid vilken lokal celldensitet sannolikheten förändras mest markant och parametererna k_{pm} och k_{mp} styr känsligheten med vilken sannolikheten förändras runt dessa tröskelvärden.

Migration

De motila cellernas rörelse bygger på två olika mekanismer; *propagering* och *repulsiv omorientering* [1]. Propagering är den rörelse de motila cellerna genomför. Vid initieringen tilldelas varje motil cell en riktning. Denna är en av fyra möjliga som beskrivs i ekvation (1). Endast i fallet då omorientering sker byter cellen sin rörelseriktning. Utan omorientering fortsätter den sin rörelse i samma riktning som vid det föregående tidssteget. Beräkningen av propageringen gjordes genom att använda cellens aktuella riktning för att beräkna nästa position genom

$$(x_{ny}, y_{ny}) = (x_{\text{gammal}} + v_x, y_{\text{gammal}} + v_y),$$

där $(x_{ny}, y_{ny}) \in L_1 \times L_2$ är positionen cellen befinner sig i vid nästa tidssteg.

Den repulsiva omorienteringsprocessen beror på två faktorer: nodflödet och densitetsgradienten. Nodflödet är den genomsnittliga rörelsen bland motila celler i en nods närhet. Den beräknas genom att summera granncellers rörelse i olika riktningar, viktat med antalet celler som rör sig i varje riktning. Detta ger en vektor $J = (J_x, J_y)$, som representerar genomsnittlig rörelsetrend hos motila celler i nodens närhet. Matematiskt beskrivs dess komponenter enligt

$$J_x = \sum_{(dx, dy) \in D} dx \cdot N_m(x + dx, y + dy),$$

$$J_y = \sum_{(dx,dy) \in D} dy \cdot N_m(x + dx, y + dy).$$

Här är $(x, y) \in L_1 \times L_2$ den aktuella positionen för cellen, D mängden av alla riktningsvektorer i ekvation (1), dx och dy förflyttningen i x respektive y -led för varje grannod, samt $N_m(x + dx, y + dy)$ antalet motila celler i grannoden.

Densitetsgradienten G uppskattar hur celltätheten förändras i närheten av en punkt (x, y) , där G_x och G_y beskriver förändringen i x - respektive y -led. Den beräknas som en viktad summa över alla riktningar $(dx, dy) \in D$, där D är mängden av alla riktningsvektorer i ekvation (1), enligt

$$G_x = \sum_{(dx,dy) \in D} dx \cdot \rho(x + dx, y + dy),$$

$$G_y = \sum_{(dx,dy) \in D} dy \cdot \rho(x + dx, y + dy).$$

I dessa uttryck motsvarar $\rho(x + dx, y + dy)$ celltätheten i varje grannposition. Resultatet är en vektor $G = (G_x, G_y)$ som pekar i riktningen av punkten med högst celltäthet och påverkar hur motila celler omorienterar sig.

Genom att kombinera J och G beräknas sannolikheten för att riktning $i \in \{1,2,3,4\}$ ska väljas enligt

$$P_O(c_{v,i}) = \frac{\exp(-\gamma \cdot (G_x J_x + G_y J_y))}{Z}.$$

Här utgör $\exp(-\gamma \cdot (G_x J_x + G_y J_y))$ ett mått på hur lämplig en riktning är, givet nodflöde och densitetsgradient. Måttet normeras över alla möjliga riktningar så att summan av alla P_O blir 1. Z representerar normeringsfaktorn.

Skalarprodukten $G_x J_x + G_y J_y$ är ett mått på överensstämmelsen mellan rörelseriktning och densitet. Om celler tenderar att röra sig mot tätare områden minskar detta värde, och sannolikheten för att välja denna riktning sjunker. Omvänt, om cellen redan rör sig bort från ett väldigt tätt område, ökar P_O . Parametern γ styr känsligheten för denna mekanism, ju högre värdet är, desto starkare blir avstötningen från de tätare områdena.

I T-modellen beskrivs migrationens sannolikhet av variabeln P_m . Detta är den konstanta sannolikheten att en motil cell rör på sig i ett givet tidssteg. Om ett slumpstal från intervallet $[0,1]$ är mindre än P_m , sker migration enligt propagering och eventuell omorientering som beskrivits ovan.

2.1.2 Modellering av cell-omgivning-interaktioner

Cellers beteende påverkas inte enbart av andra celler, utan också av den omgivande vävnaden. I den vidareutvecklade VS-modellen integreras denna påverkan genom att cellers migration görs beroende av underliggande vävnadstyp.

Vit hjärnsubstans

En hjärna består huvudsakligen av grå och vit substans. Den vita substansen består av nervfibrer kallade axon som är omgivna av myelin, ett isolerande vitt fettlager. Studier har visat att patienter vars GBM har växt i delar av hjärnan med lägre densitet av vit hjärnsubstans har en bättre prognos för överlevnad [9]. GBM kan dock även sprida sig till grå substans, men spridningen sker då vanligtvis långsammare [10]. För att modellera detta beteende definierades två olika migrationssannolikheter beroende av vilken omgivning cellen befinner sig i. P_{vs} och P_{gs} är migrationssannolikheterna i vit respektive grå substans. En illustrativ testkörning med $P_{vs} \gg P_{gs}$ visas i figur 4, där tumörcellerna i grönt syns främst sprida sig längs den vita substansen.



Figur 4: En testkörning av modellen där $P_{vs} \gg P_{gs}$. Vita pixlar representerar vit substans och varje grön pixel innehåller minst en tumörcell.

Hjärnskalet

För att förhindra att simuleringen av tumörcelltillväxten skulle ske utanför hjärnans anatomiska gränser integrerades en hjärnmask i beräkningarna. Hjärnmasken hämtades från den aktuella segmenteringen i datamängden. Hjärnskalet representerades sedan av en tvådimensionell binär matris i vilken element innanför hjärnans gränser hade värdet 1, medan övriga tilldelades värdet 0.

I de delmoment som inkluderade cellrörelse i någon form implementerades hjärnskalet för att säkerställa att simuleringen endast tillät förflyttning och celledelning till positioner inom hjärnans gränser. Detta uppnåddes genom att inkludera ett `if`-villkor vilket i varje delmoment verifierade att koordinaterna (x_{ny}, y_{ny}) låg inom gitterstorleken och att motsvarande värde i hjärnmasken var lika med 1.

2.2 Databehandling

För att kunna jämföra simuleringsmodellerna med verkliga sjukdomsförlopp användes longitudinell patientdata från den öppet tillgängliga datamängden LUMIERE [11]. Datamängden består av magnetresonansbilder (MR) från 91 patienter med GBM, där varje patient följts över tid med flera bildtagningstillfällen. Tidsintervallen mellan dessa tillfällen varierar. Totalt innehåller datasetet 638 undersökningstillfällen och 2 487 MR-bilder. För varje tillfälle finns fyra sekvenser: T1, cT1, T2 och FLAIR. Därtill tillhandahålls segmenteringsmasker av tumörerna, alltså delar av hela hjärnbilden som isolerats för att kunna skilja frisk vävnad från sjuk vid processering av data. Dessa har skapats av de två automatiserade maskininlärningsverktygen DeepBraTumIA och HD-GLIO-AUTO, och utvärderats av experter inom området.

För varje patient finns två versioner av alla sekvenser, varav den ena utgörs av bilder som transformerats till ett gemensamt koordinatsystem. Detta innebär att en given koordinat i en bild motsvarar samma anatomiska punkt i alla andra bilder, oavsett sekvens och tidpunkt. Saker som avvikelser i patientpositionering, maskininställningar och sekvensparametrar innebär att ursprungliga bilder har olika orientering och således inte är direkt jämförbara. En standardisering är alltså en förutsättning för att möjliggöra longitudinell analys, men i detta fall är den inte gjord på egen hand.

Utifrån denna omfattande datamängd har en rad bearbetningssteg genomförts för att extrahera jämförbar data gentemot simuleringsmodellerna. Detta inkluderar filtrering av operationsfria tidsintervall, extraktion av tumöregenskaper, samt slutlig transformering av data till ett format anpassat efter simuleringsmodellerna. Dessa steg redogörs i följande avsnitt.

2.2.1 Filtrering och extraktion av relevant data

Trots datamängdens storlek fanns det få patientförlopp som visade sig nyttiga att använda till jämförelse med den simulerade tumören. Eftersom modellen simulerar tillväxt över tid filtrerades först bara patienter med minst två bildtillfällen efter kirurgiskt ingrepp. Informationen om kirurgiska ingrepp fanns tillgänglig i en fil tillhörande datamängden. Denna fil benämns i arbetet ”RANO-filen” (RANO är en förkortning för *Response Assessment in Neuro-Oncology*).

Mediantiden det tar för ett GBM att dubbla i volym är cirka 50 dagar [12], varför endast sekvenser tagna med minst sex veckors mellanrum studerades. Därefter sällades patienter bort i de fall standardiseringen eller bildkvaliteten inte bedömdes vara tillräckligt bra. Slutligen gjordes ett manuellt urval för att säkerställa att inga operationer skett mellan de valda bildtillfällena.

Därefter kopierades de två bildtillfällena som motsvarade början och slut till en separat katalog för vidare bearbetning. Totalt inkluderades 2 patienter i den slutliga analysen.

En mapp innehållande endast de patienter som uppfyllde de ställda kraven, samt ett dokument som redogjorde för vilka veckor som skulle analyseras, skapades. Från dessa användes egenskriven kod för att ladda data för de relevanta tidpunkterna och utföra de manipulationer som krävdes för att senare kunna jämföra data från de verkliga tumörerna med de som simulerats. Datamängden som till slut användes kom från segmenteringarna gjorda av DeepBraTumIA och bilder av hela hjärnan. Ur datamängden extraherades exempelvis koordinater, bildintensitet och voxlarnas storlekar. Dessa kunde sedan användas för bland annat beräkning av tumörens ungefärliga masscentrum.

2.2.2 Val av analysplan och tumörkaraktärisering

De verkliga tumörerna är självfallet tredimensionella, men simuleringsmodellen begränsades till två dimensioner för att underlätta beräkningskraft. Därför behövdes ett lämpligt plan av tumören väljas ut för att möjliggöra jämförelse mellan verklighet och simulering. Detta representativa plan valdes till den transversala genomskärning i vilken tumörens masscentrum låg under den tidigaste veckan av det undersökta intervallet. Samma skiva användes för bilden av tumören som tagits under intervallets sista vecka, även om masscentrums koordinater förändrats. Detta eftersom målet var att jämföra tillväxten i ett och samma skikt av tumören till bästa förmåga, snarare än de två plan som innehåller mest tumör vid de olika tillfällena.

Beräkning av masscentrum gjordes med ett skript som räknade ut det viktade medelvärdet av voxlarna position givet en viss referenspunkt. Voxlarnas faktiska massa går inte att extrahera från datamängden. Istället uppskattades densiteten genom antagandet att T2-intensiteten är direkt korrelerad till tumörens lokala densitet [13]. Varje voxels ”densitet” multiplicerades med dess rumsliga storlek för att erhålla ”massan”. En begränsning med antagandet är att densiteten av GBM enligt vissa källor är mer eller mindre homogen [14].

För beräkning av masscentrum användes formeln

$$\mathbf{x}_{\text{mc}} = \mathbf{r}_{\text{ref}} + \frac{1}{M} \sum_i m_i (\mathbf{x}_i - \mathbf{r}_{\text{ref}}).$$

Här representerar \mathbf{r}_{ref} referenspunkten (ofta valdes den geometriska mittpunkten av hjärnan), medan \mathbf{x}_i är de rumsliga koordinaterna av voxel i , m_i samma voxels massa och M är tumörens totala massa.

För att skapa hjärnmasken extraherades data som innehöll en förbehandlad MR-volym av hjärnskalet från en segmenteringsfil. Därefter skivades masken och skivan innehållande tumörens masscentrum valdes ut som spatial begränsning för simuleringen, tillsammans med den motsvarande skivan av tumörmasken.

Även en CSV-fil innehållandes värden på en så kallad ”effektiv radie” för varje tumör vid dessa tillfällen skapades. Denna effektiva radie är radien tumören hade haft om den vore en perfekt sfär, och formeln som användes för beräkning av denna erhöles genom att lösa ut radien ur formeln för

volymen av ett klot

$$V = \frac{4\pi R_{\text{eff}}^3}{3} \iff R_{\text{eff}} = \sqrt[3]{\frac{3V}{4\pi}}. \quad (2)$$

Som volym V användes den sammantagna volymen av de voxlar som ingick i segmenteringen av tumören.

2.2.3 Intensitetsgitter, diskretisering och normering

För att initiera simuleringen behövdes ett starttillstånd som representerade den observerade tumörens fördelning. Detta konstruerades genom att extrahera intensitetsvärden från en T2-viktad MR-bild inom den segmenterade tumörvolymen. För varje vald vecka, och för både segmenteringsmasken och den T2-viktade bilden, användes den axiella skiva i vilken tumörens masscentrum låg. Innan några värden extraherades interpolerades T2-bilden till samma rumsliga upplösning som segmenteringen. Tumörområdet definierades som de pixlar där segmenteringsmasken var strikt större än noll.

De extraherade intensiteterna inom tumörområdet bildade ett tvådimensionellt intensitetsgitter $I(x, y)$. För att skalan skulle vara enhetlig normerades detta enligt

$$I_{\text{norm}}(x, y) = \frac{I(x, y)}{\max(I(x, y))}, \quad \text{där } \max(I(x, y)) > 0.$$

Det normerade gittret $I_{\text{norm}}(x, y)$ fick därefter genomgå en diskretisering i 13 nivåer enligt

$$I_{\text{disk}}(x, y) = \lfloor 12 \cdot I_{\text{norm}}(x, y) \rfloor / 12.$$

Där den normerade intensiteten multipliceras med 12, avrundas till närmaste heltal och sedan divideras med tolv. Då skalas det tillbaka till intervallet $[0,1]$ i 13 diskreta nivåer, uttryckta i tolfte delar. Just 13 nivåer valdes då de svarar mot simuleringens bärförmåga i varje nod, se 2.3.3.

2.2.4 Segmentering av vit substans

För att implementera omgivningsinteraktion i tumörsimuleringen genomfördes en segmentering av vit substans för varje patient. Denna segmentering gjordes i det plan som tumörens masscentrum låg i och genererades med hjälp av FAST-algoritmen. Denna algoritm tillämpar statistiska metoder för att analysera och klassificera vävnader baserat på bildintensitet i T1-viktade MR-bilder [15]. Ett exempel på en segmentering av vit substans med FAST-algoritmen visas i figur 5.



(a) T1-viktad MR-bild i tumörens masscentrumplan. (b) Segmenterad vit substans från samma snitt.

Figur 5: Exempel på segmentering av vit hjärnsubstans med FAST-algoritmen. Algoritmen identifierar vävnadstyper baserat på bildintensitet i T1-viktade MR-bilder. Den segmenterade vita substansen används för att modellera omgivande vävnad i tumörsimuleringen.

När segmenteringen av vit substans genomförts sparades resultatet i CSV-format.

2.3 Skapandet av den initiala tumören

Inför simulering med LGCA-modellen skapas en initial tumör baserat på bildintensiteten i MR-datan. Två huvudsakliga utmaningar försvårade omräkningen från initial bildintensitet till celldensitet. Dels saknar bildintensitet och celldensitet en linjär korrelation, och dels omöjliggör den enorma mängden celler som finns i en tumör individuell simulering med den tillgängliga beräkningskapaciteten.

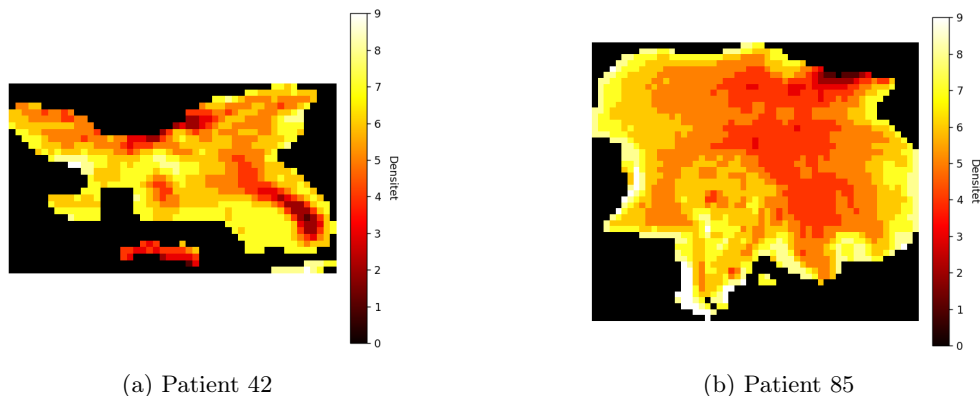
2.3.1 Uppskattning av celldensitet

Ett flertal strategier har använts för att tackla översättningen mellan intensitet och celler. En av dem är användande av AI-modeller [16]. På grund av tidsbrist och resurser antogs en negativ, linjär korrelation mellan T2-intensitet och densitet av tumörceller.

En T2-bild kan tas av en MR-maskin för att mäta magnetiseringsegenskaperna hos vätekärnorna som befinner sig i hjärnan. En radiopuls tvingar vätekärnornas spinn att synkronisera och tiden det tar för dem att återgå till sitt ursprungliga tillstånd kallas för relaxationstid [17]. En längre relaxationstid ger upphov till en högre bildintensitet. Under fältstyrkan 1,5 T har vit substans en ungefärlig relaxationstid 80 ms, grå substans har omkring 100 ms och cerebrospinalvätska har en betydligt längre relaxationstid på 220 ms på grund av den mycket höga andelen vatten [18]. Gliom består av en speciell typ av glialvävnad och är alltså skild från normal vit och grå substans [19]. Relaxationstiden av GBM varierar då tumörer karaktäristiskt är heterogena i sin histologiska karaktär, men Oh m.fl. har mätt den genomsnittliga relaxationstiden av gliom till 160 ms under samma fältstyrka [20].

GBM består av levande eller döda tumörceller, blodkärl samt omgivande vätskerika ödem [20]. Ödemens relaxationstid är förhållandevis lång på grund av deras höga vatteninnehåll och nekrotiska områden har en relaxationstid jämförbar med den av ödem [21]. Levande tumörceller har en något lägre relaxationstid än ödem, men skillnaden är inte avsevärd [22]. Densiteten av levande celler antas därför vara högre där T2-intensiteten är lägre. Ett tillkortakommande med detta antagande är exempelvis att blödningar, som ofta förekommer runt tumören som en konsekvens av utbredningen, har en kort relaxationstid [23]. Dessutom har blodkärl i sig en kort relaxationstid [24].

För att konstruera en initial fördelning av celler i modellen normerades och inverterades intensitetsvärdena inom det segmenterade tumörområdet. Sedan diskretiserades det i 13 nivåer. Den resulterande diskreta intensiteten multiplicerades därefter med det maximala antalet celler per nod för att approximativt beskriva den lokala celldensiteten. För mer detaljer om hur intensitetsgittret extraherades och bearbetades, se 2.2.3. I figur 6 plottas den approximativa celldensiteten för de initiala tumörerna hos patient 42 och patient 85.



Figur 6: Plott av den approximerade celldensiteten för de initiala tumörerna hos patient 42 och patient 85. Celldensiteten beräknades genom att normera och invertera intensitetsvärdena inom det segmenterade tumörområdet och sedan diskretisera detta i 13 nivåer. Resultaten används för att beskriva den lokala celldensiteten i tumörmodellen. En maximal celldensitet av 12 var även här tillåten men det uppnåddes aldrig en högre densitet än 9 celler per nod.

2.3.2 Förenkling av LGCA-modellen

Den genomsnittliga celldensiteten av 109 patienters GBM mättes till $3308,71 \pm 1915,32$ cells/mm² [25]. Pixelstorleken i T2-bilderna är 1 mm² och gridstorleken är 182×218 pixlar. På grund av bristande beräkningskraft sätts nodkapaciteten till 12 celler per nod och en vandrare i LGCA-modellen tilläts representera $3308,71/12 \approx 276$ celler. Med andra ord får en pixel motsvara en nod.

2.3.3 Initial fördelning av fenotyper

Eftersom celler antingen kan vara proliferativa eller motila behövs en strategi för att bestämma en fördelning av fenotyperna när den initiala tumören skapas. Ki67 är ett protein som uttrycks nästan exklusivt i celler som aktivt delar sig, det vill säga proliferativa celler [26]. Man kan använda skillnaden i ett Ki67-index för att uppskatta skillnaden i andelen av proliferativa och motila celler. För den vanliga glioblastomgenotypen U87WT var indexet 11/3 gånger större i tumörens kärna jämfört med det invasiva periferiet [4].

Ki67-indexet användes för att stokastiskt estimera en initial fördelning av fenotyperna. Eftersom andelen proliferativa celler är som högst i kärnan antogs en maximal sannolikhet, 8/12, att en cell som befinner sig precis i mitten av kärnan blir proliferativ [1]. En effektiv radie centrerad i tumörens masscentrum räknades ut av tumörgenomskärningen enligt ekvation (2). Sannolikheten att en cell blir proliferativ minskar linjärt i takt med att avståndet från centrum ökar. Vid R_{eff} är sannolikheten $3/11 \cdot 8/12 = 2/11$. Om noderna är över sin bärkapacitet för den specifika fenotypen läggs istället den lediga fenotypen till i noden.

2.4 Simulering och parameteroptimering

För att kunna anpassa modellens parametrar till verkliga observationer krävs en metod för parameterinferens. I detta arbete används ett bayesianskt angreppssätt som möjliggör inkorporering av osäkerhet och förkunskap i parameteroptimeringen.

2.4.1 Jaccard-index

För att kvantitativt kunna utvärdera hur väl de simulerade tumörsegmenteringarna kongruerar med datasegmenteringarna användes Jaccard-index \mathcal{J} , som är ett mått som används för att mäta hur lika två mängder är. Detta index definieras som kvoten mellan snittet och unionen av två mängder A och B , alltså

$$\mathcal{J}(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (3)$$

I detta fall representerar ena mängden modellerade, binäriserade tumörer som observerats och den andra mängden representerar den observerade, binäriserade tumören. Snittet $A \cap B$ motsvarar det proportionella överlappet mellan tumören och modellen, medan unionen $A \cup B$ fylls av alla gemensamma element i både simuleringen och tumörområdet. Kardinaliteten $|C|$ av en mängd har samma värde som antalet unika element i mängden. Kvoten $\mathcal{J}(A, B)$ antar värde mellan 0 och 1, som motsvarar ingen respektive fullständigt överlapp [27].

2.4.2 Approximativ bayesiansk beräkning

Modellen som har tagits fram involverar ett flertal parametrar, och deras optimala värden är allt annat än självklara. Ett bra ramverk för att ta fram värden för dem är bayesiansk inferens, som i grund och botten bygger på Bayes sats, som formuleras enligt

$$p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta)p(\theta)}{p(\mathcal{D})}.$$

Här är θ en uppsättning parametrar och \mathcal{D} observerade datapunkter. Sannolikhetsfunktionerna $p(\theta)$ och $p(\theta | \mathcal{D})$ kallas *a priori*-, och *a posteriori*- fördelningar, och beskriver kunskapen om θ

innan respektive efter den nya datan tagits i beaktande. De två andra funktionerna, $p(\mathcal{D} | \theta)$ och $p(\mathcal{D})$, kallas *likelihood* och *evidence*. Den sistnämnda är ointressant vid jämförelse av olika parameteruppsättningar till följd av dess oberoende av θ och kan tolkas som en normaliseringsfaktor. Likelihoodfunktionen uttrycker sannolikheten att observera datan \mathcal{D} givet parametrarna θ , och kan vara svår att få fram konkreta värden på [28]. Därför har metoder som kringgår likelihood utvecklats. Eftersom troligheten är svår, eller kanske omöjlig, att uttrycka explicit för den här typen av modell används i detta arbete *sequential Monte Carlo approximate Bayesian computation* (SMC-ABC).

Algoritmen går ut på att parametrarna initieras med bred *a priori*-fördelning, som sedan förbättras stegvis. Även en toleransnivå, ε , sätts i detta skede. Detta värde definierar den maximalt tillåtna skillnaden i "avstånd" mellan simulerad och observerad data för att en parameter ska accepteras i den approximativa posteriorfördelningen. För varje steg väljs ett förslag på parametrar från en viktad fördelning, och tumörens tillväxt simuleras. Efter simuleringen är färdig beräknas avståndet mellan simulation och verklighet, och parameterförslaget antingen accepteras eller avslås beroende på om avståndet är högre eller lägre än den aktuella toleransen. Efter detta uppdateras fördelningens vikter, och ε minskas, varpå processen upprepas.

Parameteroptimeringen utfördes med ett skript för parallelliserad SMC-ABC som använde resurser ur Python-modulen pyABC [29], där *a priori*-fördelningar specificerades som trunkerade normalfördelningar. Vid varje parameterförslag genereras en tumörsimulering baserad på en initial MR-bild och utvecklas över ett förutbestämt antal tidssteg. Den simulerade tumören binäriseras och jämförs därefter med den binäriserade observerade tumören med hjälp av Jaccard-index, mer specifikt $1 - \mathcal{J}$, där \mathcal{J} beskrivs i ekvation (3), eftersom ett värde nära 0 då motsvarar ett nästan perfekt överlapp mellan simulerad och observerad tumörutbredning, vilket eftersträvas i förhållande till ε . Om avståndet understiger ε accepteras parameterförslaget i den approximativa posteriorfördelningen. Den initiala toleransnivån sattes till 0,35 och efter 100 accepterade simuleringar uppdateras epsilon enligt metoden `QuantileEpsilon` ur pyABC med ett α -värde på 0,8. Detta innebär att den nya toleransen baseras på den 80:e percentilen av accepterade simuleringar i den föregående generationen. Körningar av detta skript utfördes på datorklustret Gantenbein, som finns på Institutionen för matematiska vetenskaper på Chalmers. Klustret har 112 kärnor, med två trådar per kärna.

2.4.3 Estimering av *a priori*-fördelningar

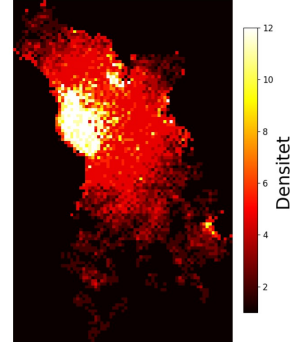
Inför ABC-inferensen specificerades *a priori*-fördelningar för modellparametrarna baserat på resultat från ett antal inledande pilotsimuleringar. Syftet var att begränsa parameterrummet till områden som gav realistisk tumörtillväxt inom rimlig beräkningstid.

Parametrarna P_p , P_m (eller P_{vs} respektive P_{gs} i VS-modellen) och tidssteg t visade sig ha en särskild påverkan på tumörens storlek och simuleringstid. För båda modeller fixerades tidssteget till 100. *A priori*-fördelningar för P_p och P_m valdes med relativt snäv spridning baserat på simuleringar där tumörutbredningen matchade eller översteg de observerade tumörområdena.

I VS-modellen leder höga värden på P_{vs} i kombination med låga värden på P_p och P_{gs} till tydlig preferens för spridning i vit substans, vilket illustreras i figur 7. I T-modellen, där tumören växer utan hänsyn till vävnadstyp, tenderar höga P_p att ge radiellt symmetriska tillväxtmönster med hög densitet, vilket exemplifieras i figur 8. En kombination av låg P_p och hög P_m valdes därför för att möjliggöra mer asymmetrisk utbredning.

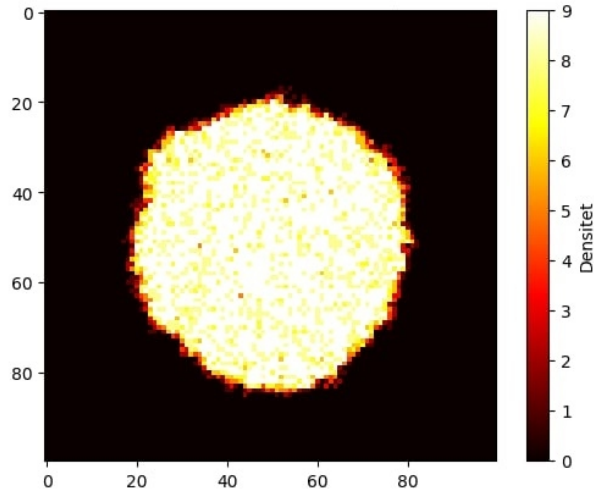


(a) Simulerad tumör med en cell som startpunkt. Varje grön pixel motsvarar minst en cell.



(b) Densitetskarta

Figur 7: Simulering av VS-modellen med manuellt valda parametrar. Tumören har en tydlig preferens att sprida sig i vit substans när man väljer 1500 tidssteg, $P_p = 0,004$, $P_{vs} = 1$ och $P_{gs} = 0,001$. Notera dessutom att den övergripande densiteten är låg. Detta är ett karaktäristiskt drag när P_{vs} dominerar över P_p .



Figur 8: Simulerad tumör med en cell som startpunkt. Tillväxten är i stort sett radiellt symmetrisk.

För att möjliggöra en rättvis jämförelse mellan modellerna standardiserades samtliga *a priori*-fördelningar, där så var möjligt. De slutliga *a priori*-fördelningarna för övriga parametrar baserades på tidigare pilotsimulationer, med relativt snäv standardavvikelse till följd av begränsade beräkningsresurser. De valda *a priori*-fördelningar för respektive modell sammanfattas i tabell 2 och tabell 3. En stor andel av parameterintervallen är sannolikheter och därför är deras definitionsmängd $[0, 1]$. Resterande definitionsmängder baserades helt på vad som beskrivs av Tektonidis m.fl [1].

Tabell 2: *A priori*-fördelningar för VS-modellen.

| Parameter | Medelvärde(μ) | Standardavvikelse (σ) | Intervall [min, max] |
|---------------|---------------------|--------------------------------|----------------------|
| P_p | 0,002 | 0,002 | [0, 1] |
| P_{gs} | 0,076 | 0,02 | [0, 1] |
| P_{vs} | 0,8 | 0,1 | [0, 1] |
| θ_{mp} | 6 | 1 | [0, 10] |
| θ_{pm} | 8 | 1 | [0, 10] |
| γ | 10 | 1 | [0, 10] |
| α_{mp} | 0,1 | 0,05 | [0, 1] |
| α_{pm} | 0,2 | 0,05 | [0, 1] |
| β_{mp} | 0,2 | 0,03 | [0, 1] |
| β_{pm} | 0,4 | 0,05 | [0, 1] |
| k_{mp} | 0,4 | 0,05 | [0, 10] |
| k_{pm} | 0,4 | 0,05 | [0, 10] |

Tabell 3: *A priori*-fördelningar för T-modellen.

| Parameter | Medelvärde(μ) | Standardavvikelse (σ) | Intervall [min, max] |
|---------------|---------------------|--------------------------------|----------------------|
| P_p | 0,0001 | 0,001 | [0, 1] |
| P_m | 0,6 | 0,1 | [0, 1] |
| θ_{mp} | 6 | 1 | [0, 10] |
| θ_{pm} | 8 | 1 | [0, 10] |
| γ | 10 | 1 | [0, 10] |
| α_{mp} | 0,1 | 0,05 | [0, 1] |
| α_{pm} | 0,2 | 0,05 | [0, 1] |
| β_{mp} | 0,2 | 0,03 | [0, 1] |
| β_{pm} | 0,4 | 0,05 | [0, 1] |
| k_{mp} | 0,4 | 0,05 | [0, 10] |
| k_{pm} | 0,4 | 0,05 | [0, 10] |

3 Resultat

Resultatet av simuleringsmodellen presenteras nedan i form av grafiska och numeriska jämförelser mellan de två modellerna för tumörtillväxt. Fokus ligger på modellernas förmåga att återge observerad tumörutbredning, vilket kvantifierades med Jaccard-index \mathcal{J} . Specifikt valdes simuleringsresultat med högst \mathcal{J} ut från den sista generationen av varje SMC-ABC. Resultatet redovisas först på en övergripande nivå, följt av patientvisa visualiseringar av celldensitet och överlapp mellan simulering och verklig data. Modellerna har anpassats för tumörutbredningen av två patienter, vilka benämns patient 42 och patient 85.

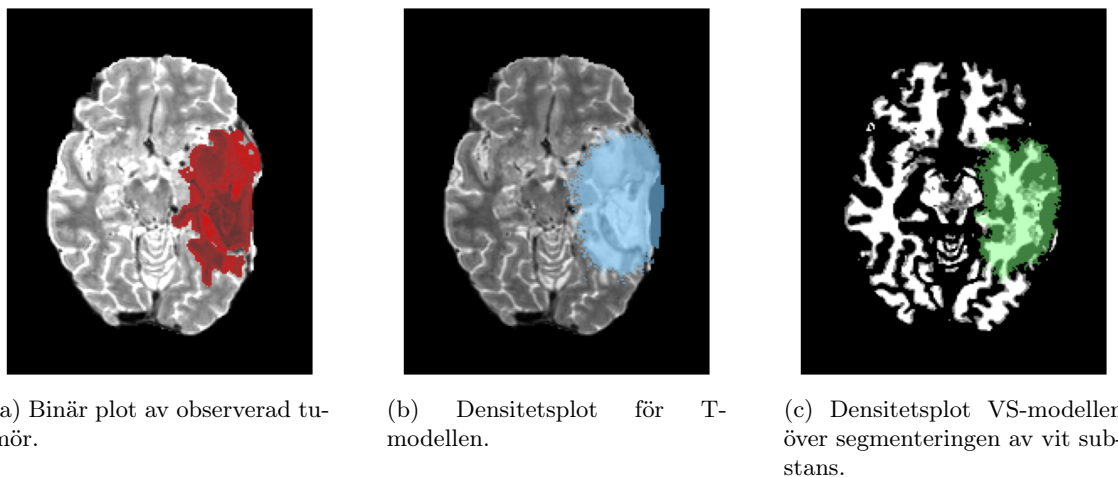
Jaccard-index beräknades med ekvation (3) för att kvantitativt mäta överlappet mellan varje modells prediktion och den segmenterade tumörregionen i MR-bilderna. Ett högre värde på Jaccard-indexet innebär en bättre överensstämmelse mellan den simulerade och den verkliga tumören. För de undersökta patienterna sammanfattas det i tabell 4, men mer djupgående analys presenteras i efterföljande delar.

Tabell 4: Jaccard-index mellan segmenterad tumörregion och simulering för respektive modell och patient. Ett högre värde indikerar större överlapp mellan simulering och verklig tumör. Skillnaden $\Delta\mathcal{J}$ anger förbättringen för VS-modellen jämfört med T-modellen.

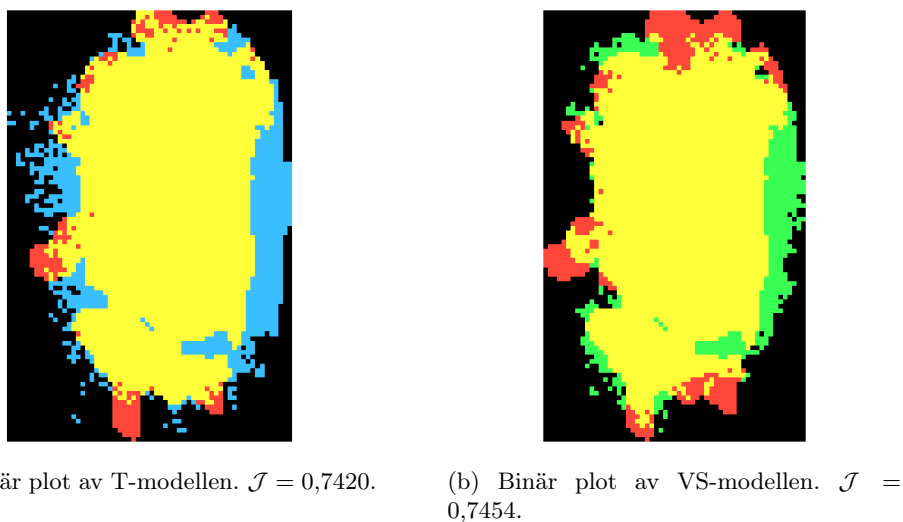
| Patient | \mathcal{J} för T-modellen | \mathcal{J} för VS-modellen | $\Delta\mathcal{J}$ |
|---------|------------------------------|-------------------------------|---------------------|
| 85 | 0,7811 | 0,8229 | 0,0418 |
| 42 | 0,7420 | 0,7454 | 0,0034 |

3.1 Patient 42

För patient 42 ses en relativt liten skillnad i modellernas överensstämmelse med den observerade tumören. I figur 9 visas den observerade tumören i 9a tillsammans med densitetsplottarna för T-modellen i 9b och VS-modellen i 9c. Dessa resultat kompletteras i figur 10, där det binära överlappet mellan modellerna och den observerade tumören med högst Jaccard-index visualiseras. Både T-modellen och VS-modellen uppvisar hög träffsäkerhet, med Jaccard-index på 0,7420 respektive 0,7454. Det gula området i figur 10 indikerar var modellernas prediktion överlappar med tumörens verkliga utbredning. Resultatet tyder på att båda modellerna fångar tumörens lokalisering på ett likvärdigt sätt, med en marginell fördel för VS-modellen.



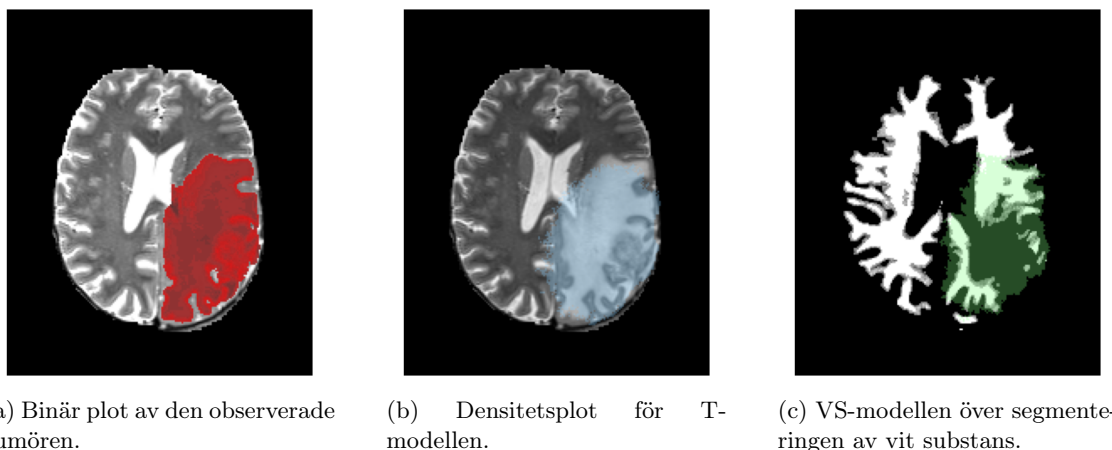
Figur 9: Bilder av observerad tumör och simulerade tumörutbredningar från både T-modellen och VS-modellen för patient 42.



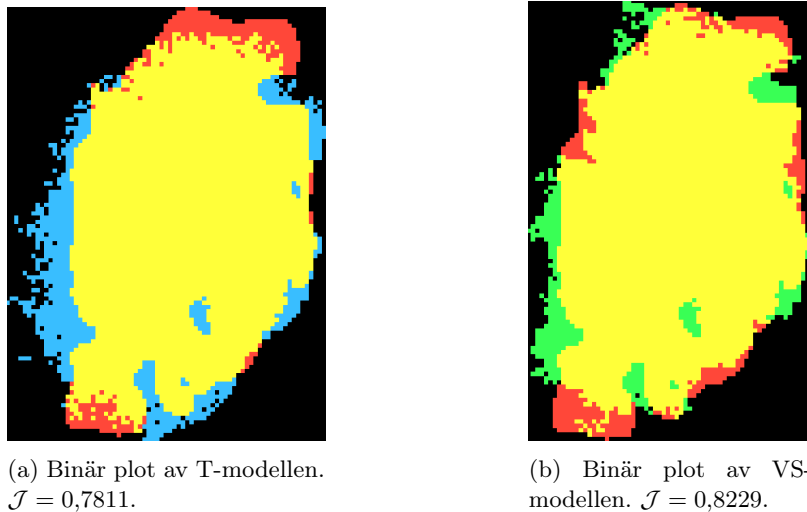
Figur 10: Överlappet mellan de simulerade tumörerna och den observerade tumören med högst Jaccard-index för patient 42. Modellerna syns i blått (T-modellen) och grönt (VS-modellen), medan den verkliga tumören gestaltas röd. Det gula området representerar överlappet mellan modellerna och den observerade tumören.

3.2 Patient 85

För patient 85 är skillnaden mellan modellerna mer markant. I figur 11 illustreras den observerade tumören i 11a tillsammans med densitetsplottarna för T-modellen i 11b och VS-modellen i 11c. Det framgår att tumörcellerna i VS-modellen följer vit substans, särskilt i 11c. Figur 12 visar den binära överlappsanalysen av körningen med högst Jaccard-index. VS-modellen uppnår värdet $\mathcal{J} = 0,8229$, medan det för T-modellen ligger på 0,7811. Detta tyder på att VS-modellen ger en mer exakt prediktion av tumörens utbredning i detta fall, särskilt där migration sker längs vit substans.



Figur 11: Bilder av observerad tumör och simulerade tumörutbredningar från både T-modellen och VS-modellen för patient 85.



Figur 12: Överlappet mellan de simulerade tumörerna och den observerade tumören med högst Jaccard-index för patient 85. Modellerna syns i blått (T-modellen) och grönt (VS-modellen), medan den verkliga tumören gestaltas röd. Det gula området representerar överlappet mellan modellerna och den observerade tumören.

3.3 ABC-SMC populationsstatistik

Tabellerna 5 och 6 sammanfattar ABC-SMC-statistiken för respektive modell och patient. Varje generation t redovisar antalet simuleringar som krävdes för att acceptera 100 partiklar (accepterade parameterkombinationer), tillsammans med den toleransnivå ε som användes. Antalet partiklar hölls konstant mellan modeller och generationer för att möjliggöra en rättvis jämförelse.

Det är tydligt att toleransnivån successivt sänks i varje iteration, vilket speglar hur modellerna successivt anpassas till de observerade data. Skillnader i antal simuleringar mellan modellerna ger en indikation på deras relativa effektivitet vid parameteranpassning.

Tabell 5: Jämförelse av ABC-SMC populationsstatistik för patient 85 med T-modellen och VS-modellen.

| T-modellen patient 85 | | | VS-modellen patient 85 | | |
|-----------------------|--------------|---------------|------------------------|--------------|---------------|
| t | Simuleringar | ε | t | Simuleringar | ε |
| 0 | 335 | 0,350000 | 0 | 159 | 0,350000 |
| 1 | 218 | 0,316021 | 1 | 175 | 0,267363 |
| 2 | 211 | 0,304973 | 2 | 171 | 0,259926 |
| 3 | 232 | 0,298360 | 3 | 178 | 0,255867 |
| 4 | 200 | 0,282639 | 4 | 185 | 0,245079 |
| 5 | 209 | 0,277720 | 5 | 216 | 0,234725 |
| 6 | 503 | 0,253409 | 6 | 258 | 0,228392 |
| 7 | 379 | 0,252758 | 7 | 302 | 0,223807 |
| 8 | 466 | 0,248699 | 8 | 466 | 0,208694 |

Tabell 6: Jämförelse av ABC-SMC populationsstatistik för patient 42 med T-modellen och VS-modellen.

| T-modellen patient 42 | | | VS-modellen patient 42 | | |
|-----------------------|--------------|---------------|------------------------|--------------|---------------|
| t | Simuleringar | ε | t | Simuleringar | ε |
| 0 | 244 | 0,350000 | 0 | 160 | 0,350000 |
| 1 | 189 | 0,335532 | 1 | 184 | 0,312341 |
| 2 | 199 | 0,318389 | 2 | 220 | 0,307696 |
| 3 | 220 | 0,309190 | 3 | 188 | 0,301837 |
| 4 | 212 | 0,297187 | 4 | 248 | 0,293637 |
| 5 | 247 | 0,288762 | 5 | 409 | 0,289922 |
| 6 | 1775 | 0,274014 | 6 | 631 | 0,284302 |
| 7 | 1518 | 0,273298 | 7 | 736 | 0,277458 |
| 8 | 1772 | 0,272482 | 8 | 1354 | 0,274767 |

4 Diskussion

I denna del analyseras resultaten från modellenanpassningen och jämförelsen mellan de olika modellerna. Fokus ligger på att utvärdera modellernas relativa prestanda, effekten av valda *a priori*-fördelningar samt metodologiska och datamässiga begränsningar. Resultatens betydelse diskuteras i relation till modellernas potentiella kliniska användning, med särskild betoning på generaliserbarhet, biologisk rimlighet och möjliga sätt att vidare utveckla dem.

4.1 Jämförelse av modellerna

Jaccard-indexet användes för att kvantifiera kvaliteten på modellerna. Ett högre värde på Jaccard-indexet indikerar ett mer signifikant överlapp mellan observerad och simulerad tumör. VS-modellen uppvisade det högsta Jaccard-indexet för båda patienterna, vilket tyder på att modellen bättre fångar tumörens utbredningsmönster.

Vid betraktandet av ABC-SMC populationsstatistiken för respektive patient framgår en bild som överensstämmer med resultatet från respektive bästa simulering. I tabell 5, som innehåller populationsstatistik för patient 85, presterar VS-modellen tydligt bättre. Den uppvisar snabbare konvergens och kräver färre simuleringar för att nå lägre värden på ε jämfört med T-modellen. Allt detta tyder på att VS-modellen är effektivare på att anpassa sig till patientdata. Tabell 6 å

andra sidan, visar mindre skillnader mellan modellernas prestanda för patient 42. VS-modellen uppvisar heller inte samma effektivitet eller konvergens, vilket stämmer överens med observerat Jaccard-index för bästa simulering för denna patient.

Skillnaden mellan Jaccard-indexen var betydligt mindre för patient 42 än patient 85, se tabell 4. En möjlig förklaring till detta är att en större mängd vit substans finns inom tumörområdet för patient 42 jämfört med patient 85, där det fanns ett hålrum. Detta syns tydligt vid jämförelse av figurerna 9c och 11c. När tumörer växer trycks vit substans undan och produceras inte internt. Att det finns vit substans inom tumören för simuleringen hos patient 42 är en konsekvens av att tumören växer utan att simuleringen tar hänsyn till att tillväxten undantränger vit substans. Det ger upphov till interaktioner mellan cellerna och en "fiktiv" vit hjärnsubstans, som i verkligheten inte skulle vara där.

Det är inte säkert att skillnaden enbart berodde på den fiktiva vita substansen. De snäva *a priori*-fördelningarna för VS-modellen kan ha passerat tumören i patient 85 bättre av ren slump, eller tvärtom att fördelningen gynnat T-modellen mer i patient 42. Parametrarna standardiserades så att de utforskade parameterrummen blev så lika som möjligt under anpassningen. Trots detta visar *a posteriori*-fördelningarna i bilaga B en viss generell osäkerhet bland vissa parametrar, medan andra uppvisar en tydligare konvergens. Vid mer storskalig simulering yrkas därmed på en bredare, och kanske till och med uniform *a priori*-fördelning över intervallen.

Eftersom studien enbart omfattade två patienter bör detta resultat tolkas med försiktighet då det inte ger ett tillräckligt underlag för att dra definitiva slutsatser. Det skulle emellertid vara av intresse att skala upp studien och applicera den på ett större antal patienter från datamängden för att potentiellt kunna få ett statistiskt signifikant resultat.

4.2 Begränsningar

Modelleringsprocessen och databehandlingen har krävt ett antal förenklingar av verkligheten. I detta avsnitt belyses begränsningarna i tillvägagångssättet. Förbättringsförslag till framtida studier kommer dessutom att diskuteras.

4.2.1 Biologiska begränsningar

Faktorer som rotar sig i de biologiska processerna som styr tillväxten hos hjärntumörer kan ha begränsat resultatet. Som nämnts i avsnitt 2.3.1 förekommer ofta blödningar i tumörens rand och dessa är hypointensiva T2-signaler. När intensiteten inverterades för att uppskatta celldensitet tolkas alla regioner av denna sort som områden av hög celldensitet. Det finns tydliga tunna randområden av hög celldensitet i tumörerna som avbildas i figur 6. Det är möjligt att dessa tätt packade randområden i själva verket är blödningar och inte områden av hög celldensitet.

Antagandet att T2-intensiteten är direkt negativt korrelerad med celldensitet är inte ett konventionellt antagande i litteraturen och bör uppdateras för att säkerställa att blödningar stör approximationen. I 2.3.1 nämns en strategi baserad på maskininlärning som hade varit intressant att utforska för framtida studier.

4.2.2 Begränsningar med datamängden

En begränsning var att LUMIERE-datamängden innehöll inkorrekt information. Datamängden inkluderade filer med information om patienterna och deras operationsveckor: RANO-filen specificerade operationstillfällen, men veckonumren stämde inte alltid överens med övrig data. Ett problem som ligger i RANO-filen refererar till ett antal veckor som inte finns med i datamängden. För somliga veckor saknades även segmenteringsfiler helt. Dessa inkonsekvenser och brister försvårade urvalet av lämpliga patienter för vidare analys, vilket ledde till att antalet distinkta sjukdomsförlopp som kunnat jämföras med simulering blev få. Modellens validitet försvagas av att den testats på få fall. De utvalda patienterna har dock tumörer väl lämpade för jämförelse, och vi ser möjligheter att extrahera fler relevanta intervall (se avsnitt 4.3.2).

Strålbehandling och cytostatika är standardbehandling för GBM och kan påverka tumörens utbredningsmönster. RANO-filen innehöll ingen information om hur patienterna behandlades så detta går inte att ta hänsyn till. Det är möjligt att en effektiv behandling med cytostatika eller strålbehandling kan försämra modellernas prestation från patient till patient. Dock finns det inget belägg för att modellernas prestation i relation till varandra skulle förändras under något av behandlingsalternativen.

4.3 Förbättringsområden

I detta avsnitt belyses förbättringsområden för modellen, databehandlingen samt anpassningsförfarandet. Fokus ligger på konkreta åtgärder som kan bidra till att öka modellens tillförlitlighet och generaliserbarhet i framtida studier.

4.3.1 Förbättringsområden för modellen

När den grundläggande LGCA-strukturen formulerades i 2.1.1 får cellerna en lika stor sannolikhet att dela sig i en diagonal riktning som att dela sig till de gränsande noderna, se figur 3. Dock motsvarar en diagonal förökning en $\sqrt{2}$ gånger större rumslig utbredning. Formuleringen ger modellen en predisposition för diagonalt orienterad spridning utan motivering. I framtida studier bör LGCA-modellerna formuleras så att cellerna endast kan dela sig till de gränsande noderna.

Vid framtida studier, bör man även säkerställa att vit substans elimineras från områden där man vet att tumören befinner sig. Både genom att se till att den initiala segmenteringen är fri från vit hjärnsubstans i tumörens område, och genom att kontinuerligt uppdatera segmenteringen och ta bort vit substans i områden dit den simulerade tumören växer. Dessa ändringar skulle kunna leda till att den simulerade cellmigrationen återger ett mer verklighetstroget spridningsmönster, och på så sätt ger ett högre Jaccard-index.

4.3.2 Förbättringsområden för databehandlingen

Ett förbättringsområde inom databehandlingen var urvalet av patienter. Den process som användes för att identifiera relevanta bildpar kunde ha gjorts mer effektiv genom en mer systematisk och automatiserad strategi. En mer omfattande initial granskning av hela LUMIERE-datasetet hade möjliggjort ett bredare urval av patienter med tydligt dokumenterad tumörtillväxt mellan bildtagningstillfällena. Om mer tid funnits till förfogande, skulle samtliga MR-bilder ha kunnat granskas noggrannare i syfte att identifiera fler representativa start- och slutbilder för simuleringen, inte bara de patienter som identifierades ha minst sex veckors intervall utan operation. Detta hade kunnat förbättra jämförbarheten med modellens resultat.

T2-intensiteten antogs vara direkt korrelerad med tumörens densitet, se 2.2.2. Exempelvis antas intensitetsvärden av 3/12 och 6/12 skiljas i densitet med 100% vilket inte är biologiskt rimligt. Överlag består tumörvävnaden till absolut störst del av vatten, proteiner och lipider och den genomsnittliga densiteten av GBM är därför, som beskrivet i 2.2.2, omkring 1,05 g/cm³ och varierar som störst med enstaka procent. Ett bättre antagande hade troligtvis varit att anta en homogen densitet för uträknandet av masscentrum.

4.3.3 Förbättringsområden för modellanpassningen

De främsta begränsningarna i parameterinferensen har varit tillgången till beräkningsresurser. Med ökade resurser hade det varit möjligt att genomföra SMC-ABC med fler generationer, större populationsstorlekar samt bredare *a priori*-fördelningar. Detta hade i sin tur förbättrat utforskningen av parameterutrymmet och minskat risken för att algoritmen konvergerar mot lokala optima snarare än det globala optimumet, om så finns. Eftersom SMC-ABC är en stokastisk metod för sannolikhetsbaserad uppskattning, är den känslig för initiala villkor, antalet generationer och populationsstorlekar som används för att successivt förfina *a posteriori*-fördelningen.

Dessutom behöver tidsstegen ej vara en fast parameter, även den kan anpassas med hjälp av ABC. Det hade varit givande att utforska då alla parametrar har en direkt påverkan av mängden tidssteg.

4.4 Samhälleliga och etiska aspekter

Då detta kandidatarbete använder patientdata är det av stor vikt att ta hänsyn till de samhälleliga och etiska aspekterna som användningen av denna medicinska bilddata medför. Det är viktigt att den data som används har samlats in på ett ansvarsfullt och etiskt sätt. Arbetet baseras på den öppet tillgängliga LUMIERE-datamängden, som är sammanställd av både forskare och experter inom neuroonkologi för att stödja utvecklingen av forskningen kring GBM. Materialet i datamängden innehåller även expertbedömningar enligt RANO-kriterierna, samt information om patienternas behandlingsförlopp och överlevnad. RANO-kriterierna är en etablerad uppsättning av riktlinjer som används för att standardisera bedömningen av behandlingssvar hos människor med hjärntumörer, framförallt GBM [11].

Vid användning av data från riktiga individer för forskningsbruk krävs ett tydligt etiskt ramverk för att skydda patientens integritet. Skaparna av LUMIERE-datasetet har tagit flera åtgärder för att säkerställa detta. Bland annat har MR-bilderna genomgått "skalleextraktion" (skull-stripping), detta för att avlägsna ansiktsdrag och huvudform för att förhindra någon form av identifiering av patienten. På detta har tidsinformationen för varje patient anonymiserats på så sätt att tidsenheten som används för samtliga undersökningar är antalet veckor efter operation i relation till det preoperativa tillfället. För att ytterligare skydda individens identitet tillhandahålls bildfilerna i standardiserat NIfTI-format utan någon form av personligt identifierbar metadata.

Utöver dataskyddsaspekterna är det av stor vikt att resultaten från matematisk modellering görs begripliga även för en bredare målgrupp utanför det matematiska fältet. För att modeller ska kunna ge faktisk insikt och stöd i medicinska beslut krävs att deras antaganden, begränsningar och osäkerheter kommuniceras tydligt. Detta ställer krav på forskare att inte bara utveckla tillförlitliga modeller, utan också att presentera dem på ett transparent och pedagogiskt vis. I förlängningen möjliggör detta utvecklingen av personaliserad medicin, där avancerade modeller kan användas för att simulera och optimera behandlingar på individnivå. Vi har av denna anledning försökt verka för att tillvägagångssätt och matematiska resonemang ska vara så tydliga som möjligt, och kunna greppas även av personer vars huvudsakliga fält inte är matematiken.

Kandidatarbetet bedöms ha en begränsad etisk påverkan, detta då det bygger på data från ett dataset där patienter givit sitt samtycke till att deras medicinska information används för forskningssammanhang. Utöver detta så har allt bildmaterial bedömts enligt RANO-kriterierna och funnits uppfylla dessa. Ett arbete på denna skalan anses inte heller medföra några etiska eller samhälleliga konsekvenser. Vid fortsatt forskning bör dock de samhälleliga och etiska aspekterna granskas mer ingående, detta då det skulle kunna medföra större påverkan på individnivå.

5 Slutsats

VS-modellen visade sig bättre reflektera glioblastomets invasionsmönster, jämfört med modellen som Tektonidis m.fl. har utvecklat för de två patienter som undersöktes. Resultatet indikerar att en liknande men uppskalad studie på LUMIERE-datamängden skulle vara mycket intressant. Om en sådan studie skulle visa att VS-modellen, även för resten av patienterna, statistiskt signifikant reflekterar tumörutbredningen bättre, skulle VS-modellen kunna utgöra en bas för framtida LGCA-modeller av GBM. Några intressanta justeringar för en sådan studie skulle vara att löpande ta bort invaderad vit substans från segmenteringsmasken, använda uniforma *a priori*-fördelningar, använda maskininlärning för att estimeras initial celldensitet, anta en homogen tumördensitet för att räkna ut masscentrum och enbart låta de proliferativa cellerna dela sig till noder som de gränsar till.

6 Användning av AI-verktyg

All text och kod som använts i detta arbete har författats av gruppens medlemmar. Under arbetets gång har emellertid det generativa AI-verktyget Chat GPT använts, men då som stöd i skriv- och programmeringsprocessen. Under skrivprocessen har generativa AI-verktyg huvudsakligen använts för att korrigera grammatiska misstag och stavfel, samt givit förslag för att omformulera eller förbättra språkliga delar av texten. Under programmeringsfasen användes AI som ett hjälpmedel

för att utveckla och felsöka koder, samt förklara implementering och användning av resurser i moduler som använts. Avsikten var att underlätta arbetet i moment där svårigheter bemöttes. Innehåll som genererats med hjälp av AI har redigerats och anpassats av författarna.

Referenser

- [1] M. Tektonidis, H. Hatzikirou, A. Chauvière, M. Simon, K. Schaller och A. Deutsch, “Identification of intrinsic in vitro cellular mechanisms for glioma invasion”, *Journal of theoretical biology*, årg. 287, s. 131–147, 2011.
- [2] S. Grochans, A. M. Cybulska, D. Simińska *m.fl.*, “Epidemiology of Glioblastoma Multiforme – Literature Review”, *Cancers*, årg. 14, nr 10, s. 2412, 2022. DOI: 10.3390/cancers14102412. [Online]. URL: <https://www.mdpi.com/2072-6694/14/10/2412>.
- [3] M. E. Davis, “Glioblastoma: overview of disease and treatment”, *Clinical journal of oncology nursing*, årg. 20, nr 5, S2, 2016.
- [4] A. M. Stein, T. Demuth, D. Mobley, M. Berens och L. M. Sander, “A Mathematical Model of Glioblastoma Tumor Spheroid Invasion in a Three-Dimensional In Vitro Experiment”, *Biophysical Journal*, årg. 92, nr 1, s. 356–365, 2007. DOI: 10.1529/biophysj.106.093468.
- [5] A. Swan, T. Hillen, J. C. Bowman och A. D. Murtha, “A patient-specific anisotropic diffusion model for brain tumour spread”, *Bulletin of Mathematical Biology*, årg. 80, nr 5, s. 1259–1291, 2018. DOI: 10.1007/s11538-017-0271-8.
- [6] A. A. Malik, K. C. Nguyen, J. T. Nardini, C. C. Krona, K. B. Flores och S. Nelander, “Mathematical Modeling of Multicellular Tumor Spheroids Quantifies Inter-Patient and Intra-Tumor Heterogeneity”, *npj Systems Biology and Applications*, årg. 10, s. 8, 2024. DOI: 10.1038/s41540-025-00492-3.
- [7] K. K. Treloar, M. J. Simpson och S. W. McCue, “Velocity-jump models with crowding effects”, *Physical Review E*, årg. 84, s. 061920, 2011. DOI: 10.1103/PhysRevE.84.061920.
- [8] K. Böttger, H. Hatzikirou, A. Chauviere och A. Deutsch, “Investigation of the Migration/Proliferation Dichotomy and its Impact on Avascular Glioma Invasion”, *Mathematical Modelling of Natural Phenomena*, årg. 7, nr 1, s. 105–135, 2012.
- [9] A. Salvalaggio, L. Pini, M. Gaiola *m.fl.*, “White Matter Tract Density Index Prediction Model of Overall Survival in Glioblastoma”, *JAMA Neurology*, årg. 80, nr 11, s. 1222–1231, nov. 2023, ISSN: 2168-6149. DOI: 10.1001/jamaneurol.2023.3284. [Online]. URL: <https://doi.org/10.1001/jamaneurol.2023.3284>.
- [10] F. Latini, M. Fahlström, A. Beháňová *m.fl.*, “The link between gliomas infiltration and white matter architecture investigated with electron microscopy and diffusion tensor imaging”, *NeuroImage: Clinical*, årg. 31, s. 102735, 2021.
- [11] Y. Suter, U. Knecht, W. Valenzuela *m.fl.*, “The LUMIERE dataset: Longitudinal Glioblastoma MRI with expert RANO evaluation”, *Scientific Data*, årg. 9, nr 1, s. 768, 2022, ISSN: 2052-4463. DOI: 10.1038/s41597-022-01881-7. [Online]. URL: <https://doi.org/10.1038/s41597-022-01881-7>.
- [12] A. L. Stensjøen, O. Solheim, K. A. Kvistad, A. K. Håberg, Ø. Salvesen och E. M. Berntsen, “Growth dynamics of untreated glioblastomas in vivo”, *Neuro-Oncology*, årg. 17, nr 10, s. 1402–1411, 2015. DOI: 10.1093/neuonc/nov029. [Online]. URL: <https://doi.org/10.1093/neuonc/nov029>.
- [13] P. Tofts, *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. John Wiley & Sons, 2003.
- [14] T.-T. Lwin, A. Yoneyama, A. Hara *m.fl.*, “Spontaneous brain tumor imaging of aged rat by crystal X-ray interferometer-based phase-contrast X-ray CT”, *Acta Radiologica Open*, årg. 5, nr 2, s. 2058460115626958, 2016.
- [15] FSL (FMRIB Software Library), *FAST - FMRIB’s Automated Segmentation Tool*, https://web.mit.edu/fsl_v5.0.10/fsl/doc/wiki/FAST.html, Accessed: 2025-05-05, 2012.
- [16] E. Gates, J. Weinberg, S. Prabhu *m.fl.*, “Estimating local cellular density in glioma using MR imaging data”, *American Journal of Neuroradiology*, årg. 42, nr 1, s. 102–108, 2021.
- [17] A. Pai, R. Shetty, B. Hodis och Y. S. Chowdhury, “Magnetic resonance imaging physics”, i *StatPearls [Internet]*, StatPearls Publishing, 2023.

- [18] J. Cramer, I. Ikuta och Y. Zhou, “How to Implement Clinical 7T MRI—Practical Considerations and Experience with Ultra-High-Field MRI”, *Bioengineering*, årg. 11, nr 12, s. 1228, 2024.
- [19] Q. Zhuang, H. Yang och Y. Mao, “The oncogenesis of glial cells in diffuse gliomas and clinical opportunities”, *Neuroscience Bulletin*, årg. 39, nr 3, s. 393–408, 2023.
- [20] J. Oh, S. Cha, A. H. Aiken *m.fl.*, “Quantitative apparent diffusion coefficients and T2 relaxation times in characterizing contrast enhancing brain tumors and regions of peritumoral edema”, *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, årg. 21, nr 6, s. 701–708, 2005.
- [21] E. Hattingen, A. Jurcoane, K. Daneshvar *m.fl.*, “Quantitative T2 mapping of recurrent glioblastoma under bevacizumab improves monitoring for non-enhancing tumor progression and predicts overall survival”, *Neuro-Oncology*, årg. 15, nr 10, s. 1395–1404, 2013.
- [22] C. Claussen, M. Laniado, W. Schörner *m.fl.*, “Gadolinium-DTPA in MR imaging of glioblastomas and intracranial metastases.”, *American journal of neuroradiology*, årg. 6, nr 5, s. 669–674, 1985.
- [23] A. Zimny, M. Neska-Matuszewska, J. Bladowska och M. J. Sasiadek, “Intracranial lesions with low signal intensity on T2-weighted MR images—review of pathologies”, *Polish journal of radiology*, årg. 80, s. 40, 2015.
- [24] N. K. R. Khurana, E. Raz, A. W. H. Mohamed *m.fl.*, “Intracranial cerebrovascular lesions on T2-weighted magnetic resonance imaging”, *Journal of Clinical Imaging Science*, årg. 14, s. 19, 2024.
- [25] O. Eidel, J.-O. Neumann, S. Burth *m.fl.*, “Automatic analysis of cellularity in glioblastoma and correlation with ADC using trajectory analysis and automatic nuclei counting”, *PloS one*, årg. 11, nr 7, e0160250, 2016.
- [26] X. Sun och P. D. Kaufman, “Ki-67: more than a proliferation marker”, *Chromosoma*, årg. 127, s. 175–186, 2018.
- [27] P. Jaccard, “Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines”, *Bulletin de la Société Vaudoise des Sciences Naturelles*, årg. 37, s. 241–272, 1901. DOI: 10.5169/seals-266440.
- [28] C. Forssén, *Bayesian Inference and Machine Learning*. Självpublicerad under CC-licens, 2023. [Online]. URL: <https://cforssen.gitlab.io/tif385-book/>.
- [29] Y. Schälte, E. Klinger, E. Alamoudi och J. Hasenauer, “pyABC: Efficient and robust easy-to-use approximate Bayesian computation”, *Journal of Open Source Software*, årg. 7, nr 74, s. 4304, 2022. DOI: 10.21105/joss.04304. [Online]. URL: <https://doi.org/10.21105/joss.04304>.

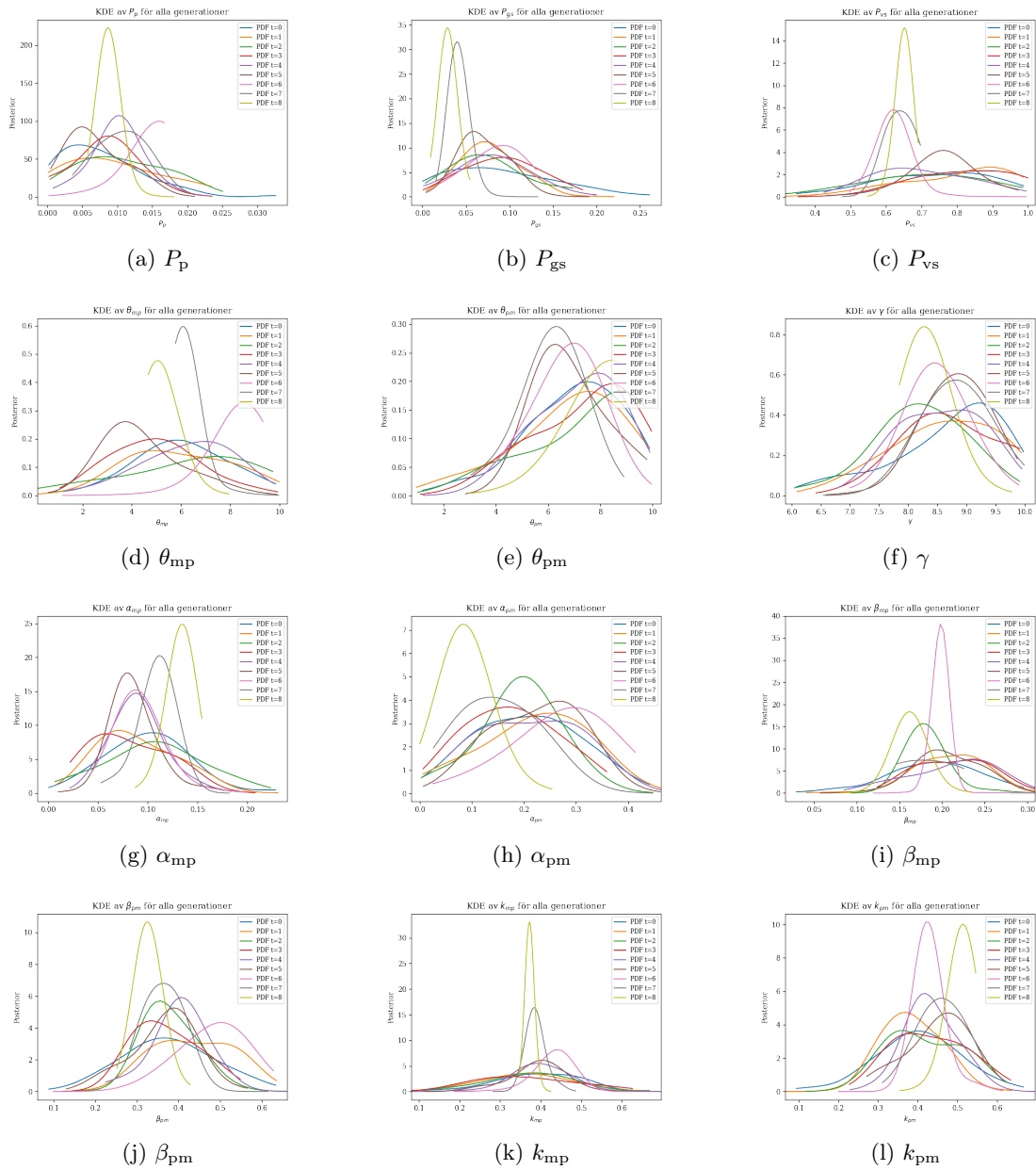
A Bilaga 1 - Källkod

All kod som använts för att utföra detta arbete, från simulering till databehandling och jämförelse finns att hitta på GitHub via <https://github.com/jarebrant/MVEX11-VT25-19>

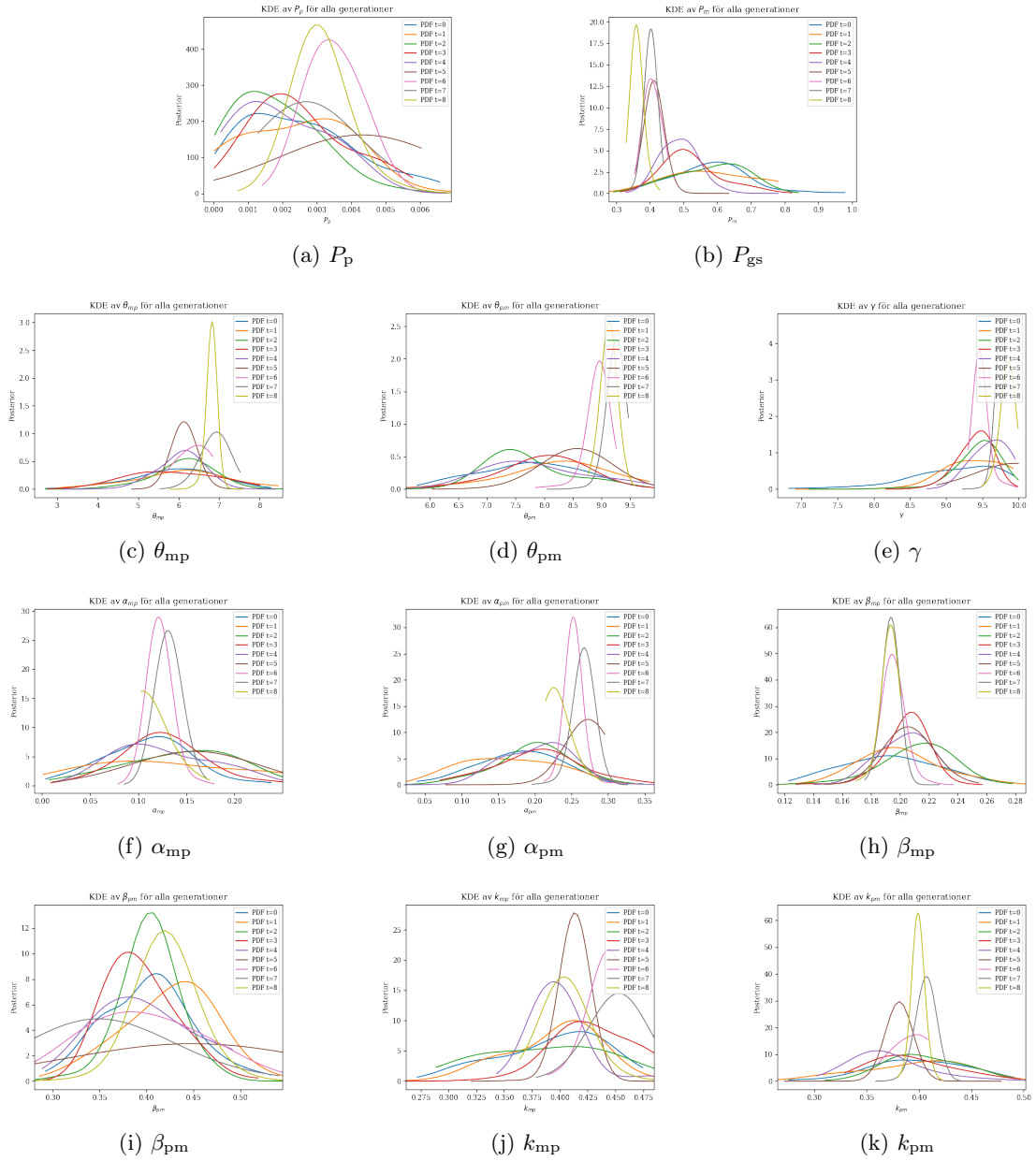
B Bilaga 2 - A posteriori-fördelningar

Nedan finns alla *a posteriori*-fördelningar för respektive parameter, modell och patient

B.1 Patient 85

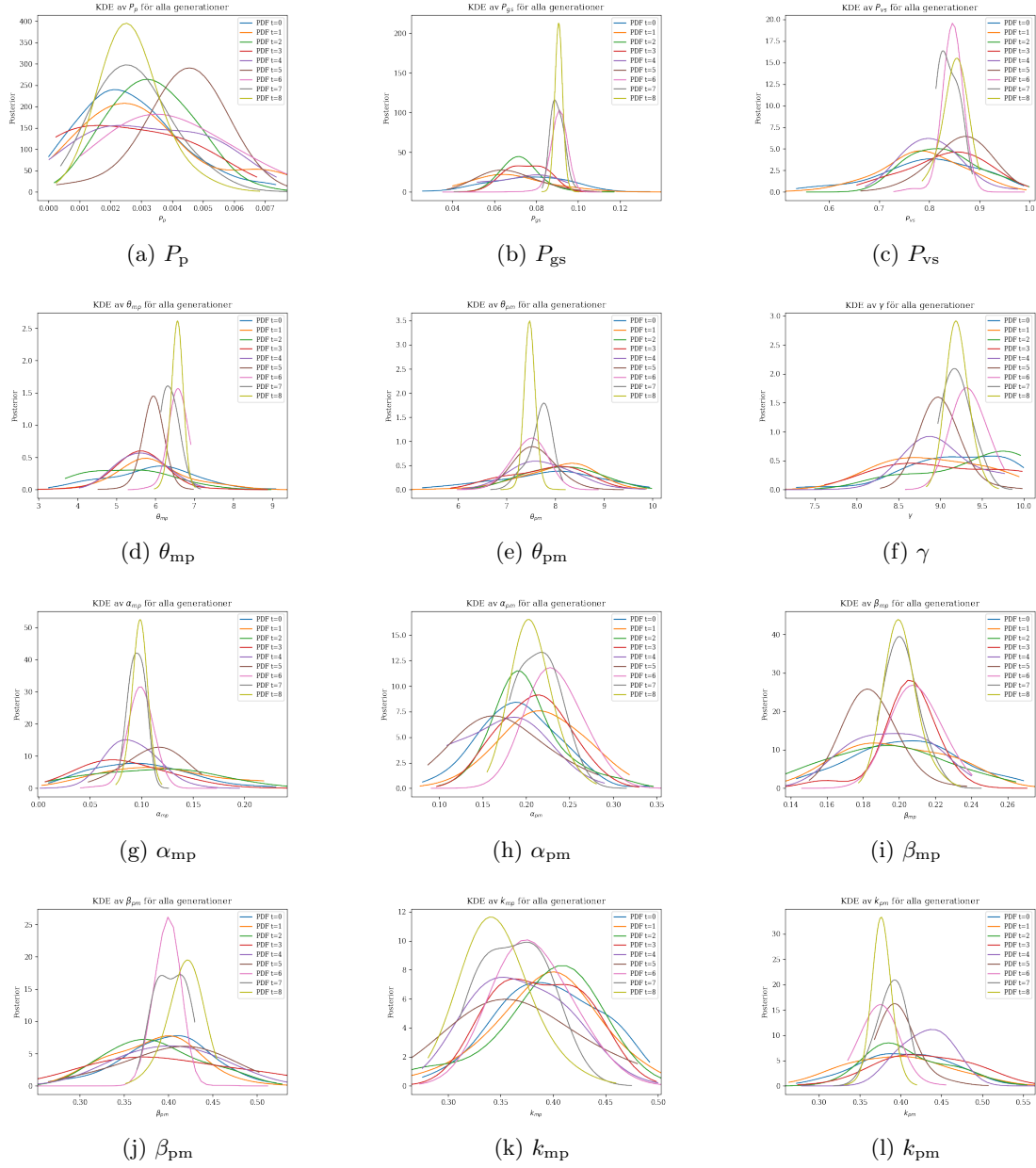


Figur 13: KDE:er för ingående parametrar i VS-modellen.

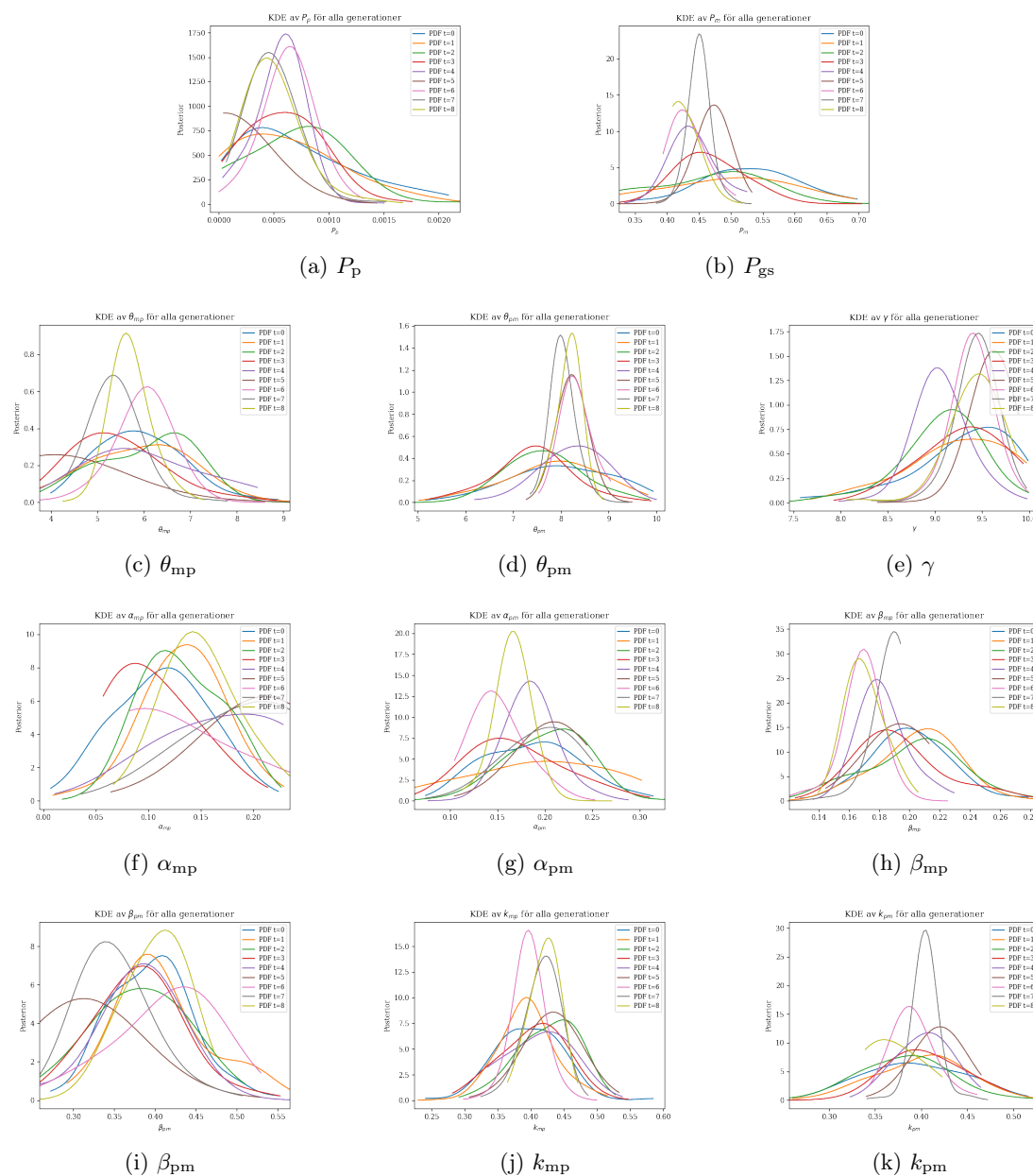


Figur 14: KDE:er för ingående parametrar i T-modellen.

B.2 Patient 42



Figur 15: KDE:er för ingående parametrar i VS-modellen.



Figur 16: KDE:er för ingående parametrar i T-modellen.