

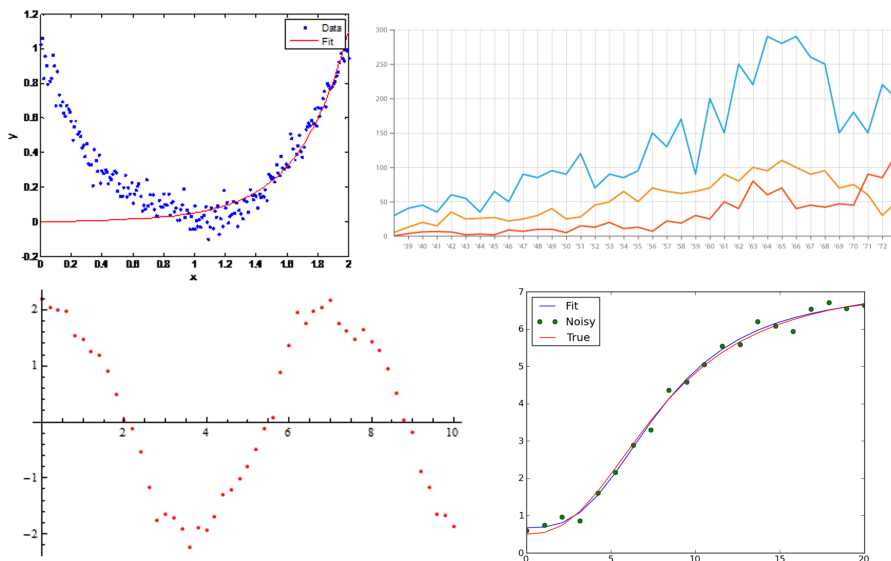
ROSKILDE UNIVERSITET

3.SEMESTERS PROJEKT

Bag om matematisk modellering

Forfattere:

Bjørn HANSEN	(62607)	Frederik GANTZEL	(62611)
Karl NORDENTOFT	(62547)	Kenny HASLUND	(63869)
Max THRANE	(62621)	Poul THRANE	(62562)



Projektvejledere:

Zamra	SAJID
Johnny	T. OTTESEN
Johanne	GUDMAND-HØYER



Naturvidenskabelig Bachelor

Roskilde Universitet

<http://www.ruc.dk>

Titel:

Bag om matematisk modellering

Tema:

Naturvidenskaben som kulturelt og samfundsmæssigt fænomen

Projektperiode:

Efterårssemestret 2018

Projektgruppe:

15

Forfattere:

Bjørn Dan Orfang Nielsen (62607)

Frederik Vedel Gantzel (62611)

Karl Skovfoged Nordentoft (62547)

Kenny Haslund Andersen (63869)

Max Thrane Nielsen (62621)

Poul Thrane (62562)

Vejledere:

Zamra Sajid

Johnny T. Ottesen

Johanne Gudmand-Høyer

Afleveringsdato:

18. december 2018

Sidetal: 95

Oplagstal: 1

Resumé:

Denne rapport vil undersøge hvor stort et omfang, de valg som tages i løbet af modelleringsprocessen, påvirker den endelige model, og på hvilke måder denne model bliver påvirket af hvert af disse valg.

Dette undersøges først teoretisk, ved at tage en masse mulige valg under matematisk modellering til diskussion af relevans for modellen, samt forskellige problematikker, som kunne have relevans at have i baghovedet. Løbende vurderes de forskellige emner, gennem kvalitativ analyse af fordele og ulemper, som senere kan refereres tilbage til.

Herefter undersøges problemstillingen gennem en mere praktisk vinkel, hvor en model case tages under mikroskopet. Her sammenlignes deres model med en model opstillet på baggrund af et enkelt anderledes valg, således at kun dette enkelte valg i modelleringsprocessen adskiller de to modeller. Derved tillades det at undersøge betydningen af lige netop dette valg i modelleringsprocessen, således at en diskussion af det optimale valg til netop denne model kan føres. Rapporten afslutter med at samle op på alle disse ting, og det diskuteres samlet, hvor stor en effekt de forskellige mulige valg kan have for modeller. Dette gøres ved at analysere hvor meget af den observerede påvirkning, som er specifikt til casen, og hvilke der er sande i en mere generel forstand.

Abstract

This paper was developed as a third semester project at Roskilde University. The purpose of the paper was to divide the formulation of mathematical models into steps and examine how the choices made in each of these steps affects the model. The problem statement behind the project is,

“What consequences do the contemplations and choices of the modeller have for the resulting model’s ability to describe a system?”.

The method chosen to answer this problem statement was to analyse a case model by Kuznetsov & Knott from the article *”Modeling Tumor Regrowth and Immunotherapy”*

based on theory for a range of different modelling tools. The analysis of this model was divided into three steps.

First we analysed the model based on the information available in the articles. This was done to evaluate the accuracy of the model and its parameters, as well as the articulation of this accuracy in the articles. It was established that most parameter values are fairly uncertain; although this is common for biological systems.

Afterwards we analysed the sensitivity of the parameters and repeated the parameter estimation. This resulted in a slightly better fit for the model and a confirmation that one of the parameter, g , is difficult to estimate accurately due to limitations of the dataset.

Lastly, we applied changes to the construction of the model on an individual basis and investigated the effects of these changes. It was found that the conditional statement could be replaced accurately by a logistic growth function to make the model continuous. It was also found that the system allowed for a certain variation in the effectiveness of the killer cells, although these variations did affect its required period for cancer cell destruction. Furthermore it was investigated whether the model could be simplified by perceiving one of the equations in a quasi-steady state. This simplification had a large effect on the system and was deemed unsuitable for most applications of the model.

We also modelled the system through difference equations for which case, we introduced a temporal displacement for one the equations. This was done

to match the description of the perceived system of the article. This change resulted in a massive difference of the model, where the populations no longer reached a steady state, but rather oscillated around it.

Overall the original model was found to be well-considered for most cases, although the articulation of the process behind the creation of the model was found to be somewhat lacking. Several of the applied modifications resulted in significant changes in the system. Thus it was seen that changing merely one perception of the system can consequently change the model significantly as well.

Indhold

1	Introduktion	1
1.0.1	Indledning	1
1.0.2	Problemformulering	2
1.0.3	Metode og afgrænsning	2
2	Teori	7
2.1	Modelleringsprocessen	7
2.1.1	Modelleringscirkelen	7
2.1.2	Mekanistisk modellering	10
2.1.3	Fænomenologisk modellering	11
2.1.4	Modellering i praksis	12
2.2	Modeltyper	14
2.2.1	Modellering med differentiaalligninger	14
2.2.2	Modellering med differensligninger	17
2.2.3	Modelkorrigerende parametre	17
2.3	Modelleringsværktøjer	20
2.3.1	Bedste fit	20
2.3.2	Parametersensitivitet	22
2.3.3	Bestemmelse af ækvilibria	23
3	Modelvalg	25
3.1	Modelanalyse	25
3.1.1	Præsentation af case	26
3.1.2	Kvalitativ analyse	31
3.2	Modelkalibrering	40
3.2.1	Sensitivitetsanalyse	40
3.2.2	Parameterestimering	44
3.3	Modelkorrigerig	55
3.3.1	Betinget erklæring	55
3.3.2	Stokastisk parameter	62
3.3.3	Kvasistabil tilstand	64
3.3.4	Differensligning	68
3.4	Delkonklusion	75

4	Diskussion	76
4.1	Case-modellen	76
4.1.1	Valg af parameterverdier	76
4.1.2	Valg af parameter- og modeltype	80
4.2	Modelleringsværktøjernes indflydelse	83
4.2.1	Analyseværktøjer	83
4.2.2	Parameterestimering	84
4.2.3	Teoretisk viden	85
4.3	Procesorienteret modellering	87
4.3.1	Formidling som middel til optimering	88
4.3.2	Formidling som middel til anvendelse	89
4.4	Foregrebet iværksætter	91
5	Konklusion	93
6	Litteraturliste	95
6.1	MATLAB	III

Forord

Denne rapport er skrevet af Bjørn, Frederik, Karl, Kenny, Max og Poul i efterårssemesteret 2018. Projektet omhandler procesforløbet under matematisk modellering. Rapporten er udarbejdet som et led i 3. semesters projektforsøg på Roskilde universitets NatBach uddannelse. Projektets formål har været at undersøge hvilke valg der foretages ved udfærdigelsen af matematiske modeller, hvilke overvejelser der ligger bag dem, samt hvilken betydning valgene får for modellen.

Rapporten er forfattet i ShareLaTeX, mens databehandling fortrinsvist er udført i Matlab R2017b. Grafer og matematiske figurer er ligeledes designet i Matlab, mens øvrige figurer er lavet i Paint; såfremt de er inspireret af andre kilder, noteres dette i figurtekst.

Der rettes tak til vores vejledere Johanne Gudmand-Høyer, Zamra Sajid og Johnny T. Ottesen for deres hjælp til udfærdigelse af projektet. Der rettes også tak til Rasmus Kristoffer Pedersen og Viggo Andreasen for bidrag til forståelse af koncepterne bag matematisk modellering. Der rettes yderligere tak til Johnny Ottesen for bidrag til projektets kildemateriale.

Rapporten henvender sig til universitetsstuderende med en interesse for matematik. Det er ikke nødvendigvis en forudsætning at studere matematik, men rapporten er henvendt til studerende som er vant til at beskæftige sig med, og inddrage, matematik, som del af deres studie. Der forudsættes samtidig en generel viden og interesse for naturvidenskab, samt forståelse af den naturvidenskabelige metode.

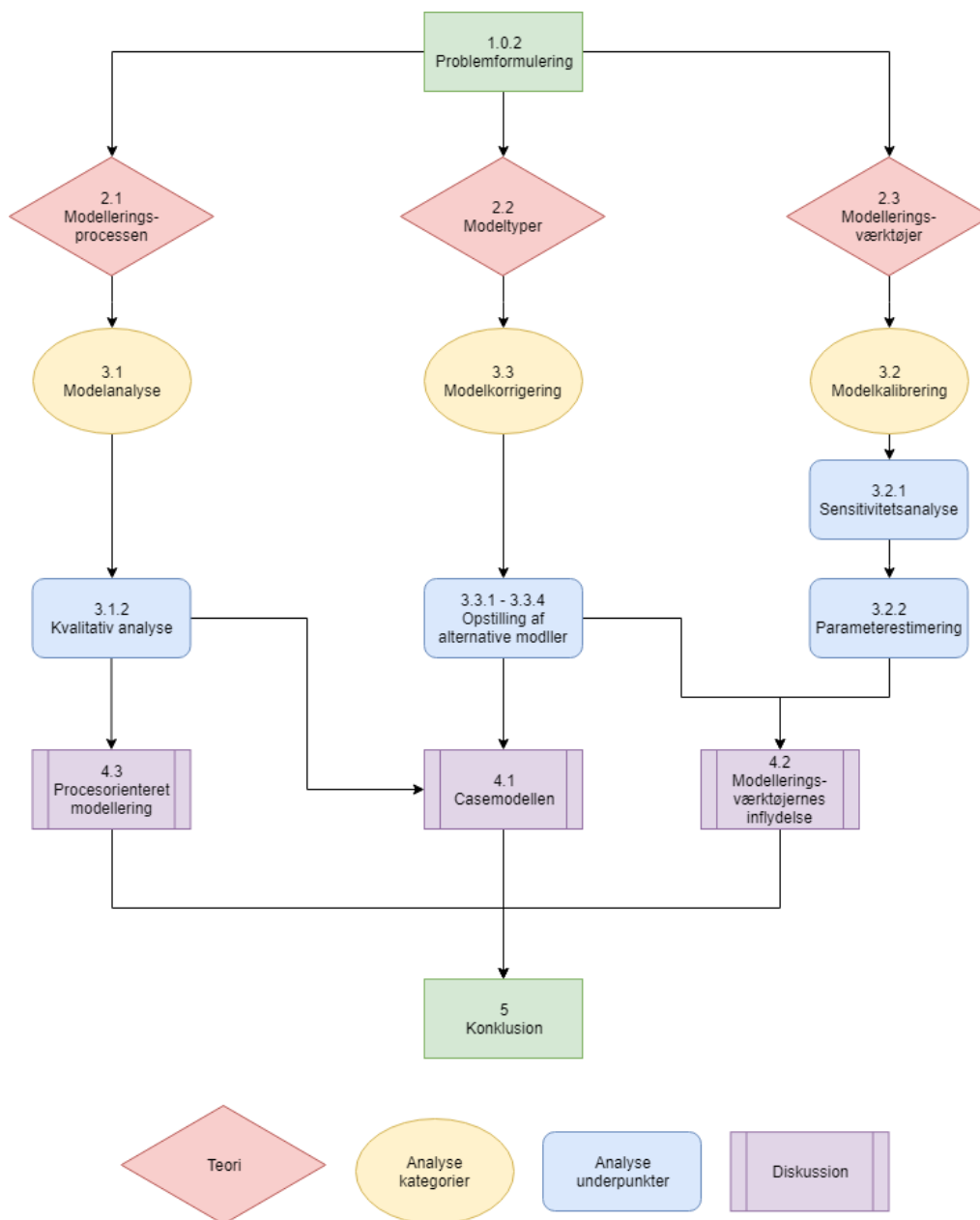
Læsevejledning

Henvisninger til figurer og tabeller er noteret således, at de er formateret som kapitelnummeret efterfulgt af nummeret på figuren i kapitlet; Eksempelvis Figur 4.3 eller Tabel 3.1. Ligninger er noteret på samme måde, blot med parentes omkring nummeret; Eksempelvis Ligning (4.2). Litteraturliste og kildehenvisning følger APA, hvorved henvisninger markeres som (Forfattere, Årstal). Notationen henviser til tilsvarende punkter i bibliografien, hvor fuld forfatterliste, titel, udgave, samt udgiver vil fremgå. Ved hjemmesider refereres der til webadresse, dato for sidste besøg, samt så vidt muligt forfatter, opretshaver og titel.

Afsnit og underafsnit indeholder metaafsnit, hvori det forklares hvad formålet med afsnittet er, og i nogle tilfælde opsummerer det som er fundet frem til i afsnittet. Disse afsnit forekommer i *skråskrift*, så de let kan adskilles fra den normale tekst. Rapportens struktur begynder med en teoretisk gennemgang af hvilke faktorer der kan påvirke en matematisk models kvalitet, egnethed og validitetsområde.

Herefter tages der udgangspunkt i en konkret model, som analyseres ud fra de pågældende faktorer. Til dette formål anvendes en model fra artiklen, *Modeling Tumor Regrowth and Immunotherapy* af Kuznetsov & Knott (Kuznetsov & Knott, 2001). Denne artikel vil gennem rapporten refereres til som **KK**. I forbindelse med analyse af modellen vil der inddrages en artikel af Kuznetsov, Perelson, Makalkin & Taylor, *Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis* (Kuznetsov, Makalkin et al., 1994), som vil forkortes til **KP**. Til sidst vil valgenes indflydelse på modellen diskuteres; både i forhold til matematisk og praktisk betydning.

Strukturen for rapporten er vist herunder på figur 2.



Figur 2: Figur som viser strukturen for rapporten. De forskellige dele af rapporten er vist med forskellige geometriske former, og farver. Nederst på figuren, er det vist hvordan de forskellige dele er markeret.

Som ses på figur 2, er rapporten struktureret således, at der først forekommer noget teoretisk viden, som derefter munder ud i en analyse. Denne analyse vil være omdrejningspunktet for rapporten, og vil også indebære

nogle umiddelbare diskussion af resultaterne. de dele som er markeret som "Diskussion" på figuren, er de sektioner hvor resultaterne fra analysen vil blive brugt, til at drage mere brede konklusioner om emnet.

1. Introduktion

1.0.1 Indledning

Naturvidenskab, og matematik i særdeleshed, har altid haft en unik rolle i academia qua dens metodiske opsøgen efter utvetydige svar. Da matematik samtidig er en meget universel disciplin, som let kan inkorporeres selv udenfor det naturvidenskabelige felt, anvendes det for eksempel ofte til at underbygge den mere fortolkende og spekulative tilgang, som ellers anvendes i samfundsvidenskaben (Noble, 1982).

Modellering fremstår i denne forbindelse som et af matematikkens mest alsidige værktøjer. Et værktøj, som kan bidrage massivt til både forståelse og beskrivelse af et system; såfremt det altså anvendes korrekt og velovervejet. Dette er især en udfordring indenfor de områder, som ikke normalvis anvender en naturvidenskabelig tilgang, men selv inden for det naturvidenskabelige felt kommer refleksioner omkring modellen ofte ikke fyldestgørende til udtryk. Dette kan for eksempel komme til udtryk ved at modeller vælges ud fra konvention eller personlig præference fremfor, hvor egnet modeltypen er til en given situation.

Under opstilling af en matematisk model, foretages der et virvar af valg, hvoraf mange har kvantificerbare konsekvenser for modellen. Nogle af disse valg tages på baggrund af forudgående viden om det system som den matematiske model skal afbilde, mens andre valg tages mere ubevidst på baggrund af tidligere erfaring med modellering. På baggrund af dette vil der i dette projekt foretages en undersøgelse af hvilke muligheder der er tilgængelige i processen bag matematisk modellering, samt hvilken effekt valgene der foretages mellem disse muligheder har. Samtidig vil det undersøges og diskuteres hvilke vilkår og kriterier der kan opsættes for hvad der udgør en ”god” matematisk model.

I projektet vil de forskellige beslutninger som tages i modelleringsprocessen blive undersøgt, med en af modellerne opstillet i: *Modeling Tumor Regrowth and Immunotherapy* af V.A Kuznetsov & G.D. Knott, som udgangspunkt. Mere præcist vil de værktøjer som er tilgængelige, og deres betydning

vil sættes under mikroskopet. Den pågældende model beskriver dynamikkerne mellem cancerceller og immunforsvar i et sygdomsforløb. Modellen har siden hen dannet udgangspunkt for forsøg på optimering af modellen (Ciancio & Quartarone, 2013). Samtidig er modellen oplagt til en analyse af de enkelte valg, da den opfylder sin funktion gennem blot to differentiallyigninger og dermed har tilstrækkeligt få parametre til at leddenes indflydelse på systemet kan betragtes enkeltvist.

Under analysen vil de steder i modelleringsprocessen, hvor KK har foretaget valg med størst betydning for den fremstillede model, fremhæves. Konsekvensen ved de foretagne valg vil analyseres med henblik på at undersøge hvordan det ville have påvirket modellen, hvis alternative valg var truffet. I den forbindelse vil det samtidig overvejes, hvorvidt alternative valg i visse tilfælde kan have været mere egnet til at opfylde modellens formål.

I modelleringsprocessen er der en række ”værktøjer” som kan tages i brug, til at opstille en model af det observerede fænomen. Her menes værktøjer at betegne alle de hjælpende redskaber, som analyser og algoritmer, samt matematiske ændringer af selve systemet. Undersøgelsen af modelleringsvalgene vil foretages med udgangspunkt i disse værktøjer og deres rolle for den resulterende model.

Motivationen for dette studie er dermed både en refleksionsproces over opstillingen af matematiske modeller, samt en vurdering af hvilken effekt en sådan refleksionsproces, under selve modelleringen, kan have for den endelige model.

Dette vil undersøges via følgende problemformulering:

1.0.2 Problemformulering

Hvilke konsekvenser kan modellørens overvejelser og valg af modelleringsværktøjer have for den fremstillede models evne til at beskrive et system?

1.0.3 Metode og afgrænsning

Vi har tænkt os at undersøge vores problemformulering via en systematisk gennemgang af forskellige valg og begrænsninger, som kan opstå under den matematiske modelleringsproces, samt den effekt de har på resultatet. Herefter vil disse overvejelser anvendes i samspil med et eksempel på en matematisk model fra en anerkendt videnskabelig artikel. Det vil i denne forbindelse

blive undersøgt både hvor stor og hvilken relevans refleksioner omkring processen har for det færdige produkt. Projektet vil have fokus på selve de overvejelser som ligger bag opstillingen af en matematisk model, og de værktøjer som kan bruges under denne proces. Alle undersøgelser som foretages, gøres med henblik på at kunne vurdere deres universelle egenskaber, således at den endelige konklusion vil være brugbart for enhver modellering, og ikke kun modeller som ligner casens i stor grad. Derved vil rapporten kunne bruges til evaluering af hvert værktøjs egenskaber. På denne måde opfyldes semesterbindingen, som lyder:

Refleksion over naturvidenskab og videnskabsteori

Semesterbindingen kræver, at der tages udgangspunkt i en specifik case, som i denne rapport vil udgøres af en opstillet model (Kuznetsov & Knott, 2001). For at afgrænse området som undersøges, vil der være fokus på modelleringsværktøjer som er relevante for ikke-lineære differentiaalligningsmodeller, som er ikke-lineære i både parametre og variable.

Ordforklaring

Algoritme

En proces som kan følges, til beregning og problemløsning

Approksimere

At tilnærme sig noget, eksempelvis værdi eller kurve

Betinget erklæring

En erklæring der beskriver at en aktion sker, såfremt at en eller flere betingelser er opfyldt, for eksempel *hvis* og *mens* erklæringer inden for programmering

Black-box modellering

En model som opfører sig på samme måde som et system, men som ikke inkorporerer noget andet viden om systemet, end datasættet

Deterministisk

Noget som er betinget af noget andets karakteristika

Diskret tid

Tid med bestemte værdier i mellem. Eksempelvis $t=1,2,3\dots 10$

Differentialligning

En ligning som beskriver hældningen af en funktion

Differensligning

En ligning som beskriver ændringen i et system over et diskret tidstrin, som funktion af den nuværende funktion. Bruges rekursivt til at beskrive en udvikling over en periode

Fit

et udtryk for korrelationen mellem teori og data

Fitting

Processen at tilpasse parameterstørrelser således at det bedste mulige fit opnås

Fænomenologisk modellering

Samme som *Black-box modellering*

Gauss-kurve

Et begreb for en normalfordeling

Grey-box-modellering

En modelleringsform, som både gør brug af praktisk data og teoretisk viden om emnet

Hyperbolsk

Hyperbolsk bruges i denne rapport, som udtryk om differentialligninger at de er løslige til alle tider

Konsekvens

Benyttes i denne rapport defineret som et resultat af en handling

Kontinuert tid

Tidlinje uden spring eller huller i tidslinjen

Kvasistabil

Pseudostabil. I konteksten her bruges det som et udtryk, hvor en lignings afledte er sat til 0, isoleret en parameter og substitueret ind i en anden ligning, for derefter at finde den lignings udvikling og steady state

Linearisering

At gøre noget ret med et 1. grads polynomie

Mekanistisk modellering

En model som er baseret på teoretisk viden om et system. Betyder det samme som *White-box modellering*

Modelleringsværktøjer

Hjælpende redskaber, som analyser og algoritmer, samt matematiske ændringer af selve systemet, som kan bruges i modelleringsprocessen

ode23s

'Ordinary differentiel equation 23 solver'. En funktion i Matlab

Operationel parameter

En styrende parameter

Parameterrum

Det rum som udspændes af parametrene og det interval af værdier de tillades at antage i modellen

Parameterestimeringen

En manøvre, hvor værdierne for parametre estimeres

Rekursiv

Gentagen påførsel af en procedure til efterfølgende mængder

Rekursiv funktion

En funktion som påføres på en startmængde, og rekursivt på resultatet af funktionen et antal gange

Residual

Afstanden mellem den observerede værdi, og den estimerede værdi

Robust

Ordet robust i denne rapport's kontekst betyder, at ved variationer af parameterværdien ændrer udfaldet meget begrænset (ingen ting)

Sensitivitetsanalyse

En analyse, som stiller parametre op mod hinanden og viser hvor sensitive de er over for hinanden

Shuffled Complex Evolution

En metode, som bruges i et analyseprogram

Stiff

Udtryk som bruges om ligninger, som ikke ændrer sig "for hurtigt". Matlab kræver stiff equations, når den skal plote ligninger

Stokastisk

Tilfældig afvigelse fra eksempelvis en parameter værdi

Søgealgoritme

Et program, som bruger en algoritme til at søge med

T-celler

Dræber celler for kroppens immunsystem

White-box modellering

En model som er baseret på teoretisk viden om et system

Ækvilibriumspunkt

Et punkt i eksempelvis en differentiaalligning, hvor den afledte af variabelen er lig 0

2. Teori

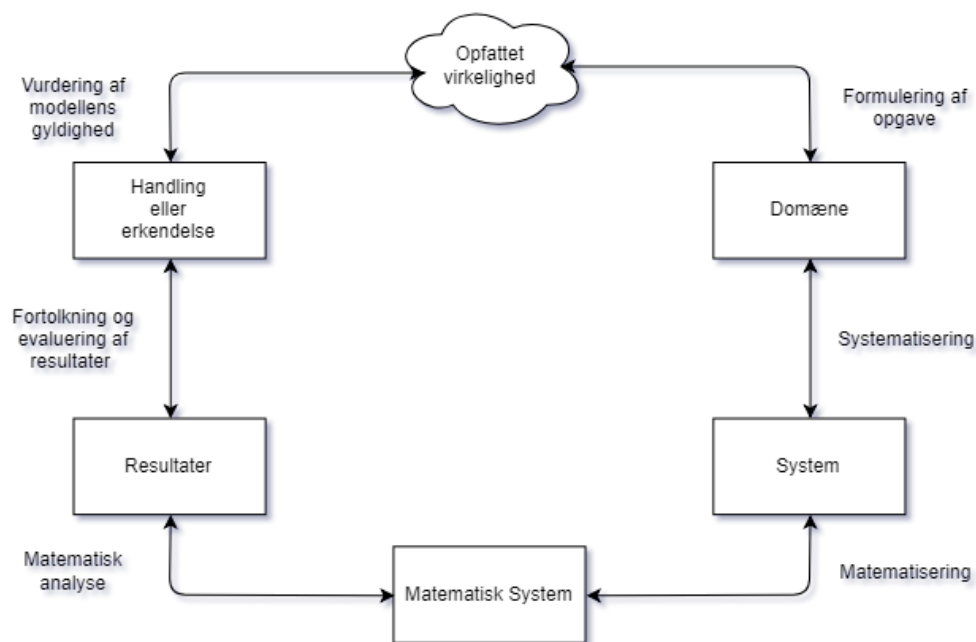
I denne rapport vil teorikapitlet tjene som en introduktion til de relevante begreber og metoder, som vil indgå i analysen af case-modellen. Dette indebærer de relevante begreber og metoder som indgår i modellen og som er nødvendige for at kunne forstå den bagvedliggende beslutningsproces. Sideløbende vil der blive introduceret nødvendig baggrundsteori for alternativer til de metoder som er benyttet i case-modelleringen, således at et grundlag for analysen kan dannes.

2.1 Modelleringsprocessen

I denne sektion vil vi forklare grundlæggende, processen bag opstilling af en matematisk model. Vi vil afklare, hvilken forståelse af matematisk modellering vi arbejder ud fra i denne opgave, samt forklare de forskellige trin der udgør modelleringsprocessen. Yderligere vil vi introducere to forskellige tilgange til modelleringen af et system: mekanistisk- og fænomenologisk modellering. Som afrunding på sektionen, vil vi uddybe hvordan disse to tilgange til modelleringsprocessen kommer til udtryk i praksis.

2.1.1 Modelleringscirkelen

Der eksisterer mange forskellige tilgange og perspektiver på processen bag matematisk modellering. Generelt betragtes modellering, som et middel til at systematisere et fænomen på en måde, som gør fænomenet mere tilgængeligt for forståelse og fortolkning (Noble, 1982). Der er dog delte meninger om, hvilke trin denne proces involverer, hvordan de vekselvirker, samt hvor tidligt i processen der kan siges at være opsat en matematisk model. I denne rapport tages udgangspunkt i en definition af modellering fra Blomhøj & Højgaard, 2003.



Figur 2.1: De forskellige stadier og processer som forekommer indenfor udformningen af en matematisk model. Figuren er inspireret af (Blomhøj & Højgaard, 2003)

Som ses på figur 2.1, består modelleringsprocessen af forskellige stadier for modellen (repræsenteret af boksene på figur 2.1), der er forbundet af pile. Pilene repræsenterer de handlinger der skal foretages, for at komme fra det forrige stadie, videre til næste stadie i modelleringsprocessen. Man kan se pilene som små processer i sig selv, som tager stadiet på den ene side som input, og giver stadiet på den anden side som output. Modelleringsprocessen starter med en *opfattet virkelighed*, og modellen bliver så videreudviklet, i urets retning, for hvert trin man går videre.

De forskellige trin i modelleringsprocessen lyder således:

- **Opfattet virkelighed**

Et observeret fænomen eller problem, som danner grundlaget for den forekommende modelleringsproces.

- ⇕ **Formulering af opgave**

Der opstilles et spørgsmål eller en opgave, som den matematiske model skal belyse eller udføre.

- **Domæne**

En præcis afgrænsning af hvilket aspekt af den opfattede virkelighed, der skal modelleres, for at besvare opgaven.

- ⇕ **Systematisering**

Der laves en definition af de relevante faktorer, der indgår i det system som modellen skal belyse.

- **System**

Et system der indeholder forskellige faktorer og elementer der påvirker hinanden på en bestemt måde.

- ⇕ **Matematisering**

De relevante variabler evalueres, og oversættes til matematiske værdier og termer.

- **Matematisk system**

En ligning, eller et system af ligninger, der matematisk beskriver de relevante aspekter ved det system der modelleres.

- ⇕ **Matematisk analyse**

Ved hjælp af det matematiske system, og andre matematiske metoder, opnås nogle matematiske resultater og konklusioner.

- **Resultater**

Resultater opnået ud fra matematisk behandling af det matematiske system.

- ⇕ **Fortolkning og evaluering af resultater**

Resultaterne fra modellen diskuteres, der drages konklusioner, og der bestemmes hvilken betydning resultaterne har for det område, der ønskes beskrevet med modellen.

- **Handling eller erkendelse**

Ny viden om emnet, som man har tilegnet sig ud fra resultaterne af den matematiske model.

- ⇕ **Vurdering af modellens gyldighed**

Den nye videns validitet undersøges ved at sammenligne med data og teoretisk viden fra virkeligheden, og man er altså tilbage til første trin.

Det kan ske at man løber ind i problemer med et af trinene i processen, for eksempel kan man få resultater der ikke stemmer overens med data, eller man kan tilegne sig ny viden om emnet, der modsiger den tidligere viden. I så fald

kan man gå ét eller flere trin baglæns i processen, for at gentage et bestemt trin, og tilpasse modellen. Dette resulterer herefter i, at alle følgende trin skal genovervejes med denne nye basis. Modelleringsprocessen er opsat som en cirkel, idet at man aldrig vil kunne nå frem til en model der er fuldentt til et punkt hvor ingen mulige forbedringer kan tænkes at forekomme. Dette skyldes at ny viden om emnet altid kan forekomme, hvorefter en ændring i modellen kan være nødvendig (Blomhøj & Højgaard, 2003).

2.1.2 Mekanistisk modellering

I en fuldstændigt mekanistisk model er den matematiske model en nøjagtig og fyldestgørende systematisering af den teoretiske viden som findes om emnet. I en sådan model er alle parametre og dynamikker udledt fra en dyb forudgående viden om det system som modelleres (DiStefano, 2013). Modellen opstilles på en sådan måde, at den teoretiske viden og den matematiske model, er ækvivalente. En sådan ækvivalens kan kun opstilles, såfremt at al information om systemet er kendt. Derved vil modellen ikke kunne afvises, uden også at afvise den teoretiske viden om emnet.

Det kan tænkes at alle trin i den matematiske modelleringsprocess, som leder op til stadiet **System**, allerede er foretaget under udviklingen af den teoretiske viden. Såfremt udviklingen af den teoretiske viden om emnet, ses som en separat process fra opstillingen af modellen, er selve modelleringsprocessen, som vist på figur 2.1, altså begrænset til **Matematisering** og frem.

Mekanistiske modeller er udbredt i fysik, hvori en stor teoretisk baggrundsviden kendes, hvilket kan præsenteres i en model af ækvivalent gyldighed. Her kan foreksempel ses på Isaac Newtons model af tyngdekraft (Newton, 1687):

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2} \quad (2.1)$$

Som beskriver, at tyngdekraften, F , mellem to legemer kan findes som den reciprokke kvadrerede distance mellem legemerne, $\frac{1}{r^2}$, ganget med masserne, m_1 og m_2 , og den gravitationelle konstant, G .

Denne model er funderet i teoretisk viden om, hvordan legemer tiltrækker hinanden. Skulle det påvises at denne model ikke er korrekt, ville det betyde at den teoretiske viden om dette fænomen ikke var korrekt, eller i hvert fald ikke var komplet.

2.1.3 Fænomenologisk modellering

Fænomenologisk modellering er den mekanistiske modellerings direkte modsætning. I kontrast til de mekanistiske modeller, bliver fænomenologiske modeller opstillet uden noget krav om at gøre brug af forudgående viden om systemet. I en sådan model tages kun den observerede udvikling i betragtning, hvorefter en model som opnår samme opførsel opstilles (DiStefano, 2013).

Dette kan opnås ved at tilpasse polynomier, normalfordelinger og lignende til det valgte systems opførsel. Denne modelleringsproces adskiller sig dermed også fra den, som blev præsenteret på figur 2.1. Her vil trinene mellem **domæne** og **matematisk system** reduceres til en analyse af den observerede opførsel, da den teoretiske viden om systemets opførsel ikke tages med i modellen.

Med udgangspunkt i samme fænomen som i forrige afsnit, kan der ses på tyngdekraft. En fænomenologisk model af tyngdekraften er:

$$F = 9.82 \frac{N}{kg} \cdot m \quad (2.2)$$

Hvor tyngdekraften F , findes som produktet af legemets vægt, m , og en konstant. Dette er den observerede tyngdekraft på jorden, og er en fænomenologisk model af tyngdekraft. Den anvendes til beregninger af tyngdekraft i situationer på jorden, og er en stærkt forsimplet version af ligning (2.1), som blev præsenteret i forrige afsnit.

Den store fordel ved denne type modellering er, at den nødvendige viden om emnet mindskes til et minimum. Da modellen ikke har til formål at forklare de observerede dynamikker i systemet, men blot at vise udviklingerne, så kræver modelleringsprocessen ikke nogen dyb teoretisk forståelse. En fænomenologisk model er et godt udgangspunkt for områder, hvor der er begrænset teoretisk viden. Modeller vil i sådanne tilfælde kunne danne udgangspunkt for at opstille teoretiske hypoteser, hvor modellens matematiske system tolkes som naturvidenskabeligt plausible mekanismer. Disse hypoteser kan herefter efterprøves, og eventuelt accepteres som teoretisk velbegrundede.

Fænomenologiske modeller opstilles sjældent inden for områder, hvori det er muligt at opstille en mekanistisk model. Den eneste pointe ved at opstille en fænomenologisk model i en sådan situation, er for at opnå en mindre kompleks model. I sådanne tilfælde vil der ofte tages udgangspunkt

i en mekanistisk model, hvor visse af de beskrevne mekanismer forsimples. Sådanne forsimplinger vil medføre at modellen bliver mere fænomenologisk.

2.1.4 Modellering i praksis

Med udgangspunkt i definitionerne for disse to modelleringstilgange, kan det praktiske aspekt af modellering tages til betragtning. I praksis er modellering ikke opdelt så sort og hvidt mellem mekanistiske og fænomenologiske modeller. Specielt fordi det oftest er tæt på umuligt at opnå en fuldstændigt mekanistisk model. For eksempel er ligning (2.1) ikke en perfekt mekanistisk model, da den ikke tager højde for generel relativitet (Charles W. Misner, 1973). Yderligere er mange systemer meget komplekse og kræver en grad af forsimpelse for at være anvendelige. I praksis er der derfor snarere tale om en hybrid mellem disse to tilgange, hvori det forsøges at udvide den kendte viden om systemet. Mekanistisk og fænomenologisk modellering betegnes også som henholdsvis *white box* og *black box* modellering. I denne terminologi kaldes denne hybrid mellem de to tilgange for *grey box* (Keesman, 2011).



Figur 2.2: Figuren viser en udvikling i forståelse af data og klarhed af model

Når man opstiller en matematisk model, er målet typisk at skabe en model der er så mekanistisk som muligt. Som det fremgår af figur 2.2, vil der inden for nye forskningsområder, hvorom der ikke findes noget kendt teori, begynde med en fuldstændig fænomenologisk model. Efterhånden som der tilegnes en større forståelse af de bagvedliggende mekanismer, vil modellen langsomt blive mere mekanistisk, da den nye teoretiske viden inkorporeres i den matematiske model. Sådan vil det fortsætte, indtil der er opnået en fuldstændig forståelse af området, hvorefter en ren mekanistisk model kan opstilles. Denne model kan herefter efterprøves, for at undersøge om den teoretiske viden om emnet er dækkende.

Inden for områder som biologi, kan det dog være problematisk at skulle opstille en helt mekanistisk model, da det kan føre til at modellen bliver så

kompleks, at den ikke vil være praktisk brugbar. I sådanne situationer kan det være mere brugbart at have en *grey box* model som endemål. Derved opnås en model som ikke er for kompleks, men som stadig inkorporerer en stor del af den teoretiske viden som findes. Modeller af denne type tager ofte udgangspunkt i hvordan det system som modelleres fungerer i virkeligheden, men er samtidig forsimplet ved at estimere parameterstørrelser gennem data (Keesman, 2011).

En god *grey box* model skal ideelt set kunne kombinere de virkelighedstro resultater, som findes gennem en mekanistisk model, og den simplistiske struktur som opnås gennem en fænomenologisk model. En simpel *grey box* model, som giver tilfredsstillende resultater, vil typisk være at foretrække frem for en meget kompliceret mekanistisk model, som giver perfekte resultater. Ved at simplificere en model kan dens generalitet ofte øges, hvormed modellen kan bruges mere fleksibelt til at beskrive flere systemer (Haefner, 2005).

2.2 Modeltyper

I denne sektion vil vi præsentere differentiaalligninger og differensligninger; to ligningstyper som kan anvendes til at matematisere et system. Vi vil, for hver af de to ligningstyper, komme ind på den grundlæggende matematiske struktur, hvilke specielle tilfælde der findes for ligningstypen, samt hvordan alt dette påvirker ligningstypens anvendelighed i praksis for forskellige typer af systemer. Derefter vil vi introducere parametertyperne stokastiske led, samt betingede erklæringer og undersøge hvordan disse parametertyper kan benyttes til at modificere ligningerne i en matematisk model.

2.2.1 Modellering med differentiaalligninger

Differentiaalligninger bliver brugt i stor stil i matematisk modellering. Differentiaalligninger beskriver hældningen til ethvert tidspunkt for en funktion som er ukendt. Differentiaalligninger benytter kontinuert tid, således at data for ethvert tidspunkt under forløbet er tilgængeligt gennem modellen (Haefner, 2005). Derved antager ligningen altid formen:

$$x' = f(x)$$

Hvor hældningen på et givent tidspunkt for en funktion x , beskrives ved en funktion $f(x)$. Da funktionen selv ikke kendes, men blot hvordan den ændrer sig på et givent tidspunkt, anvendes ofte computer til at numerisk approksimere funktionens udvikling, såfremt en startpopulation kendes. Dette er grundet i, at funktionen sjældent er mulig at udlede teoretisk, og det derved er nemmere blot at undersøge funktionens udvikling.

En af de største fordele ved differentiaalligninger er, at de bliver så hyppigt brugt. Inden for langt de fleste områder, vil det være muligt at finde andre opstillede differentiaalligningssystemer, som muliggør sammenligning og læring ved undersøgelse af disse modeller. Modellerne kan således tages i brug som skabeloner, for opstillingen af modellen, så andre beskrevne biologiske processer kan iagttages.

Til at beskrive et givent system tages dog sjældent kun en enkelt differentiaalligning i brug. Til dette bruges ofte differentiaalligningssystemer, hvor et antal differentiaalligninger forekommer samtidigt. Her kan det forekomme at hver differentiaalligning er en funktion af mere end én variabel, således at de to differentiaalligninger kan påvirke hinanden, og ikke blot køre sideløbende for hinanden. Dette ville for eksempel kunne se således ud for et system af

to koblede differentialligninger:

$$x' = f(x, y)$$

$$y' = g(x, y)$$

Hvor hældningen til et givent tidspunkt for funktionen x , er en funktion af både x og y , og ligeledes for funktionen y . Dette ville for eksempel være nyttigt til at beskrive et system, hvor to populationer, x og y , har indflydelse på hinandens vækst, gennem for eksempel konkurrence mellem populationerne i systemet. Det er dog ikke altid nødvendigt at alle funktionerne påvirker hinanden, differentialligningssystemer kan også udtrykke ensidig påvirkning mellem ligningerne.

En af ulemperne ved differentialligninger er, at de ikke altid kan løses analytisk (Haefner, 2005) og at større differentialligninger hurtigt kan blive uoverskuelige af denne grund. Selv i differentialligningssystemer med udelukkende lineære led, vil den enkelte vækst hurtigt opføre sig ikke-lineært, når antallet af ligninger stiger, da væksten påvirkes af en lineær kombination af øvrige variable og parametre. Derved kan det være svært at danne overblik over systemet, og danne nogle generelle regler for, hvordan systemet vil opføre sig under givne forhold.

Selvom nutidens computere kan løse differentialligninger på kort tid, kan der stødes på differentialligninger, som de fleste differentialligningsløserere ikke vil kunne løse, uden at bruge meget lang tid. Denne type differentialligninger kaldes *stiff* differentialligninger (Haefner, 2005), og fænomenet forekommer når en differentialligning udtrykker en så hurtig ændring i løsningen, at differentialligningsløseren bliver nødt til at tage meget korte tidstrin, og derved ender med at tage meget lang tid (Haefner, 2005), og kan i nogle situationer gå helt i stå. Dermed forekommer en usikkerhed ved denne modeltype, da man risikerer at skulle bruge en større mængde arbejdskraft på at komme rundt om dette problem.

Ikke-lineære og ikke-autonome differentialligningssystemer

Ikke-lineære differentialligningssystemer, er systemer hvori en af ligningerne, for eksempel x , opfører sig ikke-lineært. Dette ville forekomme, såfremt x er ganget med en anden funktion y , eller x er opløftet til en potens forskellig fra 0 og 1. I en sådan situation stiger leddets værdi ikke som en lineær funktion af x , og systemet siges derved at være ikke-lineært. Ikke-lineære differentialligningssystemer følger mange af de samme regler som andre differentialligningssystemer, for eksempel med hensyn til kontinuitet, men der

er også vigtige forskelle som der bør tages højde for.

For mange ikke-lineære ligningssystemer vil det for så vidt være muligt at finde en analytisk løsning, men løsningen vil i så fald være specifik for hver funktion, og det vil være for omfattende og uoverskueligt at have uendeligt mange løsningsmodeller. Man siger derfor at der i ikke-lineære ligningssystemer ikke findes en general analytisk løsning. Med dette menes, at det ikke altid er muligt at finde frem til den funktion x , som differentialligningen $x' = f(x)$ beskriver. Derfor løser man typisk de ikke-lineære ligninger numerisk, via computerprogrammer. (DiStefano, 2013).

Der findes tilfælde hvor det godt kan lade sig gøre at finde frem til en generel analytisk løsning, til et ikke-lineært ligningssystem, fx hvis ligningssystemet er lineære i parametrene. Fx kan man sige at funktionen $y = a \cdot x^2$ er en ikke-lineær funktion, der er lineær i parametrene, idet at parameteren a ikke er involveret i et ikke-lineært udtryk, selvom at variabelen x er. Har man med en sådan funktion at gøre, kan det godt lade sig gøre at finde en løsning analytisk, fx med ikke-lineær eller polynomisk regression (Haefner, 2005). Der vil dog ikke forklares yderligere om dette, da rapportens case er en ikke-lineær dynamisk model, som også er ikke-lineær i parametrene.

De ikke-lineære ligningssystemer er også unikke, i forhold til forudsigelighed. Sammenlignet med lineære ligningssystemer, vil udviklingen i et ikke-lineært ligningssystem, være meget mere afhængig af startbetingelserne. Varierer man derfor startbetingelser i et ligningssystem, vil det være tydeligt hvordan en lineær ligning opfører sig, mens det kan være mere komplekst for ikke-lineære ligninger (Hirsch et al., 2013).

En anden type differentialligning, hvori der ikke kan findes en generel analytisk løsning, er en ikke-autonom differentialligning. En differentialligning siges at være ikke-autonom hvis den afhænger eksplicit af tiden t ; det vil sige hvis t optræder på højre side af lighedstegnet. Et eksempel kan være:

$$x'(t) = f(t, x(t)) = a \cdot x(t) - \sin(\pi \cdot t)$$

Her kan det ses at $x'(t)$ afhænger eksplicit af t i leddet $\sin(\pi \cdot t)$. Selv hvis differentialligningen er lineær, vil man ikke kunne finde en generel analytisk løsning, idet at en ikke-autonom differentialligning ikke er delelig. Grundet dette bliver ikke-autonome differentialligninger ikke benyttet lige så ofte som autonome differentialligninger (Hirsch et al., 2013).

2.2.2 Modellering med differensligninger

Differensligninger gør brug af diskrete tidstrin, og beskriver hvorledes en population vil ændre sig, over dette tidstrin. I modsætning til differentialligninger, beskriver ligningen ikke ændringen i populationen, men i stedet selve populationens størrelse. I deres essens beskriver alle differensligninger en population således:

$$P_{t+1} = f(P_t) \quad (2.3)$$

Hvor antallet til næste tidspunkt, $t + 1$, findes som en funktion f af antallet ved det nuværende tidspunkt t .

Denne type ligningssystem kan uden større problemer løses numerisk uden maskinel hjælp, og denne egenskab gør selve ligningssystemet langt mere overskueligt. Derudover tillader differensligninger let adgang til diskrete tidsforskelle, eller andre diskrete værktøjer i modellen.

Ligesom en af de største fordele ved differentiaalligninger var, at de var brugt i så vidt et omfang, gør det modsatte sig gældende for differensligninger. Differensligninger bliver sjældent brugt, og derved er det tilgængelige materiale af modeller til sammenligning og inspiration langt mindre, end ved opstilling af differentiaalligningsmodel. Altså er denne type model ikke at foretrække, såfremt at der ønskes en model, som har størst potentiale for udbredelse.

En anden oplagt ulempe ved differensligninger er, at det ikke er muligt at opnå information om systemet ved andre tidspunkter, end lige netop de som er bestemt af det diskrete tidstrin. Altså er det muligt at vide hvor stor populationen er ved tidspunktet $t = 1$ og $t = 2$, men ikke ved tidspunktet $t = \frac{1}{2}$. Disse tidstrin kan være af vilkårlig størrelse; det vigtige er blot af de er ens for den enkelte model.

2.2.3 Modelkorrigerende parametre

Stokastiske led

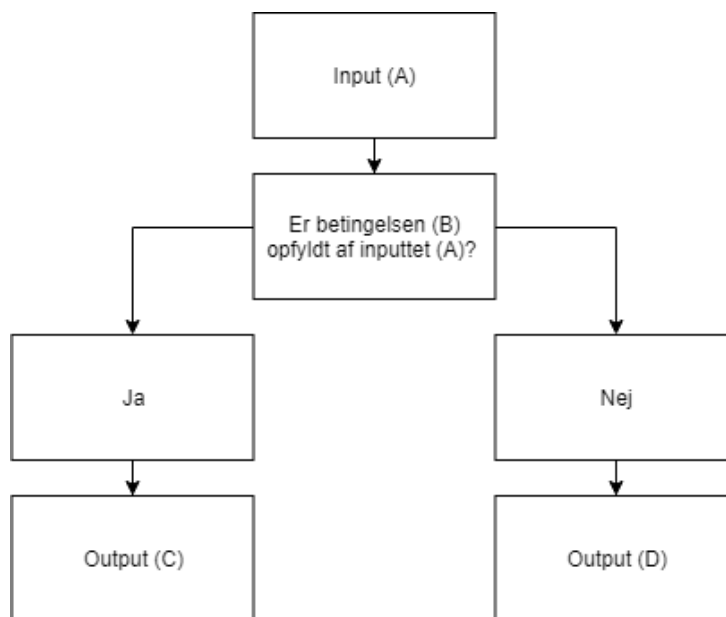
Når genstandsområdet for modelleringsprocessen ikke er et strengt kontrolleret miljø, men derimod et system udgjort af en samling af komplekse interaktioner, udfordrer dette en models præcision og korrekthed. En del af årsagen kan skyldes den tilfældige natur af den dynamiske udvikling i systemet, hvor data hyppigt er normalfordelt (Haefner, 2005). For at øge en models brugbarhed er det derved oplagt at give dens forudsigelser et stokastisk element,

således at den afspejler biologien mere realistisk.

I en differentiallyigningsmodel tages der hensyn til denne systemkarakteristik ved at variere parametrene eller variableerne stokastisk, ved at fitte til en Gauss-kurve (DiStefano, 2013). Det stokastiske led varieres i sandsynlighed fra en sandsynlighedsdistribution over mulige sandsynligheder for bestemte udfald. Dette gøres over gentagende simulationer, således at den sandsynlige respons fås gennem statistisk behandling, enten som tendens eller afvigelse i modellens output (Haefner, 2005).

Betingede erklæringer

En betinget erklæring er en sætning som beskriver, at en aktion sker såfremt at en betingelse er opfyldt. Sådanne ting er udbredt blandt programmering, i foreksempel *hvis* og *mens* erklæringer. For nemmere at forklare hvordan disse fungerer, er den grundlæggende funktion af en betinget erklæring vist på figur 2.3.



Figur 2.3: Figuren viser en betinget erklæring i dets mest basale form, med én betingelse. Det undersøger hvorvidt en betingelse er opfyldt, og skaber en reaktion efter hvorvidt det er opfyldt eller ej.

Som det ses på figur 2.3, fungerer betingede erklæringer således, at de kan undersøge hvorvidt et krav er opfyldt, og skabe en bestemt reaktion til dette.

Betingede erklæringer kan også involvere mere end én betingelse, således at der for eksempel undersøges for en række af betingelser, og skaber en reaktion specifikt til hver af dem.

Dette kan tages i brug i modellering til at beskrive en lang række fænomener. Det kan blandt andet bruges, til at definere et startkrav for et led, således at en proces først starter, når et startkrav er opfyldt.

2.3 Modelleringsværktøjer

I denne sektion vil vi belyse nogle af de matematiske værktøjer, der anvendes til at estimere parametrene i en matematisk model. Til dette formål vil vi diskutere begreberne bedste fit, og parametersensitivitet. Vi vil først definere disse begreber, hvorefter vi vil komme ind på i hvilke situationer det er relevant at kigge på begreberne, samt hvilke metoder man benytter til at undersøge dem. I denne forbindelse vil det også undersøges hvilken rolle overfitting har for disse processer. Afrundende vil vi diskutere ækvilibriumpunkter, samt hvad disse siger om et system, og hvordan disse punkter findes.

2.3.1 Bedste fit

Estimering af parametre er trin i matematiseringsprocessen der har til formål at bestemme modelparametrenes værdier, således at modeloutput er korreleret med data. Det er at foretrække at kunne bestemme modelparametre gennem et uafhængigt eksperiment, således at disse optræder som såkaldte operationelle parametre eller at finde dem gennem litteratur. Ofte er dette dog ikke muligt og det er derfor nødvendigt at estimere de ukendte parametre kvantitativt med matematiske værktøjer. Det er herunder begrebet *bedste fit* træder ind i billedet.

Der findes forskellige optimeringskriterier som kan anvendes, og valget af metode afhænger af flere faktorer. Disse faktorer inkluderer maskinel kraft, software til rådighed, hvilket system man modellerer og hvilken model der er valgt til at beskrive dette (DiStefano, 2013). I de fleste koncepter til estimering af parametre forsøges det at finde den kortest mulige afstand mellem parametrene og modeloutput, summeret over alle datapunkter, således at residual afvigelsen er mindst muligt. En residual er forskellen mellem en models output og den afhængige variabel. Der er flere teoretiske måder at beregne denne afstand på, og af dem er the least square kriteriet, *LS*, oftest anvendt (Keesman, 2011). Denne metode kvadrerer afstanden mellem punkterne, og større afstande vægtes derfor højere. For eksempel vil én afvigelse på 2, resultere i et dårligere fit, end to afvigelser på 1.

Der findes dog adskillige metoder der kan bestemme hvor godt en model, med nogle fastsatte parametre, afspejler et datasæt. Hvilken metode der er mest egnet til at estimeringen, afhænger af hvor interessen for modellen ligger. Least square tages ofte i brug, da den vægter negative afvigelser i lige så stor grad som positive, og vægter store afvigelser højest. I nogle situationer er det

dog ikke i modellens interesse at vægte afvigelserne forskelligt, hvilket ville resultere i et andet krav, som ikke kvadrerer afstanden.

Overfitting

Overfitting er et begreb som betegner en model, hvori analysen som den bygger på (eksempelvis parameterestimeringen), læner sig for nært op ad et datasæt, således at modellens evne til at beskrive det generelle fænomen forringes. Derved opnås en model som kun er lokalt beskrivende (Küppers et al., 2006), og som ikke kan beskrive andre datasæt over samme fænomen, hvilket ville gøre det globalt gældende. Dette kan hænde i situationer hvor modellens bedste fit til datasættet ikke formår at beskrive fænomenets generelle opførsel, hvorved fittet ikke vil gælde universelt for andre datasæt. I mekanistisk modellering er målet for modelleringsprocessen, at opstille en model som kan beskrive de underliggende mekanismer, og derved ville kunne beskrive ethvert datasæt af samme fænomen. Såfremt dette er tilfældet, ville en mekanistisk modellering nemlig kunne agere som en forklaring på fænomenet som er observeret og ikke blot en beskrivelse af dets udvikling (Küppers et al., 2006). Derved ødelægger overfitting pointen med modelleringen, og forsøges altid at blive undgået.

Under parameterestimering, er overfitting altid en risiko. Man kan aldrig helt undgå muligheden for overfitting, men blot reducere sandsynligheden for det. Overfitting skal betragtes analogt til lineær regression af et polynomie af meget høj grad (Nannen, 2003). Såfremt at et polynomie af $(n - 1)$ 'te grad, hvor n er antallet af datapunkter, tages i brug til beskrivelse af et datasæt, vil et perfekt fit af polynomiet til datasættet være muligt at finde. Som funktion af et reduceret antal polynomiegrader, vil fittet forværres. Samme sammenhæng findes mellem antallet af parametre som varieres i en parameterestimering og antallet af datapunkter. Jo flere parametre der introduceres i modellen, jo bedre et fit vil det være muligt at finde i parameterestimeringen (Nannen, 2003).

En model som beskriver et datasæt godt, udelukkende på grund af et stort antal parametre, er ikke en universelt anvendelige model. Modellen skal ideelt set i stedet beskrive dataen godt på baggrund af en god repræsentation af de bagvedliggende mekanismer. Dette er den evige problematik med overfitting, da det aldrig er muligt at bestemme hvorvidt det fit som er opnået er et resultat af korrekt fortolkede mekanismer i modelleringen eller om nogle af de fundne parameterværdier blot tillader modellen at imitere dette datasæts opførsel. For at bestemme hvorvidt modellen er overfittet eller

ej, skal modellen sammenlignes med et datasæt som er adskilt fra det, som parameterestimeringen er udført over (Nannen, 2003). Derved vil sandsynligheden for at modellen er globalt gældende stige, som antallet af datasæt som den nøjagtigt kan beskrive, stiger.

Derved er det optimalt, at reducere antallet af parametre som varieres i parameterestimeringen, da det både forsimples modellen, men samtidigt reducerer sandsynligheden for at overfitte modellen til datasættet. Derudover vil en reduktion i antallet af parametre i parameterestimering også resultere i et mindre parameterrum, som skal undersøges. I parameterestimering kan hver parameter som skal varieres, ses som en ekstra dimension som skal undersøges. Foretages en parameterestimering af én enkelt parameter, vil alle mulige værdier som kan undersøges ligge på en linje. Undersøges to parametre vil løsningerne ligge på et plan, og ved tre vil der være tale om et tredimensionelt rum. Efter tre parametre bliver parameterrummet vanskeligt at visualisere, men området som skal undersøges stiger på en lignende måde. Reduceres parameterrumets størrelse, formindskes området som skal undersøges, og algoritmen som foretager parameterestimeringen vil derved have større sandsynlighed for at finde det globale minimum for optimeringen. Der er altså stort incitament, for at reducere antallet af parametre som undersøges, eller det interval de undersøges i.

For at undersøge hvilke parametre det ville være meningsfyldt at betragte som konstanter i parameterestimering, kan en undersøgelse af parametrenes sensitivitet udføres.

2.3.2 Parametersensitivitet

Parametersensitivitet kan undersøges gennem en såkaldt sensitivitetsanalyse. En sådan analyse kan anvendes til at undersøge effekten på modellens output, som en variation af hver parameter vil have. Denne analyse begrundes med to formål (Haefner, 2005);

1. at finde frem til hvor følsom modellens respons er
2. størrelsen af spændet af modellens respons på parametreændringerne

Analysen er en naturlig del af modelleringsprocessen, som parameterestimeringen bygger videre på, da resultatet kan anvendes til at reducere sandsynligheden for overfitting. Derudover undersøges det også hvorvidt modellen har en entydig løsning, som følge af parametre som påvirker modeloutput.

Basalt set er kernen ved en parametersensitivitetsanalyse et numeriske eksperiment hvor der ændres på parametreværdierne (Δp) således at den samlede effekt (Δy) af dette kan analyseres.

Det er muligt at undersøge denne effekt ud fra ændringen i forskellige kvantitative størrelser: variablerne, værdier som modellen kan beregne, eller modellens output. Omfanget af dette afhænger proportionalt med antallet af valgte værdier som parametrene ændres til, samt det antal af parametre der undersøges (Haefner, 2005). Hvis der ved en lille ændring i en parameterværdi forekommer usandsynligt store ændringer i modellens output, kan dette være et tegn på at der er problemer med den overordnede model, da det skaber usikkerhed (Haefner, 2005).

Hvis en ændring i en parameter Δp forårsager en meget lille eller ingen ændring i model variablerne, så betegnes denne parameter som *robust*. Hvis Δp forårsager væsentlige ændringer i den variable respons, at parameteren er *sensitiv*. Er en parameter robust kan denne være svær at identificere ud fra et datasæt, og det kan derfor vise sig nødvendigt at erstatte denne med en konstant i stedet. Dette er også med til at simplificere modellen ved at reducere parameterrummet og kan gøre dens anvendelse mere universel.

Det kræver mere regnekraft at processere et stort område af parametre og derfor er det oftest at foretrække at ændre hver af parametrene enkeltvist. Dette har dog den ulempe, at viden om parametrenes interaktioner går tabt, hvilket er ofte er u hensigtsmæssigt for ikke-lineære differentiaalligninger (Haefner, 2005). Afvigelsen måles med en simpel metode hvor S angiver ratioen af respons i modeloutput relativt til det nominelle output.

Der er flere metoder end denne, og i analysen vælges det at sammenligne de to størrelser i grafernes toppunkt fremfor den gennemsnitlige værdi.

$$S = \frac{n}{n_0} \tag{2.4}$$

n_0 er oprindelig model respons i toppunkt, og n er model respons i toppunktet for grafen hvor én parameter er ændret procentvist.

2.3.3 Bestemmelse af ækvilibria

I et ækvilibriumspunkt er løsningen til en differentiaalligning, eller et system af differentiaalligninger, konstant og er at betragte som en af de vigtigste løsninger for systemet (Hirsch et al., 2013). Viden om et systems opførsel

omkring dets ækvilibrium afslører, hvordan systemet udvikler sig når det bevæger sig igennem punktet. Dette er vigtig viden, da det er meget følsomt overfor ændringer i et ækvilibriumpunkt; ofte med signifikante biologiske konsekvenser (Hirsch et al., 2013).

I et ækvilibriumpunkt er $x'(t) = 0$, således at $x(t)$ forbliver det samme, uagtet udviklingen i tid. Omkring ækvilibriumpunktet vil løsningen bevæge sig imod eller væk fra punktet. Tilfælde hvor løsningen nærmer sig punktet, kaldes *sinks*, og den modsatte situation, kaldes *sources*. Det første tilfælde kaldes et stabilt ækvilibrium, og det andet et ustabil (Hirsch et al., 2013). Den metodiske fremgangsmåde for ikke-lineære systemer er en linearisering i ækvilibriumpunkterne, således, at der opnås en tilnærmet forståelse af systemopførelsen omkring punktet. Derefter kan udviklingen analyseres for forskellige start- og parameterværdier, som giver information om hvordan de bevæger sig imod eller væk fra ækvilibriumpunkter. Når der modelleres med ikke-lineære modeller er deres løsninger kun numerisk løselige for bestemte startværdier, inputs og parameterværdier. Som konsekvens er det ikke muligt at finde alle de eksakte løsninger for systemet da sådanne systemtyper ofte er af kaotisk karakter.

3. Modelvalg

I dette kapitel vil der foretages en analyse med udgangspunkt i en konkret model. Der anvendes en model over populationsudviklingen af henholdsvis cancer- og T-celler fra KK. Vi vil foretage en analyse af modelleringsprocessen bag KK's model, og til dette formål vil vi fokusere på systematiserings-, og matematiseringsprocessen samt den matematiske analyse, som beskrevet i figur 2.1. Vi vil altså hovedsageligt fokusere på en analyse af hvordan KK har formuleret de specifikke mekanismer og parametre i modellen, og hvordan der er opnået matematiske værdier for disse parametre.

Analysen er struktureret således, at der først fokuseres på selve modellen, hvorefter modelleringsprocessen vil blive analyseret med henblik på, hvilke valg der er truffet, samt hvilke alternativer der findes dertil. Under analysen af selve modellen, vil data samt model først introduceres samt forklares. Herefter vil de involverede parametre, og deres sammensætning, analyseres kvalitativt, i forhold til i hvilken grad de kan betragtes som mekanistiske.

Herefter analyseres modellens parametersensitivitet, hvorefter resultaterne af denne vil blive inddraget i en ny parameterestimering. Med baggrund i både sensitivitetsanalysen og parameterestimering vil det blandt andet undersøges, hvordan og hvor meget de forskellige parametre påvirker modellens opførsel. Efterfølgende vil modellen konverteres til en differensligning, som mere meningsfuldt kan håndtere den tidsforskydning, som modellen indeholder.

3.1 Modelanalyse

I denne sektion beskæftiger vi os med case-modellen, og sektionen har derved til formål at gøre læseren bekendt med selve modellen, samt alt det som den er bygget på. først vil vi introducere selve artiklen af KK, således at case-modellens formål er klart. Efterfølgende dette, sætter vi fokus på datasættet som KK fitter modellen til, da den ikke blot spiller en stor rolle for KK's model, men også vil blive brugt løbende gennem rapporten i fits.

Med disse områder inddækket, afrunder vi afsnittet ved at introducere selve case-modellen. Differentialligningssystemet introduceres, med en tilhørende forklaring af parametrene størrelse og betydning for systemet. Til sidst i afsnittet præsenterer vi case-modellens fit til det præsenterede datasæt, da alt nødvendig information på dette tidspunkt er givet.

3.1.1 Præsentation af case

Vi vil præsentere det system, som KK modellerer, samt selve differentialligningerne som de opstillede under nævnte modellering. Dette indebærer en undersøgelse og forklaring af selve modellen med fokus på parametrene betydning for systemet, samt hvad de repræsenterer i den situation som skal modelleres. Først introduceres læseren til artiklen skrevet af KK for at give et indblik i modellens funktion og det fænomen den skulle beskrive. Herefter introduceres læseren til datasættet som modellen blev sammenlignet med og som er brugt i KK's artikel. Til sidst præsenteres læseren for selve modellen, med kendskab til datasættet, samt hvad KK ville modellere, i mente.

Præsentation af artikel

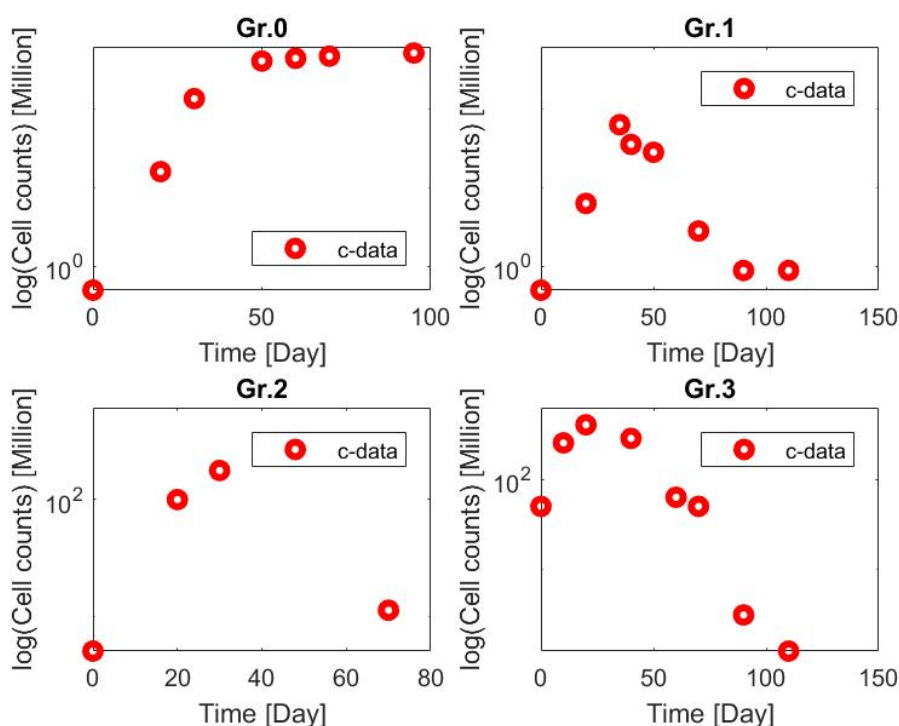
Da KK udgav deres artikel i 2001 var målet at lave en matematisk model som beskrev hvordan cancerceller udvikler sig i en mus, samt musens T-cellers reaktion på angreb, og hvordan de reagerer med hinanden. Der blev særligt fokuseret på et fænomen der hedder *cancer dormancy*. Cancer dormancy er hvor cancer tumoren hverken vokser eller bliver mindre. Dette sker fordi cancer- og T-cellerne, når et ækvilibrium hvor ingen af populationerne stiger eller falder. De ville også modellere hvad der sker når en tumor ikke længere er i cancer dormancy stadiet. En tumor kan begynde at gro igen af mange grunde som alle relaterer til et svækket immunforsvar. De simulerede hvordan musens immunforsvar bliver svækket med alderen ved at reducere visse parametre som tiden gik. De kunne på den måde vise sammenhængen mellem redueringen af parameterene som relaterer til immunforsvaret og hvor lang tid tumoren befinder sig i cancer dormancy stadiet.

Den case, som vil blive analyseret i denne rapport vil være den model, som er anvendt som basismodel i KK. Det vil sige den model, som modellerer cancercellernes vækst ved et velfungerende immunforsvar, hvor cancercellerne rammer, og bliver i, dormancy. Dataen som de lavede modellen ud fra, fik de fra en anden artikel (Siu & Vitetta, 1986) og vil præsenteres i næste underafsnit.

Præsentation af data

I artiklen af KK, brugte de et datasæt fremstillet af Ellen S. Vitetta og Henry Siu (Siu & Vitetta, 1986), som de præsenterede i artiklen "*Tumor dormancy. I. Regression of BCL1 tumor and induction of a dormant tumor state in mice chimeric at the major histocompatibility complex*".

Dette datasæt er vist i figur 3.1.



Figur 3.1: Figuren viser en gennemsnitslig udvikling af cancerceller i mus, med forskellige mængder af indsprøjtede cancerceller ved tidspunktet $t=0$. Gr. 0 er en kontrolgruppe hvori musene ikke har noget immunforsvar, Gr. 0 og Gr. 1 modtog en startkoncentration på 0,5 millioner cancerceller, Gr. 2 fik 5 millioner og Gr. 3 fik 50 millioner. Figuren er lavet gennem appendix A.

Datasættet som analyseres anerkendes som at være et velegnet datasæt i forhold til modellering. Da antallet af cancerceller der indsprøjtes i musene varierer i tre ud af fire grafer, beskytter det en model der beskriver disse situationer beskyttelse mod overfitting. Kontrolgrafene der viser vækstraten af cancer i en mus uden immunforsvar, giver en god case for parameterne k og b . Disse faktorer gør at dette datasæt er nemt at modellere på. Yderligere kan det ses via antallet af citationer som KK's artikel har, at datasættet er

velrespekteret og dermed må antages at være indsamlet under velkontrollede forsøgsforhold.

Præsentation af model

I KK's ligninger, har de en række parametre som i dette underafsnit vil forklares i forhold til hvilken rolle de har i modellen, samt hvordan disse relaterer til systemet. De to differentielle ligninger som udgør modellen, er:

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot e(t) \cdot c(t) \quad (3.1)$$

$$e'(t) = r + \frac{f(t) \cdot c(t)}{g + c(t)} \cdot e(t) - d \cdot e(t) - a \cdot (1 - p) \cdot c(t) \cdot e(t) \quad (3.2)$$

c er mængden af cancerceller der er i systemet til tiden t , målt i millioner. Mængden af cancerceller der er i systemet i starten, c_0 , refererer til hvor mange der blev sprøjtet ind i musene. e er antallet af cytotoic T-celler til tiden t , også målt i millioner. Disse celler er en del af immunforsvaret, og deres job er at bekæmpe cancercellerne.

Når en cancercelle og en T-celle mødes, antager modellen to potentielle scenarier. Enten dræber T-cellen cancercellen, eller cancercellen dræber eller deaktiverer T-cellen. Chansen for at T-cellen vinder kampen er betegnet som p , og chancen for at cancercellen vinder er derfor $(1 - p)$. Da p bestemmer chancen for et udfald kan den naturligvis kun være mellem 0 og 1. Den værdi som KK har fundet frem til er $p = 0.998$, hvilket betyder at hver gang en cancercelle og en T-celle kommer i kontakt, er der 99.8% chance for at T-cellen dræber cancercellen. Parameteren a bestemmer antallet af møder mellem T- og cancercellerne ud fra mængden af de to celletyper ved at blive ganget på produktet af c og e til tiden t . $a(1 - p)$ er derfor raten af T-celler der dør, og at $a \cdot p$ er raten af cancerceller der dør. a er sat til at være $a = 0.139$

Konstanten k bestemmer den maksimale vækstrate som cancercellerne kan have. Dette medfører, at jo større k er, desto hurtigere kan cancercellerne vokse. b er den reciprokke af bærekapaciteten for cancerceller, hvilket betyder at $\frac{1}{b}$ er det højeste antal cancerceller der kan være i systemet. Dette har den effekt på systemet, at når b mindskes, stiger bærekapaciteten for cancerceller. b har fået værdien $b = 0.002$, hvilket betyder at den maksimale mængde af cancerceller, er 500 millioner. Ligningen for væksten af cancerceller (ligning (3.1)), kan opdeles i to led. Det første led er:

$$k \cdot c(t) \cdot (1 - b \cdot c(t)) \quad (3.3)$$

Dette led bestemmer vækstraten af cancerceller når de ikke har noget modstand. Hvis man skulle beskrive cancers vækst i et system uden noget immunforsvar, ville dette led være tilstrækkeligt. Det andet led i ligningen, som inkluderer T-cellernes påvirkning, er:

$$a \cdot p \cdot e(t) \cdot c(t) \quad (3.4)$$

Som beskriver T-cellernes påvirkning på antallet af cancerceller. De resterende parametre, er dem som kun indgår i ligningen som beskriver vækstraten af T-celler (ligning (3.2)).

r betegner her den naturlige tilførsel af T-celler som vil forekomme uden for sygdomsforløbet. Når immunforsvaret opdager canceren, sender det en besked om at der er behov for flere T-celler. Herefter tager det et vist antal dage for nye T-celler at modnes. Antallet af dage defineres som u og er i artiklen bestemt til 28,0545. f er en betinget erklæring som beskriver netop dette. Fra sygdommens start, t_0 , indtil u dage er forløbet, er $f = 0$. Efter dette er $f = v$, hvor v er raten af nye T-celler som ankommer i respons til sygdommen, og er bestemt som $v = 0.525$. Parameteren g , er en parameter som bestemmer, at tilvæksten af T-celler vil være størst, når antallet af cancerceller er størst. Dette går sammen i leddet:

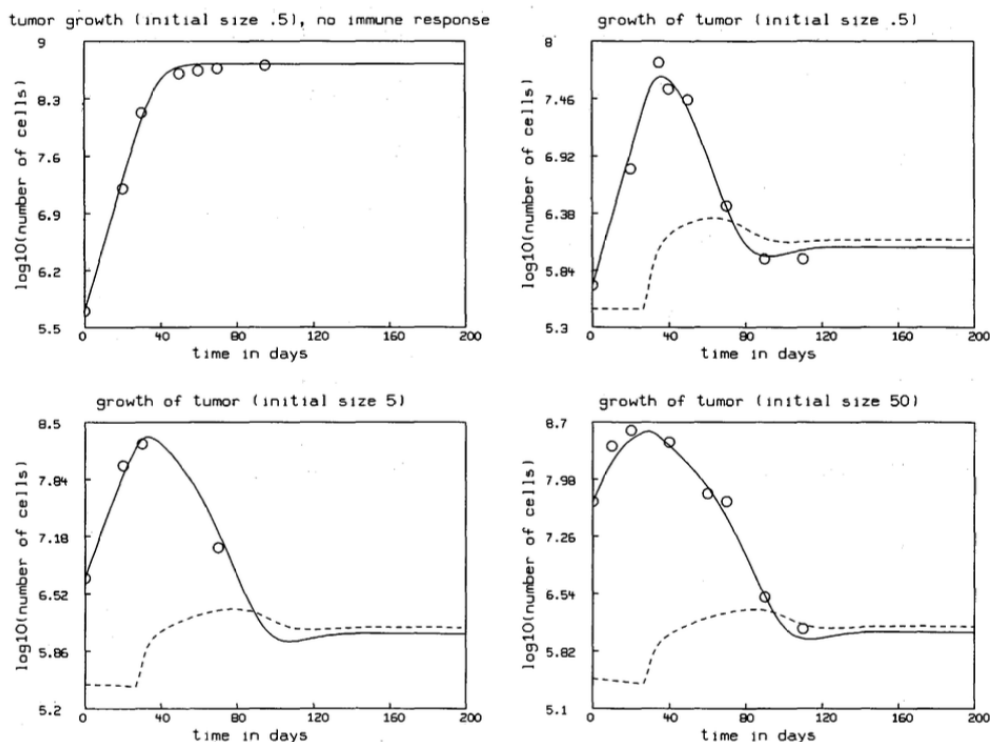
$$\frac{f \cdot c}{g + c} \cdot e \quad (3.5)$$

Dette beskriver T-cellernes vækst, i respons til cancercellerne, og er en skalar for antallet af cancerceller i systemet. Raten af nye T-celler, stiger dog ikke lineært med antallet af cancercellerne, da leddet opfører sig som en logistisk funktion. Dette betyder at der er en bærekapacitet for T-celler, og som antallet af disse nye celler nærmer sig bærekapaciteten falder vækstraten for T-cellerne.

T-celler dør med raten d . d og r hænger sammen da r er bestemt som $r = d \cdot e_0$. Her er e_0 antallet af T-celler som er tilstede, ved sygdommens start. Før musen får cancer, vil antallet af T-celler altid være e_0 , da $r - d \cdot e_0 = 0$, og der derved ikke vil være nogen ændring i populationen.

Ydermere er der intet i strukturen for modellen, som gør at parametrene ikke vil være entydige. I alle led er der kun én korrigerende parameter, med undtagelse af $apec$ leddet i ligning (3.4). a kan ikke bestemmes her da p også er i leddet. I ligning (3.2) optræder a igen, men denne gang i enkeltstående form, hvori den kan bestemmes adskilt fra andre parametre.

KK's fundne parametre, og deres fit til det førnævnte datasæt, er vist på figur 3.2.



Figur 3.2: Figuren viser KK's fit til de fire datasæt, med de parameterværdier som de præsenterer. Den stiplede linje viser antallet af T-celler, den fuldt optrukne er antallet af cancerceller og punkterne viser datasættets beskrevne antal af cancerceller. Figuren er taget fra Kuznetsov & Knott, 2001.

På grafen der beskriver musen uden et immunforsvar ses der en stigning af cancerceller indtil de rammer systemets bærekapacitet. Konstanterne k og b er blevet estimeret ud fra denne graf, da de er de eneste to parametre som indgår i led uden e .

De tre andre grafer illustrerer hvad der sker når man indsprøjter 0.5, 5 og 50 millioner cancerceller ind i en mus med normalt immunforsvar. I disse tre grafer ses det, at antallet af cancerceller stiger hurtigt, indtil de rammer et toppunkt hvor de begynder at falde igen. Toppunktet rammes når f , systemets betingede erklæring aktiveres og bliver til en ny kilde af T-celler. De falder indtil systemet rammer et ækvilibrium hvor canceren går i cancer dormancy stadiet hvor mængden af cancerceller hverken vokser eller falder (Kuznetsov & Knott, 2001).

3.1.2 Kvalitativ analyse

I dette afsnit vil vi foretage en vurdering af de parameterverdier, som KK har fundet frem til, samt de enkelte matematiske led som benyttes til modellen. Vi vil fokusere på hvilke argumenter der er benyttet i forhold til startgæt og grænser; både ud fra biologisk mekanistiske overvejelser, samt matematisk analyse af systemet.

Artiklen, som denne rapports case-model stammer fra, har haft som fokus at undersøge hvordan case-modellen kan tilpasses til at beskrive væksten af kræftceller, når de forlader en "hvilelignende tilstand". Case-modellen er derfor mest anvendt som et udgangspunkt for dette, og er ikke beskrevet med stor detaljegråd i artiklen. KP har modelleret samme datasæt med en model, som grundlæggende set har samme struktur, men hvor fokus har omhandlet parameterestimering og bifurcation analyser. Det medfører at KP indeholder mere dybdegående overvejelser og argumenter for de valg der er truffet undervejs. Før selve den kvalitative analyse foretages, vil modellen fra KP introduceres og sammenlignes med modellen fra KK.

Dette afsnit vil derfor inkludere referencer og sammenligninger med konklusionerne fra KP, i tilfælde hvor formidlingen af overvejelser om parameterverdiernes størrelse kommer til kort i KK.

Sammenligning med KP's model

I KK modelleres datasættet ud fra en mere generaliseret model som udtrykkes via,

$$\frac{dT}{dt} = a \cdot T(1 - b \cdot T) - n \cdot E \cdot T \quad (3.6)$$

og

$$\frac{dE}{dt} = s + \frac{p \cdot E \cdot T}{g + T} - m \cdot E \cdot T - d \cdot E \quad (3.7)$$

Sammenlignet med ligning (3.1) og ligning (3.2) fra KK, ses to væsentlige forskelle. Først og fremmest er visse af parametrene forsøgt at blive udvidet og relateret til biologiske sammenhænge i KK. Yderligere introduceres tidsforskydningen, u , først i KK.

KP	T	a	m	n	p	s
KK	C	k	$a \cdot (1 - p)$	$a \cdot p$	v	r

Tabel 3.1: Tabellen illustrerer hvordan de enkelte parametre i modellen fra KP tilsvares dem som ses i KK

Opsættes ligning (3.6) og ligning (3.7) med en notation som er konverteret til tabel 3.1 opnås,

$$\frac{dC}{dt} = k \cdot c \cdot (1 - b \cdot c) - a \cdot p \cdot E \cdot C \quad (3.8)$$

og

$$\frac{dE}{dt} = r + \frac{v \cdot E \cdot C}{g + C} - a \cdot (1 - p) \cdot E \cdot C - d \cdot E \quad (3.9)$$

Disse ligninger tilsvare tydeligt ligning (3.1) og ligning (3.2) med den undtagelse at $f = v$ til alle tider. Udover tidsforskydningen tilføres KK's model også et mere mekanistisk perspektiv for KP's parameter n , som i KK udvides til $apec$.

”Thus, ap is the “kinetic” constant that multiplies e and c to form the overall tumor-cell destruction rate term $apec$ ”

(Kuznetsov & Knott, 2001)

I KP er parameteren n blot en matematisk beskrivelse af den faktor som produktet af c og e har for systemet. Egentlig er det en sammenlægning af flere parametre (k_{-1} , k_1 , k_2 og k_3), som beskriver hver deres biologiske proces, men hver af disse processer beskrives blot gennem en konstant, som faktor, uden yderligere biologisk baggrund for udformningen af leddet. KK forsøger at beskrive dette led mere mekanistisk gennem ændringen af leddet til ap . Her beskrives en konkret rate for celledestruktion ud fra en sammenlægning af raten for hvor ofte cellebindinger resulterer i den ene celledes død og procentsatsen for at den celle som dør ved denne interaktion er en cancercelle. De øvrige parametre er direkte sammenlignelige i deres matematiske udtryk for begge artikler. Fremover i denne analyse vil begge modeller beskrives ud fra notationen fra KK med udgangspunkt i konvertionen fra tabel 3.1

En anden iøjnefaldende forskel mellem de to artikler er i antallet af betydende cifre for de enkelte parametre. e_0 , k og b er estimeret til to cifre, d er estimeret til tre, mens de resterende er estimeret til fire cifre. I KK er alle parametre noteret med enten ni eller ti betydende cifre. Det er ikke forklaret i KK hvilken baggrund valget om at inddrage så mange betydende cifre er taget på, idet forsøgsdata fra artiklen (Siu & Vitetta, 1986) blev opgivet med ned til et betydende cifre fra indsprøjtet cancerceller - eksempelvis $5 \cdot 10^6$ celler. Inddragelsen af det store antal cifre kan umiddelbart betragtes som tvivlsom med tanke på, at biologiske meget sjældent kan forudsiges med så stor nøjagtighed, da der ofte er tale om meget komplekse systemer. Ved behandling i denne rapport, vil parameterstørrelserne derfor begrænses til

tre betydende cifre. Hvor stor mekanistisk nøjagtighed de enkelte parametre i systemet har, undersøges dybere ved at betragte henholdsvis cancer- og T-celle funktioner for sig.

Cancercelle funktionen

Som nævnt i afsnit 3.1.1, så består funktionen for cancercellernes udvikling af to led; et logistisk udtryk for egenvæksten, ligning (3.3) , og et udtryk for T-cellernes angreb på cancercellerne, ligning (3.4). Parametrene til det logistiske led, k og b , bestemmes i artiklen ved at fitte en gruppe mus uden immunforsvar. På denne måde kan leddet betragtes uafhængigt af det resterende system.

”Group 0 mice have no immune response; tumor-size data from this group will be used to help estimate the parameters that define tumor growth by fitting the pure growth model

$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t))$ with $c(0) = 0.5.$ ”

(Kuznetsov & Knott, 2001)

I KK uddybes denne proces ikke yderligere. I KP, hvor samme metode er anvendt, bruges metoden til at argumentere for at parameterværdierne er mekanistisk troværdige

”the logistic model is biologically reasonable and the two parameters that need to be estimated from the data determine the slope and asymptote of the tumor growth curve.”

(Kuznetsov, Makalkin et al., 1994)

Den logistiske vækstmodel betragtes, som biologisk meningsfuld for cellevækst, hvor en øget population ved lave populationer resulterer i større vækst, qua større potentiale for celledeling. I takt med at populationen nærmer sig bærekapaciteten begynder væksten at falde. Ved at bestemme disse to parametre isoleret fra resten, ud fra et matematisk system, som er biologisk meningsfuld, så kan parametrene betragtes som mekanistiske og værdierne antages at være relativt præcise.

For det led som anvendes til at beskrive interaktionerne mellem killer- og T-celler, *apcc*, er det mindre klart i hvilken grad, parametrene kan opfattes som mekanistiske. Parametrenes funktion i modellen er opsat til at tilsvare en meningsfuld biologisk mekanisme. I og med at mekanismen er afhængig af $e(t)$ er den samlede sammensætning af denne mekanisme dog meget kom-

pleks i forhold til den rene logistiske vækst og det er derfor mere vanskeligt at efterprøve med data, hvorvidt den antagede mekanisme afspejler virkeligheden. Yderligere er selve værdierne for parametrene svære at betragte som mekanistiske. KK har fundet frem til værdierne udelukkende gennem parameterestimering af det fulde system; altså en fænomenologisk tilgang.

T-celle funktionen

For T-celle funktionen anvendes der fem yderligere parametre for systemet, henholdsvis r , d , u , v og g .

r er i denne forbindelse defineret som $d \cdot e_0$. e_0 betragtes af KK, som biologisk velbegrundet. Værdien tager udgangspunkt i, at der findes et kendt estimat for det samlede antal splenocytter i milten på de pågældende forsøgsmus, samt et estimat for hvor mange af disse celler der er reaktive overfor tumoren (Kuznetsov, Makalkin et al., 1994). Estimerne er dog præget af en vis usikkerhed. Især estimatet for antallet af reaktive celler, der beskrives som,

”The frequency of CTL precursors reacting to alloantigens forms several tenths of a percent of the total number of lymphocytes (Bronz, 1987)”

(Kuznetsov, Makalkin et al., 1994)

Frekvensen beskrives altså som en fraktion af ”adskillige tiendedele af en procent”. Ordet ”adskillige” er både bredt og upræcist og skaber dermed en stort spænd af værdier, som opfylder dette krav. Estimatet kan i den forstand snarere betragtes, som et estimat for en størrelsesorden, end en konkret værdi. Det samme gør sig, til dels, gældende for estimatet af det samlede antal af splenocytter. Dette antal estimeres som,

”The spleen of a BALB/c mouse contains approximately 10^8 splenocytes.”

(Kuznetsov, Makalkin et al., 1994)

Antallet af splenocytter vurderes altså som ”tilnærmelsesvist 10^8 ”. Hvor præcist dette estimat er tiltænkt at være, er ikke umiddelbart klart fra formuleringen. Estimatet kan både fortolkes som et estimat på 10^8 for antallet af splenocytter og som et estimat på 10^8 for størrelsesordenen af antallet af splenocytter. e_0 kan dermed siges at være mekanistisk velbegrundet, omend værdien for parameteren vil være præget af stor usikkerhed.

KK's værdi for e_0 er valgt på baggrund af et estimat for e_0 i KP på $3,3 \cdot 10^5$ celler. I KK er dette estimat tilpasset til $3 \cdot 10^5$, hvilket er en meningsfuld afrunding med tanke på usikkerheden bag de betragtninger som ligger til grund for estimatet. Antages det at 10^8 udtrykker et estimat for den konkrete værdi, så vil $3 \cdot 10^5$ tilsvare at "adskillige tiendedele" tolkes som tre tiendedele.

Den anden del af udtrykket r udgøres af parameteren d som også indgår egenrådigt i ligningen. Denne parameter, som beskriver dødsraten, er direkte relateret til cellernes levetid via et reciprok forhold. Cellernes levetid er der biologisk baggrund for at estimere, som ses i KP

"The lifetime of T lymphocytes from the spleen and the blood is not known precisely but can be estimated to be approximately 30 days or more (Reynolds et al., 1985; Gray & Leanderson, 1990)."

(Kuznetsov, Makalkin et al., 1994)

I KP er værdien for d , via parameterestimering, fundet til $d = 0,0412$, hvilket tilsvare en levetid på $\frac{1}{d} = \frac{1}{0,0412 \text{dage}^{-1}} = 24,3$ dage. Dette resultat afviger altså en smule fra det biologiske estimatet, men er dog indenfor en lignende størrelsesorden. I KK er d derimod bestemt til $d = 0,591$. Dette tilsvare $\frac{1}{d} = \frac{1}{0,591 \text{dage}^{-1}} = 1,69$ dage, som afviger markant i størrelsesorden fra den levetid som estimeredes at være realistisk i KP.

Model-fitting predicts that the life time of killer cells is short (about two days).

(Kuznetsov & Knott, 2001)

Det klargøres ikke tydeligt her, hvorfor denne værdi alligevel accepteres som et validt estimat for modellen.

u introduceres som en ny parameter i KK, hvor den dikterer en betinget erklæring for modellen. Parameterens størrelse afgøres af hvor længe det tager for en precursor celle at udvikle sig til en killercelle.

"Note particularly that 28 days is an appropriate estimate of the maturation time for CD8+ cytotoxic T lymphocytes."

(Kuznetsov & Knott, 2001)

Et estimat for værdien af denne periode er ikke let tilgængeligt i litteratur, men det til trods, så vurderes det i KK at 28 er et passende estimat for u . Ved denne vurdering anvendes der ingen direkte kildehenvisning til at underbygge deres vurdering. Såfremt vurderingen er baseret på litteratur, kan det antages, at vurderingen er foretaget på baggrund af KP, som generelt for KK udgør deres referencepunkt for uddybende begrundelser for KK's parameterverdier. Der er dog ingen biologisk uddybning der relaterer direkte til cellernes modningstid i KP.

Vurderes uklarheden omkring værdien for u i samspil med uklarheden omkring værdien for d i samspil, kunne det dog tyde på, at introduktionen af u under matematiseringen indebærer et ændret perspektiv på systemet i KK. Mens KP ikke indeholder nogen argumentation for en periode for cellernes modningstid, så indeholder den trods alt et argument, som relaterer til deres livsforløb, som estimeredes til over 30 dage. Det fremgår ikke tydeligt i beskrivelsen af dette livsforløb, hvorvidt levetiden på 30 dage skal betragtes som 30 dage efter udviklingen til T-celle eller 30 dage fra dannelsen af precursor cellen. Betragtes det som sidstnævnte, kan cellernes levetid betragtes som todelt; en udviklingsperiode og en aktiv periode. d er beskrevet som selve T-cellernes dødsrate og den samlede levetid vil dermed kunne beskrives som,

$$\text{levetid} = \text{precursor levetid} + \text{T-celle levetid} = u + \frac{1}{d} \quad (3.10)$$

hvor u udtrykker perioden for udviklingsfasen af precursor cellerne, mens $\frac{1}{d}$ udtrykker den periode, hvor de færdigudviklede T-celler aktivt påvirker systemet. Denne måde at forstå parametrene kræver en grad af fortolkning af den information der er tilgængelige i KP, da det ikke formidles klart i artiklen. Betragtes parametrene således, vil levetiden være

$$\text{levetid} = u + \frac{1}{d} = 28,05\text{dage} + 1,69\text{dage} = 29,74\text{dage}$$

Fortolkningen medfører altså en levetid for cellerne, som er meget tæt på den estimerede værdi på 30+ dage. Fortolkningen indebærer dog også, at det ikke åbenlyst kan vurderes, hvor realistiske parameterverdierne for u og d er ved en parameterestimering. Denne vurderingen kan kun foretages ved at betragte parametrene samlet.

Uklarheden omkring parameteren u har endnu en konsekvens, som kommer til udtryk ved den matematiske fortolkning af systemet.

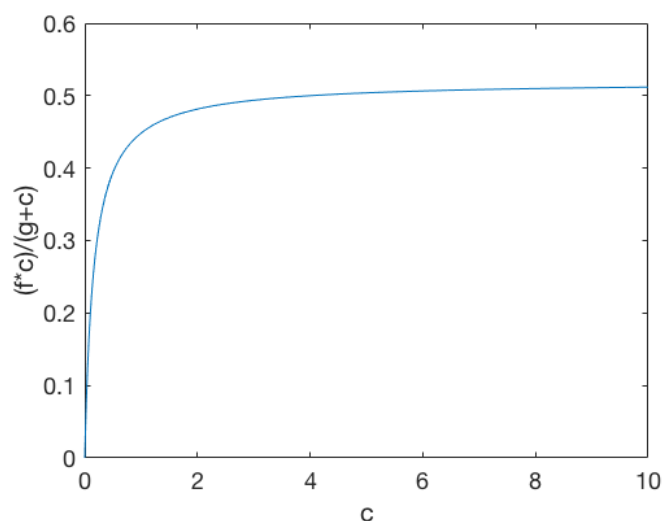
the parameter u is the delay time of the immune response before “new” killer cells can be applied to attack the tumor; i.e., u is the time for precursor cells to mature into killer cells. The function f “cancels” the additional growth term until $t > u$.
 (Kuznetsov & Knott, 2001)

Dette citat underbygger den tolkning af u , som er foretaget i denne rapport, hvor u udtrykker tiden det tager for precursor celler at udvikle sig til T-celler. Denne tolkning indebærer at responsen på tilstedeværelse af cancerceller er øjeblikkelig og at forsinkelsen i stedet forekommer ved effekten af denne respons. Betragtes ligning (3.4), ses det dog at dette ikke afspejles i den matematiske konstruktion. Når leddet aktiveres, så afhænger det af den samtidige population af cancerceller. For at opsætte et matematisk udtryk, som er tro mod den mekanisme, som er beskrevet, så burde udtrykket i stedet referere forsinket til cancercellepopulationen. Uden denne forsinkelse beskrives der matematisk set i stedet en situation, hvor selve responsen er forsinket.

De sidste parametre i systemet er parametrene v og g , som begge indgår i den betingede erklæring, som styres af u . Parametrene i denne del af systemet er umiddelbart ikke specielt biologisk velbegrundede. I KP argumenteres der ganske vist for hvilken biologisk mekanisme leddet beskriver, men det klargøres ikke hvorfor det er oplagt at beskrive denne mekanisme gennem lige præcis dette matematiske udtryk. Yderligere er der ved bestemmelsen af parameterværdierne udelukkende taget højde for parameterestimeringens output. Dette har resulteret i meget store forskelle mellem estimererne for v og g mellem KK og KP. v estimeres til 0,525 i KK og 0,125 i KP, mens g estimeres til 0,16 i KK og 20,19 i KP. Især forskellen i g , som er mere end en faktor 100, er iøjnefaldende. Uklarheden omkring denne parameter understreges også i KK under sensitivitetsanalysen.

”However, increasing the parameter g from 0.16 to 5 did not change approaching the tumor dormancy steady state.”
 (Kuznetsov, Makalkin et al., 1994)

g ses dermed ikke blot at være besværlig at estimere biologisk, men også ud fra det anvendte datasæt. Ved at analysere leddet, $\frac{v \cdot c}{g+c}$, ved c 's størrelse i forhold til g , ses at når $c \gg g$, så $\frac{v \cdot c}{g+c} = v$, mens $\frac{v \cdot c}{g+c} = \frac{v \cdot c}{g}$, når $c \ll g$. Det vil sige at g 's indflydelse på systemet kun kan undersøges dybdegående ved lave værdier for c . Ligning (3.5)'s indflydelse på systemet ved forskellige c -værdier kan ses på figur 3.3.



Figur 3.3: Her ses en graf over hvordan funktionen $y = \frac{f \cdot c}{g+c}$ udvikler sig som c stiger. Her har f og g fået de samme værdier som KK har givet dem.

Figur 3.3 viser at $g = 0,167$ tidligt i modellens udvikling mister sin indflydelse på systemet. Allerede ved $c = 1$ er leddet ved $\frac{1}{0,167+1} = 85,7\%$ af dets maksimale værdi og ved $c = 10$ $98,3\%$. Siden leddet først aktiveres længe efter cancercellepopulationen er begyndt at vokse (efter u dage), så betyder det i praksis for dette system, at g først er relevant for systemet, når c nærmer sig sit ækvilibrium. Jævnført figur 3.1, så er selv de laveste datapunkter efter u i størrelsesorden 10^0 og langt de fleste over 10^1 .

Værdierne i størrelsesordenen 10^0 ses kun omkring cancercellepopulations ækvilibrium og det vil primært være for denne del af kurven, at g har relevans for systemet. Da datasættet er yderst sparsomt for denne del af c 's udvikling, så kan værdien for parameteren ikke rigtig estimeres med nogen form for præcision. Der kræves derfor meget store ændringer i g for at ændre væsentligt på modellen.

En yderligere betragtning i forhold til ligningens betingede erklæring er opfattelsen af parameteren u . u 's indvirkning på systemet i den opstillede model medfører, at systemet opfattes sådan, at alle de celler, som udvikles i respons til cancercellerne tager præcis lige længe om at modnes. Denne betragtning er ikke biologisk intuitiv, omend den dog potentielt kan betragtes som en valid forsimpning af systemet. Dog ville samme egenskab kunne opnås for udtrykket ved at benytte et logistisk led fremfor et conditional statement. Et logistisk led vil kunne beskrive tilvæksten af T-celler via en fordeling over

flere dage. Dette tillader en vis fleksibilitet i forhold til at hvordan tilvæksten af T-celler opfattes over tid. Samtidig vil dette led kunne tilnærme sig præcis samme opførsel som en betinget erklæring ved $w \rightarrow \infty$, hvor w er den logistiske vækstrate.

I denne sektion har vi undersøgt modellen, som den er blevet præsenteret af KK. Vi har gennemgået alle parametre og variabler og forklaret deres betydning, både med et biologisk- og matematisk udgangspunkt. De vigtige led i de to ligninger er blevet redegjort for matematisk. Vi har også sammenlignet KK's differentialligninger med de differentialligninger som KP har lavet. Det er disse ligninger som KK bygger videre på for at lave deres model. Da parameteren u kan forstås på flere forskellige måder, forklares her hvordan, og hvorfor, vi har valgt at forstå u som tiden det tager precursor celler at modnes og blive til T-celler.

3.2 Modelkalibrering

I denne sektion vil vi lave en parameterestimering på KK's model ud fra det data som de har brugt. Dette gøres først ved at lave en sensitivitetsanalyse, hvor vi finder ud af hvilke parametre der er mere sensitive end andre. Parametrenes sensitivitet analyserer vi via en procentvis ændring af alle parametrene enkeltvist. Derefter ser vi hvor meget modellens output varierer i forhold til de ændrede parametre. Når modellens output ændres meget i forhold parameterens ændring, så bliver parameteren set som at være sensitiv. Hvis det omvendte sker sætter vi parameteren som robust.

I parameterestimeringen erstatter vi de parametre som vi fandt robuste med konstanter, og estimerer dem som vi anser at være sensitive. Estimeringen udføres med et program i MATLAB. Programmet bliver ikke redegjort for. De nye parameterverdier vi finder sammenlignes med KK's værdier hvorefter en række overvejelser bliver udført på baggrund af hvad analysen betyder

3.2.1 Sensitivitetsanalyse

I dette afsnit vil vi lave en sensitivitetsanalyse af case-parametrene, udført i MatLab, og analysere hvordan modellens output ændres når de individuelle parametre ændres procentmæssigt. I afsnittet vil vi træffe valg undervejs for at øge nøjagtigheden af identifikation af parametertype, og tydeliggøre hvilke parametre der har størst effekt på modellen.

Strukturen for afsnittet er bygget op således at den anvendte metode og procedure bliver introduceret kort. Herefter påbegyndes analysen af hvilke parametre der er mere sensitiv end resten, sammen med løbende konklusioner og overvejelser herom. Resultaterne bliver præsenteret i grafer og tabeller.

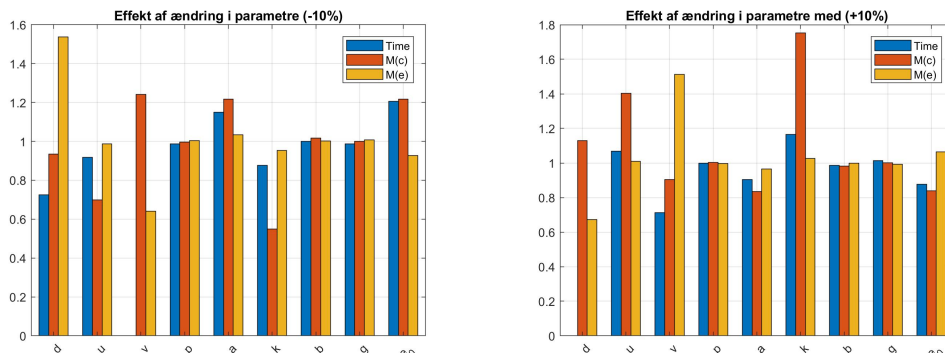
Sensitivitetsanalysen er simuleret med *MatLab* med udgangspunkt i case-modellens parametre. Første skridt i analysen er en enkeltvis ændring af hver parameter på $\pm 10\%$.

Parameteren p er en procentsats og bør derfor altid ligge i et interval $p \in [0 : 1[$ for at være meningsfuld. Da p 's værdi er på 99,8% vil en +10% ændring af p resultere i en værdi uden for dette interval. Der varierer derfor i stedet på $1 - p$, fremfor selve parameteren p . Fremover gennem analyse afsnittene, vil værdien $1 - p$ præsenteres i p 's sted som p_e , da det er denne værdi som varierer og dermed den relevante værdi at analysere på.

Differentialligningerne er simuleret for hver af de ændrede parameterverdier over et tidsinterval $t \in \mathbb{N}$ med tidsskridt på 1 dag; én gang med det originale fit, og én for hver gang vi ændrer en parameter, hvis afvigelse kaldes z . Sammenligningen foretages i punktet hvor der er et globalt maksimum på kurven for det originale fit og på kurverne for hver parameter som er ændret. Således vil to punkter altid sammenlignes relativt til hinanden som i ligning (2.4).

Resultaterne vises med to søjlediagrammer, én for hver af de procentvise variationer. Henad den horisontale akse optræder parametrene, og på den vertikale akse vises værdier for s . På den vertikale akse svarer værdien 1 til ingen ændring i modelrespons.

Følgende figurer har tre søjler i blå, orange og rød farve som viser parameter respons på procentvist variation af værdi. Den røde og orange søjle repræsenterer cancer- og T-celler. Den blå søjle angiver punktet hvor graferne for populationerne af cancer- og T-celler krydser hinanden til tidspunktet t , i forskel fra det originale fit.



(a) Søjle diagram der viser resultatet af en sensitivitsanalyse med en -10% ændring i alle parametrene. Hen af den horisontale akse er de enkeltvise parametre for hver af populationerne og på den vertikale akse angives den relative forskel fra modeloutputtet

(b) Søjle diagram der viser resultatet af en sensitivitsanalyse med en $+10\%$ ændring i alle parametrene. Hen af den horisontale akse er de enkeltvise parametre for hver af populationerne og på den vertikale akse angives den relative forskel fra modeloutputtet.

Figur 3.4: blå, rød og orange farve viser parameterrespons på procentvist variation af værdi. Den røde og orange søjle repræsenterer cancer- og T-celler. Den blå søjle angiver punktet hvor populationerne krydser til tidspunktet t , i de respektive ligninger, i forskel fra det originale fit. På den vertikale akse vises værdier for s , med 1 svarende til ingen ændring i modeloutput. Figurene er lavet gennem appendix B.1.

Parametrenes robusthed vurderes ved at sammenligne afvigelserne numerisk i tabel 3.2. For at vægte negative ændringer i samme grad som positive, sammenlignes disse ud fra forskellen relativt til den højeste værdi af s og n_0 . I praksis gøres det ud fra følgende ligning,

$$\text{Hvis } s > 1, \text{ så } z = s - 1$$

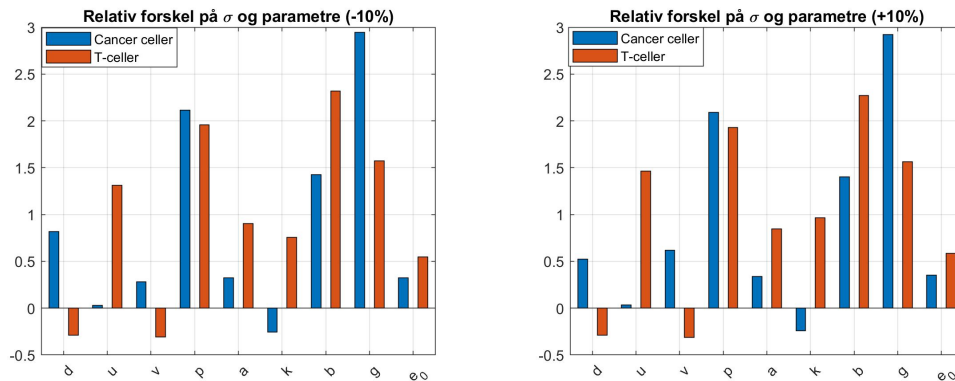
$$\text{Hvis } s < 1, \text{ så } z = \left(\frac{1}{s}\right) - 1$$

For at give et bedre overblik vises de relative forskelle i tabel 3.2

Afvielser	c -10%	e -10%	c +10%	e +10%
d	0.0697	0.5380	0.1293	0.4876
u	0.4297	0.0134	0.4031	0.0087
v	0.2409	0.5583	0.1048	0.5132
p_e	0.0035	0.0030	0.0035	0.0030
a	0.2172	0.0344	0.1984	0.0356
k	0.8222	0.0481	0.7532	0.0272
b	0.0172	0.0013	0.0171	0.0013
g	0.0005	0.0074	0.0005	0.0068
e_0	0.2166	0.0785	0.1932	0.0649

Tabel 3.2: Tabel over hver enkelt parameters relative forskel fra det originale fit, sammenlignet i toppunktet for alle simulationer

Ud fra figur 3.4b og 3.4a tyder det på at parametrene g , b og p_e er af robust karakter, da z i disse tilfælde er tæt på 0. Derudover fremgår det fra tabeloversigten, at ingen af parametrene er fuldstændigt robuste. Dette ville have resulteret i at modellen ikke havde nogle entydige løsninger, da alle værdier for en fuldstændig robust parameter, ville resultere i samme modeloutput. For at definere et kriterie for robusthed, anvendes et kriterie baseret på gennemsnit og *standardafvigelse* (σ) af z .



(a) Søjlediagram der viser forholdet mellem σ og parametrene for tumor- og T-celler, ved en ændring af z med -10% (b) Søjlediagram der viser forholdet mellem σ og parametrene for tumor- og T-celler, ved en ændring af z med +10%

Figur 3.5: Figuren viser et søjlediagram i logaritmisk skala. På den horisontale akse vises $\frac{\sigma}{z}$ og med parametrene henad den vertikale akse. Figurene er lavet gennem appendix B.2.

Figur 3.5 viser et søjlediagram i logaritmisk skala. På den horisontale akse

vises $\frac{c}{z}$ og med parametrene henad den vertikale akse. Dette trin bekræfter os i at de førnævnte parametre er robuste.

De beregnede skæringspunkter har to tilfælde, hvor populationerne ikke krydser i et punkt. Dette gælder for v når denne ændres negativt med 10% og i det modsatte tilfælde for d . I begge situationer er effekten af parameter ændringen at $c(t)$ og $e(t)$ ikke krydses som det vises i det originale fit.

Med denne analyse er parametrene blevet opdelt parametrene i to kategorier, for henholdsvis robusthed og sensitivitet. b, g og p_e blev bestemt til robuste parametre, mens de resterende er oplagte at estimere på med det formål at finde et bedre bud på fit end det originale.

3.2.2 Parameterestimering

I rapporten af KK, som bliver brugt som opgavens case, fremgår metoden til parameterestimeringen ikke. I dette afsnit vil de værdier som præsenteres af KK, sammenlignes med dem som vi kan findes gennem en veldefineret og anerkendt søgealgoritme. Vi vil i den forbindelse undersøge hvor meget parameterværdierne, som vi har fundet, og den som KK har brugt, varierer fra hinanden. Dette gøres for at undersøge søgealgoritmernes indflydelse på parameterstørrelser

I forbindelse med parameterestimeringen vil vi inddrage resultaterne fra sensitivitetsanalysen. I det afsnit fandt vi ud af hvilke parametre der var robuste. Disse parametre bliver nu sat som konstanter, og en parameterestimering bliver udført på de mere sensitive parametre. Parameteren p_e vurderede vi også som robust i sensitivitetsanalysen, men den vil ikke blive fastholdt under parameterestimeringen. Dette skyldes til dels, at p_e er en procentsats, hvis spænd ligger ved $0 < p_e < 1$, og samtidig har en værdi meget tæt ved dens grænse. Dette besværliggjorde analysen, hvis resultat for p_e dermed ikke kan opfattes med samme sikkerhed.

Da formålet med denne rapport er at undersøge de enkelte trin der er foretaget undervejs i modelleringsprocessen for sig, så foretrækkes der, at undersøge hvert af disse trin med udgangspunkt der er så tæt på den oprindelige model som muligt. Derfor vil det ikke være hensigtsmæssigt at vurdere p_e som robust under parameterestimeringen.

Parameterestimeringen er udført via *Shuffled Complex Evolution*, udviklet ved *University of Arizona* (SCE-UA). Algoritmen placerer et antal punkter

i parameterrummet, baseret på nogle givne startværdier. Rundt om disse punkter konstrueres nogle komplekser af punkter, som herefter optimeres på.

Til optimeringen undersøges det, hvorvidt der for det dårligst estimerede punkt, opnås et bedre fit, når punktet flyttes tættere på centrum eller længere væk. Denne procedure gentages herefter for det værst estimerede punkt i det nykonstruede kompleks og for hver gang et af disse punkter rykkes udføres en simulering af differentiaalligningernes løsning.

Efter at have optimeret på disse punkter et stort antal gange, vil hvert af kompleksene have lukket sig rundt om det minimum, med mindst gennemsnitlig afvigelse, omkring dets startposition. Herefter blandes de fundne punkter for hvert kompleks med hinanden, hvorved der dannes nye komplekser. Herefter gentages fremgangsmåden fra før, således at hver af disse komplekser igen finder hen imod et minimum.

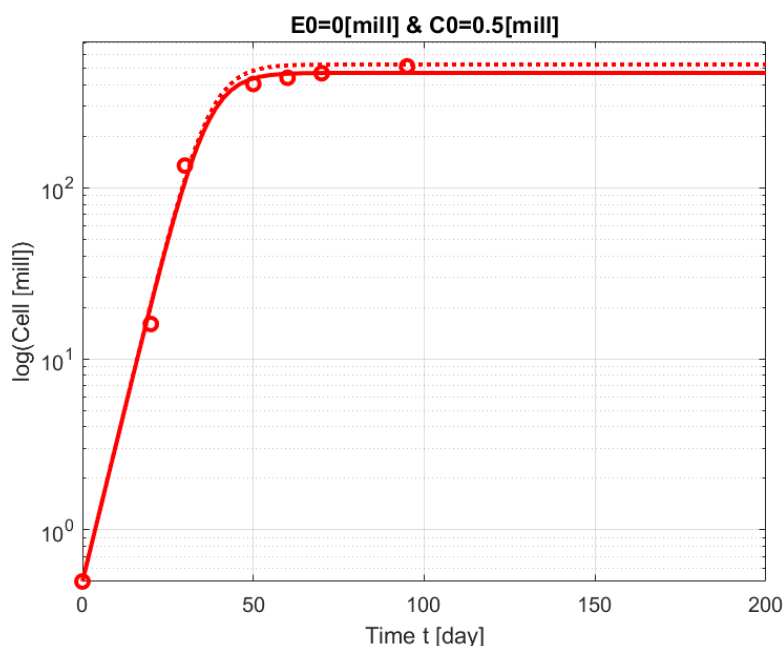
Blandingerne udføres da det ikke altid kan afgøres om et kompleks har fundet det globale minimum eller blot et lokalt. Ved at blande de fundne punkter øges sandsynligheden for at finde frem til globalt minimum. For hver blanding undersøges det, hvor stor forbedring der er i forhold til det forrige minimum, således at et krav for en minimumsforbedring kan opstilles.

Parameterestimeringen udførtes således, at programmet fortsatte indtil, at et helt loop af algoritmen resulterede i en forbedring under 0.1% eller at programmet havde kørt 10.000 simulationer af modellen. Differentiaalligningernes løsning blev approksimeret med Matlabs *ode23s* differentiaalligningsløser. Kvaliteten af fittet vurderes ud fra en korrigeret standardafvigelse, hvor optimeringer med få variable parametre vægtes som bedre fit, end optimeringer med mange.

$$\sum_{n=1}^i \sqrt{\frac{\log(x_i(data)) - \log(x_i(fit))}{\log(\bar{x}(data)) - N}} \quad (3.11)$$

Hvor $(x_i(data)-x_i(fit))$ er afstanden mellem datapunkt og fit, $\bar{x}(data)$ er den gennemsnitlige dataværdi, mens N er antallet af variable parametre.

Med samme fremgangsmåde som KK, estimeres parametrene k og b indledningsvist til det første datasæt, hvori de er de eneste betydende parametre. De fundne størrelser er: $k = 0.188$ og $b = 0.00213$.



Figur 3.6: Parameterestimering af k og b , til datasættet uden nogle T-celler, sammenlignet med KK's fit til samme datasæt. Den stiplede linje repræsenterer KK's fit til datasættet, mens den solide linje er det nye fundne fit. Figuren er lavet gennem appendix C.1.

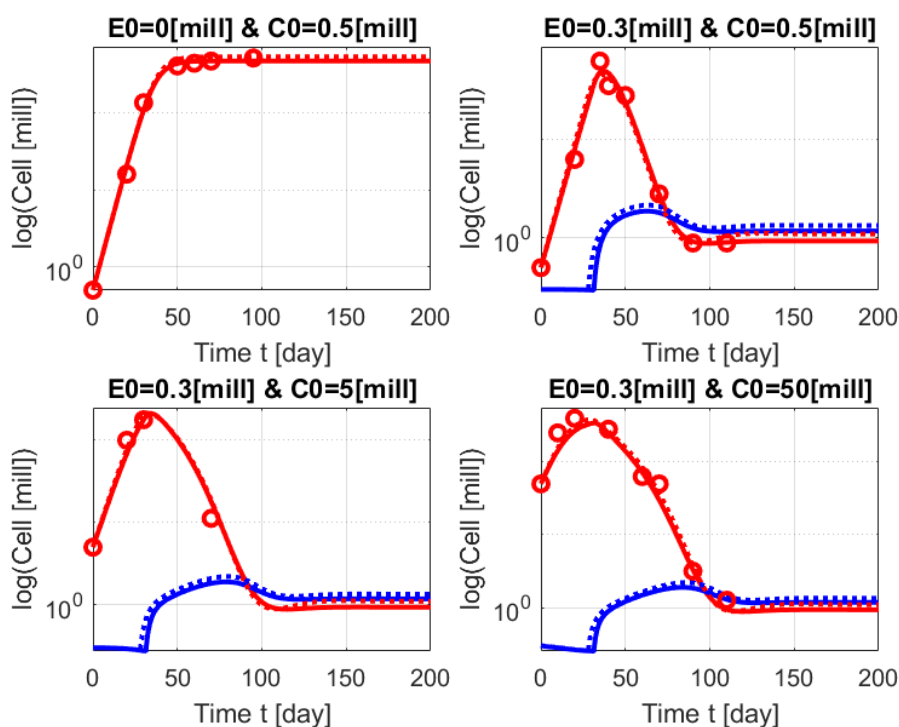
På figur 3.6 ses det, at KK har fundet frem til en b -værdi som resulterer i et lidt højere toppunkt for grafen. Den fundne k -værdi er kun anderledes på fjerde decimal, mens den nye b -værdi er anderledes med omkring 15%, sammenlignet med KK's.

At udviklingerne er synligt anderledes som resultat af blot en ændring i b , er dog lidt i modstrid med sensitivitetsanalysens bestemmelse af denne som en robust parameter. Dette er grundet at parameterens indflydelse på modellen er i bestemmelsen af dets maksimum. Da de resterende datasæt falder kort efter at have nået op på den mængde hvor b 's størrelse er relevant, er modellen derved robust overfor ændringer i denne. Det er dermed kun i datasættet uden nogen T-celler, at der forekommer nogen mærkbare ændringer i modellens output, når en ændring i b forekommer.

Parametrene k og b vil ikke tages med i overvejelserne i de resterende afsnit, da de nu er estimeret, og vil derved blive set som konstanter med de fundne værdier fremover. Da de eneste to parametre som påvirker datasæt nu også er konstante, vil der ikke forekomme nogle ændringer i fit til dette

datasæt fremover.

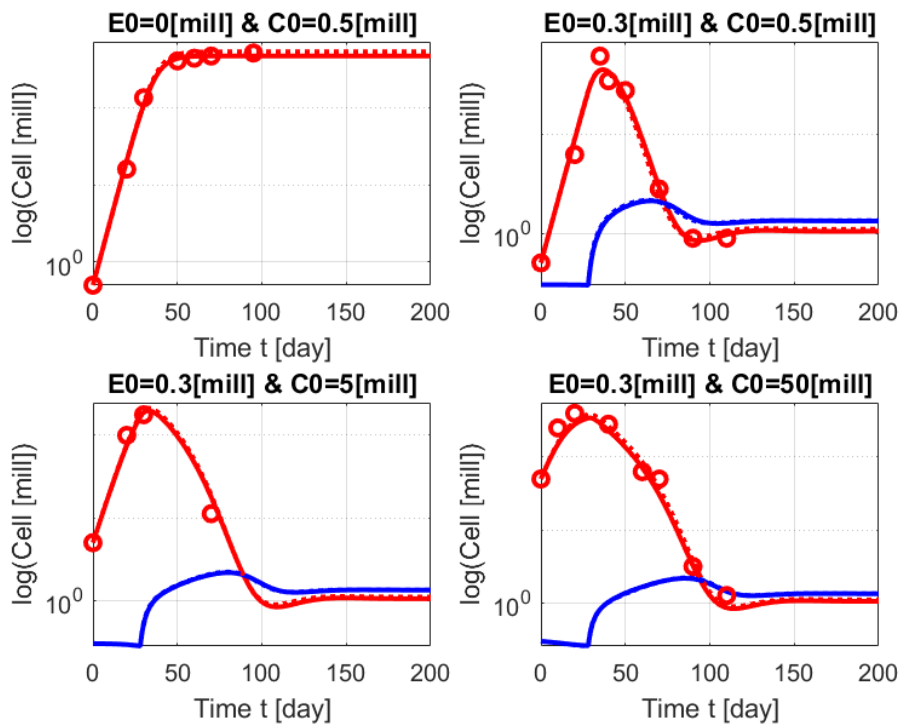
Med metoden beskrevet tidligere, udføres først en parameterestimering, hvor alle parametre holdes variable, hvorved der opnås en direkte sammenligning med den parameterestimering, som er foretaget af KK. Det resulterende fit, det originale fit, samt det tilhørende datasæt er vist på figur 3.7.



Figur 3.7: Parameterestimering med alle parametre som variable. De røde linjer viser mængden af cancerceller, mens de blå viser T-celler. De stiplede linjer er fittet med KK's fundne parametre, mens den solide linje er fittet med de nye fundne parametre. Cirklerne viser de eksperimentelt bestemte antal af cancerceller. Figuren er lavet gennem appendix C.2.

Figur 3.7 viser, at afvigelserne mellem de to udviklinger er minimale. Den mest synlige ændring, er ændringen af parameteren u , som ses gennem en forskydning langs x-aksen. Her ses det, at den store stigning i antallet af T-celler, og det resulterende fald i antallet af cancerceller, forsinkes. Derudover ses en ændring i det ækvilibrumpunkt, som de to populationer søger hen imod, hvor begge værdier er mindsket en smule.

Det ses i tabel 3.3, at ændringen i u i forhold til KK er konsekvent for parameterestimeringen ved alle de anvendte startbetingelser. Afvigelser i u er for disse ligninger ganske afgørende for modellen, da parameteren korrigerer tilførslen af et ekstra led til modellen. Dette ses også på figur 3.4, hvor u fremgår som den næstmest sensitive parameter i modellen. Ved at fastsætte u til værdien fra KK kan det undersøges hvor stor lighed der er mellem bestemmelsen af de øvrige parametre for de to estimeringsalgoritmer. Herved kan det undersøges hvorvidt KK har gjort brug af en optimeringsalgoritme som blot har fundet frem til et lokalt minimum for denne bestemte u -værdi, eller om den systematisk afviger fra den algoritme som anvendes i denne rapport.

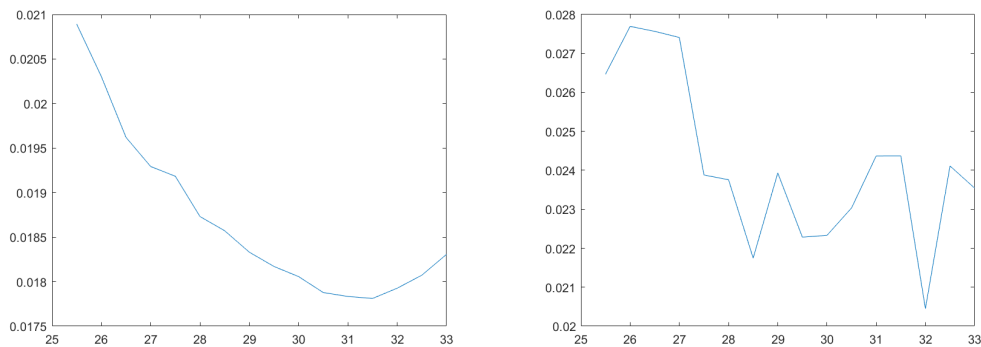


Figur 3.8: Parameterestimering med u som konstant. De røde linjer viser mængden af cancerceller, mens de blå viser T-celler. De stiplede linjer er fittet med KK's fundne parametre, mens den solide linje er fittet med de nye fundne parametre. Cirklerne viser de eksperimentelt bestemte antal af cancerceller. Figuren er lavet gennem appendix C.3.

Figur 3.8 og tabel 3.3 viser, at dette fits parameterværdier er langt tættere på værdierne fra KK, end når u holdes variabel. Dette ville kunne forklares

ved, at KK har fundet parametrenes størrelse ved et lokalt minimum for u , men ikke ved det globale minimum. En sådan situation vil for eksempel forekomme ved søgning af parameterrummet. Parameterrummet vil aldrig kunne undersøges absolut fyldestgørende og jo mere dybdegående det undersøges, jo længere tid kræves det for algoritmen at eksekvere til ende. Dette kompromis mellem grundighed og effektivitet kan dermed resultere i, at afsøgningen af parameterrummet er for begrænset til at finde det globale minimum.

For nærmere undersøgelse, kan en mindre grundig parameterestimering udføres, ved forskellige fastholdte værdier af u . Herved kan det undersøges hvorvidt denne mere udførlige parameterestimering løber ind i lignende problematikker med lokale minima omkring KK's u -værdi, når den tvinges til at foretage en mindre dækkende undersøgelse af parameterrummet. Dette er illustreret på figur 3.9b og kan sammenlignes med en lignende parameterestimering, som tillades en mere dækkende undersøgelse ved hver af disse u -værdier på figur 3.9a. På begge figurer i 3.9 er dette undersøgt ved u -værdier i intervallet $[25.5 : 33]$ med trin på 0.5. De fundne *bedste-fit*-værdier tilhørende hver u -værdi er vist på y-aksen.



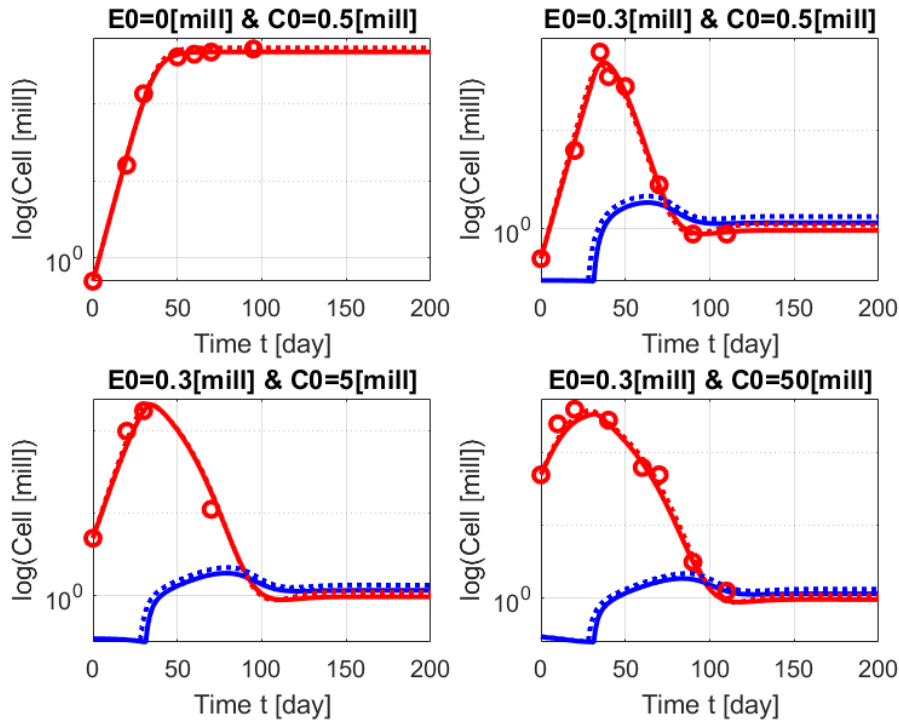
(a) *bedste fit til specifikke u -værdier, hvor algoritmen er tilladt op til 3.000 simulationer, i stedet for de 10.000 som er brugt til de resterende figurer i afsnittet.* (b) *bedste fit til specifikke u -værdier, hvor algoritmen er tilladt op til 1.000 simulationer, i stedet for de 10.000 som er brugt til de resterende figurer i afsnittet.*

Figur 3.9: *Undersøgelse af parameterestimeringens bedst fundne fit ved forskellige fastholdte u -værdier. På figur 3.9a er algoritmen tilladt 3.000 simulationer, mens der på figur 3.9b er tilladt 1.000 simulationer. Det fremgår af figur 3.9a at 3000 simulationer resulterer i en tæt på glat, parabellignende form, mens figur 3.9b har mange lokale minima. Begge grafer har globalt minimum i intervallet $[31:33]$. Figurene er lavet gennem appendix C.4.*

Som det fremgår af figur 3.9, så vil der, ved en mindre grundig analyse af parameterrum, findes et lokalt minimum tæt på den værdi som er præsenteret af KK. Dette ville kunne forklare forskellene på de fundne u -værdier, såfremt at KK har anvendt en algoritme, som undersøgte parameterrummet mindre fyldestgørende, end den som benyttes i denne rapport. Dette virker plausibelt, da artiklen blev udgivet i 2001 og den tilgængelige regnekraft i moderne computere er steget meget siden da.

Proceduren bag parameterestimeringen kan justeres på baggrund af sensitivitetsanalysen, hvor g og b blev vurderet som robuste. I forbindelse med sensitivitetsanalysen blev $(1 - p)$ også vurderet som robust. Da analysen dermed ikke er foretaget direkte på baggrund af p , så vil denne parameter ikke undersøges grundigt i denne parameterestimering, som har til formål at undersøge parametrene med udgangspunkt i den oprindelige model fra KK.

Som tidligere nævnt er b blevet estimeret i isolation, på datasættet uden nogen T-celler. Derved er kun parameteren g tilbage som en robust variabel i parameterestimeringen. Såfremt denne vurdering er fyldestgørende, så vil fittet være næsten lige så godt, når denne parameter fastsættes som konstant, som når den er variabel. I kraft af at den numeriske bestemmelse af fittets kvalitet i denne estimering tager højde for antallet af variable parametre, så bør denne værdi derfor være bedre for dette fit. g fastsættes derfor til den værdi som blev præsenteret i KK og det resulterende fit kan ses på figur 3.10.



Figur 3.10: Parameterestimering med g som konstant. De røde linjer viser mængden af cancerceller, mens de blå viser T-celler. De stiplede linjer er fittet med KK's fundne parametre, mens den solide linje er fittet med de nye fundne parametre. Cirklernerne viser de eksperimentelt bestemte antal af cancerceller. Figuren er lavet gennem appendix C.5.

Med en konstant g , finder algoritmen altså nogle parameterværdier, som resulterer i en meget lignende udvikling af cancerceller. Igen ses ændringen i værdien af u dog tydeligt, i form af forskydningen af T-cellernes stigning langs x-aksen. Jævnført analysen af parameteren g i afsnit 3.2.1 og 3.1.2 er denne overensstemmelse som forventet. Parameteren er ved alle udførte estimeringer i størrelsesordenen 10^{-1} , hvilket gør parameteren negligerel ved $t = u$ for alle graferne, hvor c er mange gange større end g . Efter tidspunktet $t = u$, er væksten af cancerceller dog faldende, således at g 's indflydelse på leddets størrelse begynder at stige. Alt i alt er der dog meget få datapunkter med lave c -værdier efter u . Kun ved fem datapunkter, befinder c sig i en størrelsesorden $[10^0 : 10^1]$, som trods alt stadig er omkring en faktor 10 – 100 større end g -værdien.

Af tabel 3.3 fremgår parameterværdierne, samt kvaliteten af deres respek-

tive fit i forhold til datasættet, for begge disse grafer, samt for KK's originale parameterestimering. Her ses det, at selvom graferne visuelt ligner hinanden, så ændres de øvrige parametre (med undtagelse u) sig væsentligt, alt efter om g fastholdes eller ej. Dette forventes ikke umiddelbart af en robust parameter og er derfor en kraftig indikation af, at g er en mere sensitiv parameter, end sensitivitesanalysen umiddelbart antydede.

Jævnført tabel 3.3 er det dog værd at bemærke, at fastsættelse af g som konstant resulterer i den bedste værdi for fittet. Konsekvensen af g 's sensitivitet er dermed ikke i sig selv et dårligt fit, men derimod en stor indflydelse på de øvrige parametres størrelser. Ydermere er der, som tidligere nævnt, mangel på data i det område, hvor g 's effekt på modellen ville ses tydeligst. Det er ikke garanteret at g ligeledes ville vurderes som robust i en sensitivitetanalyse ud fra et datasæt, som dækker perioden med cancer i *dormancy* fyldestgørende.

De fundne parameterverdier for hver af de udførte parameterestimeringer kan sammenlignes med udgangspunkt i tabel 3.3

Fit	Bedste fit	a	p_e	u	v	g	d
Original	0.1070	0.139	0.180%	28.1	0.525	0.161	0.591
Fastsat:							
Ingen	0.0944	0.159	0.259%	31.4	0.630	0.148	0.729
u	0.0971	0.140	0.191%	28.1	0.488	0.152	0.550
g	0.0925	0.162	0.243%	31.3	0.586	0.161	0.678

Tabel 3.3: *Parameterverdier og bedste fit, fundet gennem parameterestimering med forskellige variable fastsat. Disse sammenlignes med parameterverdier fundet af KK, samt bedste-fit-værdier tilhørende disse parameterverdier. Bedste fit for KK er beregnet gennem appendix C.6.*

Som det fremgår på tabel 3.3, er *bedste fit* i alle situationer bedre med de korrigerede parameterverdier, end de som blev præsenteret af KK. Dette underbygger antagelsen om, at KK har anvendt en algoritme som foretog en ringere undersøgelse af parameterrummet, end den som er benyttet i denne rapport. Derudover ses det, at der forekommer minimale ændringer i parameteren k gennem de mange estimeringer, hvilket er et stærkt tegn på, at det er den korrekte parameterværdi for denne model, som er fundet.

I begge estimeringer hvor u har været tilladt variation, har den søgt hen til værdier meget tæt på hinanden. Da der kun er foretaget to estimeringer

med denne parameter fri, er dette dog ikke nok til at kunne konkludere noget entydigt omkring denne parameter.

Parametrene a , p , v og d ser store ændringer mellem de forskellige estimeringer, hvilket fortæller os at dette nok er de korrigerende parametre, som kan gøre op for en fastsat parameter. Altså kunne det tænkes, at der var en slags sammenhæng mellem disse parametres effekt på modellen, og de parametre som fastsattes. Såfremt en parameter fastsættes, som har en negativ effekt på antallet af cancerceller, ville det tyde på, at parameterestimeringen bruger en kombination af ændringer i disse parametre til at korrigere udviklingen.

Det fremgår tydeligt i tabel 3.3, at den robuste parameter g antager forskellige værdier, blandt de forskellige fits. Derudover finder modellen også hen til vidt forskellige parameterstørrelser, for for størstedelen af de resterende parametre, med en gennemsnitslig ændring af 8,18%. Dette kunne tyde på, at selvom modellen er robust overfor ændringer i kun denne parameter, er den sensitiv overfor g 's størrelse, når den varieres i kombination med andre parametre. Dette er ikke en mulighed som blev undersøgt i sensitivitetsanalysen, da parametrene kun blev varieret enkeltvist. Som tidligere nævnt, ses modellens sensitivitet overfor g 's størrelse samtidig ikke ved størstedelen af datapunkterne. Såfremt flere datapunkter blev introduceret for perioden hvor modellen er stabiliseret, ville det ses mere tydeligt, at den er sensitiv overfor ændringer i dennes størrelse.

På baggrund af disse observationer, kan det konkluderes, at til trods for resultaterne fra sensitivitetsanalysen, så kan g ikke betragtes som en parameter af lav betydning eller sensitivitet for det system som KK forsøger at modellere. I samme ombæring er det anvendte datasæt dog dårligt egnet til at estimere g 's værdi, da der er mangel på datapunkter i det område af systemet, hvor g 's indflydelse kommer klarest til udtryk. Det vil sige at g er en vigtig parameter for modellen, men at tilhørende parameterværdier, som estimeres ud fra det givne datasæt vil være relativt upålidelige.

I denne sektion har vi undersøgt hvad der sker når der ændres på de individuelle parameters værdier. Vi startede med at vise at parameterene d , u , v og k alle er sensitive, hvilket indebære at de resterende parametre p_e , b og g , relativt set, alle er robuste. Ydermere viser vi visuelt via grafer og numerisk, hvor meget modellens output ændres når parameterene ændres med $\pm 10\%$. Med denne viden laver vi vores egen parameterestimering. Vores og KK's estimering ligger tæt op af hinanden, men vores parametre giver et bedre fit.

3.3 Modelkorrigerings

I denne sektion vil vi undersøge en række valg, som KK har taget i løbet af deres modelleringsproces. Først vil vi undersøge et alternativ til den betingede erklæring. Alternativet består i indførelsen af et logistisk led, som er i stand til at emulere den betingede erklæring funktionelt. Herefter vil vi foretage en overfladisk undersøgelse af hvordan modellen ville påvirkes af tilføjes af et stokastisk led. Der vil derefter foretages et forsøg på en stor forsimples af modellen ved at foretage en analyse ved kvasistabil tilstand af $e(t)$. Til sidst vil vi analysere hvordan differensligninger vil fitte til datasættet. I den forbindelse vil systemet samtidig undersøges med en tidsforskydning i forhold til KK's model. Såfremt at der opstår interessante ændringer i modellen, kan dette tages op til diskussion i næste kapitel.

3.3.1 Betinget erklæring

KK valgte at tage en betinget erklæring med i udformningen af deres matematiske model. Formålet med dette er at inkorporere en tidsforsinkelse der beskriver en tilvækst af T-celler som starter efter en vis periode, hvori denne tilvækst ikke forekommer. Dette er gjort med f -leddet som er beskrevet tidligere i rapporten.

Afsnittet er bygget op omkring en analyse af betingelsen, og dets matematiske betydning for modellen, for derefter at vise hvordan et andet valg forsimpler deres tilgang. Det alternative valg er et logistisk led. Med det alternative led vil vi foretage en parameterestimering, således at vi kan sammenligne kvaliteten af dette valg på baggrund af den originale model. Alternativets implikationer på forskellige aspekter af modellens virke, såsom overfitting og om det er en mekanistisk eller fænomenologisk model, vil også udforskes.

For at kunne finde et alternativ til den betingede erklæring, skal den effekt som dette har på modellen, samt den bagvedliggende tankegang som det bygger på, analyseres.

Denne procedure kan betragtes som et eksempel på en forsimples af en model. Argumentet er at alternativet skal beskrive data, i tilfredsstillende grad, sammenlignet med KK's model, således at dette ikke forringer modellens egnethed. Resultatet er derfor en model som er mere transparent og

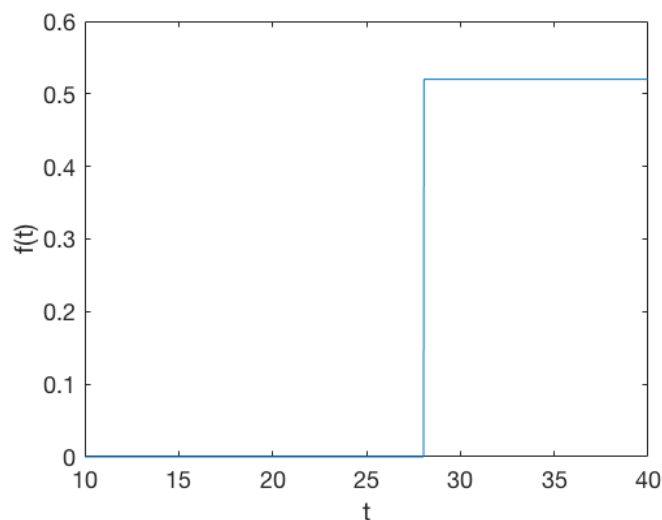
bedre i syntaks med det fænomen den modellerer.

I KK's model, kan den betingede erklæring som definerer funktionen f , beskrives som en funktion som antager værdien 0, på intervallet $t \in [0 : u]$, og værdien v på intervallet $t \in [u : t_{slut}]$.

Denne opførsel kan tilnærmes med en meget stejl logaritmisk funktion. En logaritmisk funktion har den matematiske form:

$$f(t) = \frac{L}{1 + e^{-h \cdot (t - t_0)}} \quad (3.12)$$

Her bestemmer parameteren L funktionens toppunkt, parameteren h bestemmer hældningen af funktionen og t_0 beskriver x-værdien for punktet hvor $f(t) = \frac{L}{2}$. Denne funktion vil tilnærme sig funktionen f som tidligere i afsnittet blev defineret, med $L = v$ og $t_0 = u$ ved $\lim_{h \rightarrow \infty}$.



Figur 3.11: Logistisk funktion med $L = 0.525$, $h = 1000$ og $t_0 = 28$. Figuren er lavet gennem appendix D.1.

På figur 3.11, ses en logistisk funktion med $L = v$, $h = 1000$ og $t_0 = u$. Dette eksemplificerer, hvordan den logistiske funktion med et højt h , tilnærmer sig opførslen af KK's betingede erklæring. Anvendes denne logistiske funktion, som erstatning for den betingede erklæring, vil modellen forsimples. Den betingede erklæring fjernes, og den eneste nye parameter som introduceres, kan ses som en konstant, så længe en stor værdi vælges.

Med implementering af denne logistiske funktion i modellen, ville diffe-

rentialligningssystemet se således ud:

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot e(t) \cdot c(t) \quad (3.13)$$

$$e'(t) = r + \frac{\frac{v}{1+e^{-1000 \cdot (t-u)}} \cdot c(t)}{g + c(t)} \cdot e(t) - d \cdot e(t) - a \cdot (1 - p) \cdot c(t) \cdot e(t) \quad (3.14)$$

Herved er det klart, at dette vil forsimplere modellen, da den betingede erklæring fjernes, til fordel for et udtryk som ikke introducerer nogen nye parametre, men blot en enkelt konstant. Ydermere er der en vis uklarhed omkring den tiltænkte funktion og formål af den betingede erklæring i KK; specielt i forhold til dets værdi ved dag 0. Denne uklarhed vil undgås ved konvertering af den betingede erklæring til denne logistiske funktion, hvis opførsel er tilnærmelsesvist ens. Ulempen ved dette nye udtryk er, at det giver modellen en eksplicit tidsafhængighed, og det derved bliver et ikke-autonomt system. Derved udelukkes brugen af de analytiske værktøjer, som ikke er mulige at bruge på sådanne systemer.

Variation i modningstiden af T-celler

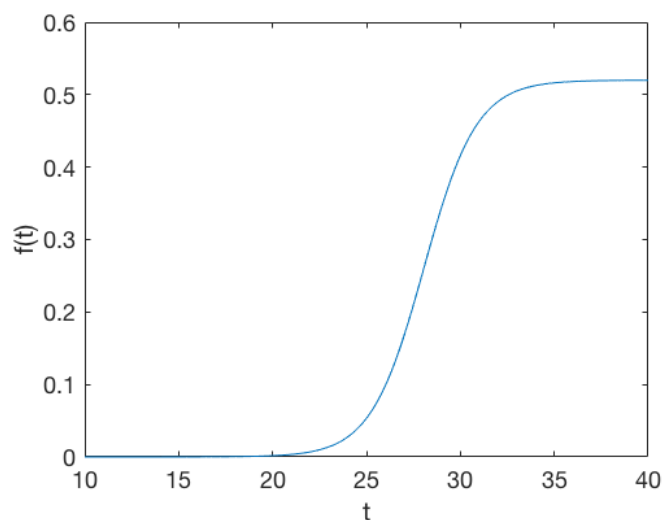
Ved denne erstatning for udtrykket af f er det muligt at introducere en smule spredning i modningstiden af T-celler. Da der er tale om en biologisk proces, er det usandsynligt at processen altid tager samme tid. Ved at gøre den logistiske funktions parameter h variabel, vil dette kunne inkorporeres i det logistiske led, afspejlet ved en mindre stejl kurve. Derved distancerer funktionen sig fra tilnærmelsen af det conditional statement f , som dikterede at der ikke forekom nogen modnede T-celler før perioden u var overskredet. Ved ændring af h , vil der forekomme en tilvækst af T-celler gennem modning af nye celler i respons til cancercellerne, som starter før den gennemsnitlige modningstid. Derved er det muligt at undersøge hvordan en variation i modningstiden af T-celler, vil påvirke modellens output.

Først er det nødvendigt at tage stilling til, hvor stor en betydende variation omkring KK's 28 dage, det ville være meningsfyldt at introducere. Da variationen ikke introduceres på baggrund af fastlagt teoretisk viden om systemet, kan den ikke bestemmes på baggrund af dette.

I parameterestimeringsafsnittet blev det fundet, at KK's u -værdi af cirka 28 dage, ikke stemte overens med den værdi som parameterestimeringen kom frem til. Denne værdi blev fundet til omkring 31 dage i parameterestimeringen, og efter undersøgelse af KK's tidligere udgivelser blev det fundet at der

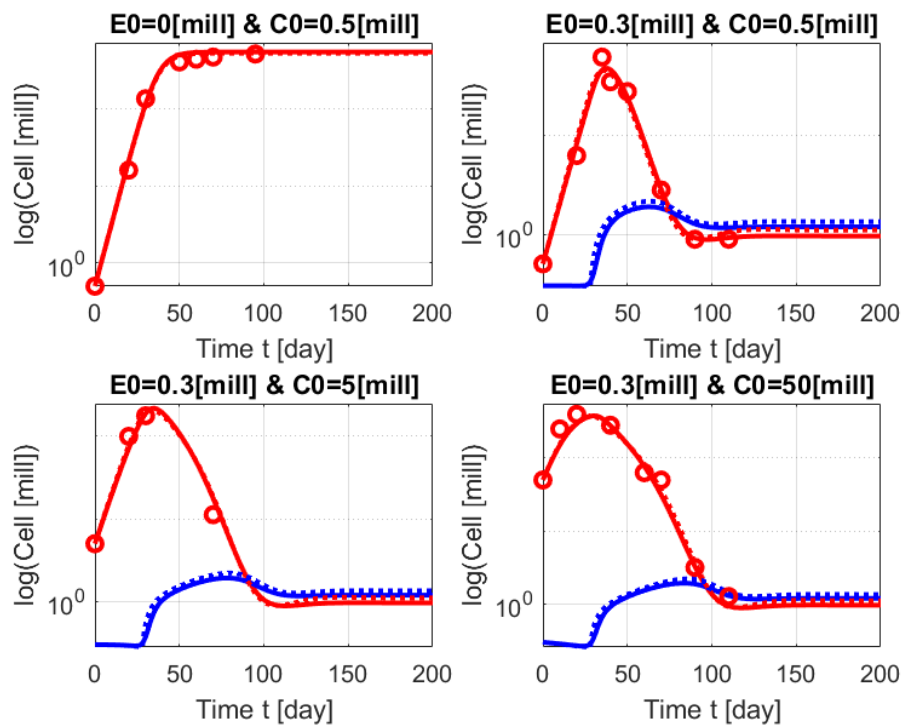
tidligere har været brugt en værdi af 24, til at beskrive noget lignende. Med udgangspunkt i den manglende teoretiske viden, benyttes disse overvejelser som grundlag for at anvende en eksemplarisk spredning på tre dage.

Den logistiske funktion vil aldrig kunne være en helt præcis imitation af den betingede erklæring, da funktionen aldrig når 0 eller 1, men blot tilnærmer sig dem uendeligt nært. Funktionens spredning begrænses i praksis af, at funktionen ikke bør have stor effekt på modellen udenfor grænsen på 3 dage omkring u . Dette krav opstilles således, at mængden i perioden $t \in [0 : 25[$ skal være 5% af mængden i perioden $t \in [0 : 31]$. Punktet hvor dette krav for den logistiske kurve er opfyldt, er ved en h -værdi på 0.711.



Figur 3.12: Logistisk funktion med $L = 0.525$, $h = 0,711$ og $t_0 = 28$. Figuren er lavet gennem appendix D.1.

Figur 3.12 illustrerer hvordan væksten fordeler sig over tid ved denne h værdi. Ved at foretage en parameterestimering for $h = 0,711$, kan spredningens effekt på den samlede model undersøges. Det resulterende fit ses i figur 3.13, sammenlignet med KK's fit, med parameterværdierne og *bedste fit* vist i tabel 3.4.



Figur 3.13: Plot af KK's model, med en logistisk funktion som erstatning af den betingede erklæring, med parameterestimering udført. De blå linjer repræsenterer T-celler, mens de røde linjer repræsenterer cancerceller. De stiplede er KK's fit, mens de solide er den nye optimering. Datasættet er vist som punkter. Figuren er lavet gennem appendix D.2.

Fit	Original	Estimering	Logistisk
Bedste fit	0.107	0.0925	0.0993
k	0.188	0.188	0.189
b	0.00188	0.00213	0.00184
a	0.139	0.162	0.158
p_e	0.180%	0.243%	0.192%
u	28.1	31.3	29.7
v	0.525	0.586	0.585
g	0.161	0.161	0.150
d	0.591	0.678	0.676

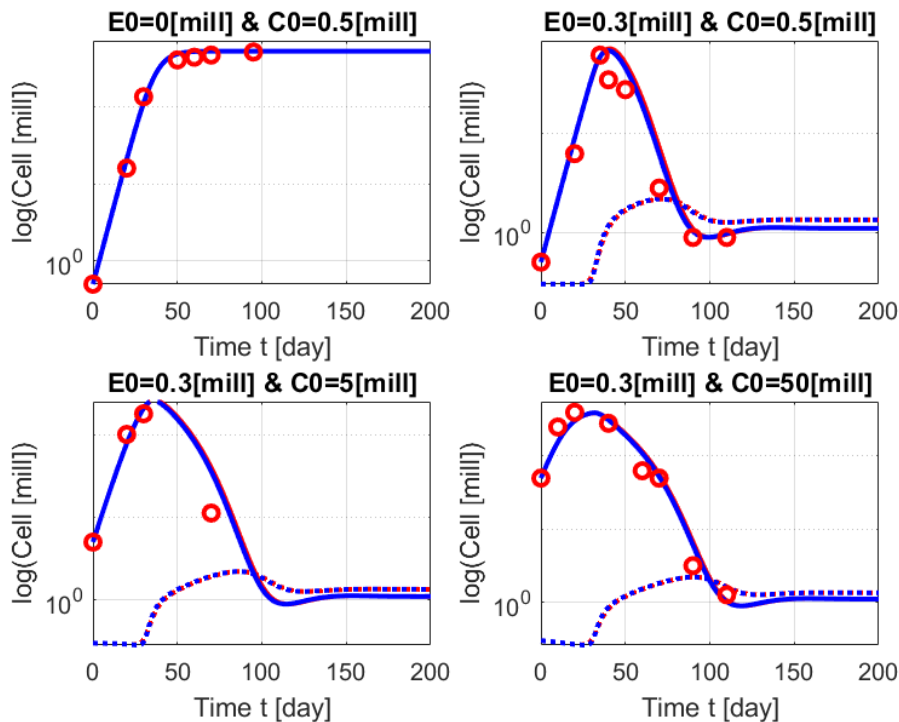
Tabel 3.4: *Parameterverdier og bedste fit, fundet gennem parameterestimering af KK's model med en logistisk funktion som erstatning af den betingede erklæring. Disse værdier sammenlignes med parameterverdier og bedste-fit-værdier fra KK, samt den estimering som resulterede i den laveste bedste-fit-værdi i parameterestimeringsafsnittet. Hældningen af det logistiske udtryk (parameteren h), er fastsat til 0.711, i det logistiske udtryk.*

Af figur 3.13 fremgår det, at der ikke forekommer væsentlige forskelle i modellen ved implementering af denne variation i modningstiden. De primære forskelle som ses på figur 3.13, er forskydningen af væksten af T-celler. Denne ændring kan dog tilskrives selve parameterestimeringen, da dette også er den primære ændring som blev set i parameterestimeringsafsnittet.

På tabel 3.4 ses det, at denne forskel i u -værdi ikke er lige så stor, som den der blev set i parameterestimeringsafsnittet. Dette kunne oplagt være et resultat af den logistiske ligning. Selvom indsættelsen af det logistiske led ikke forårsagede særligt store synlige ændringer på modellens evne til at beskrive systemet, så ændredes modellen dog nok til, at de fleste af parameterstørrelserne er forskellige fra den tidligere estimering. Derved kan de forskellige parameterstørrelser mellem **Logistisk** og **Estimering** i tabel 3.4, ses som et tegn på, at ændringen ikke er fuldstændigt irrelevant for modellen.

For endegyldigt at kunne konkludere, hvorvidt det er en egnet forsimpning, ikke at tillade variation i modningstiden, skal endnu en undersøgelse foretages. Det kan nu undersøges hvordan h i sig selv påvirker modellen, ved at variere h omkring 0.711, med de parameterstørrelser, som er tilpasset det logistiske led, fastholdt. Disse værdier er vist i tabel 3.4 under **Logistisk**. Dette gøres for at undersøge dens indflydelse på udviklingen i det spektrum som tænkes relevant for modellen. For at indsnævre undersøgelsens område, da der ikke vides hvilket spektrum den faktiske værdi ville være teoretisk,

varieres h med $\pm 20\%$. Disse to grænsetilfælde for parameteren, er vist på figur 3.14, med de resterende parameterværdier værende dem, som blev fundet i parameterestimeringen. De fundne parameterstørrelser er vist i tabel 3.4, hvor den parameterestimering med ingen fastsatte variable også indgår som referencepunkt.



Figur 3.14: Plot af KK's model, med en logistisk funktion af forskellig hældning som erstatning af den betingede erklæring. De solide linjer repræsenterer T-celler, mens de stiplede linjer repræsenterer cancerceller. De blå linjer har gjort brug af den laveste hældning og den højeste værdi til de røde linjer. Datasættet er vist som punkter.

I figur 3.14 er der vist udviklingen ved tre forskellige værdier af h . I de blå linjer, er en h -værdi af 0.853 blevet brugt, hvilket tilsvarende, at leddets værdi ved grænserne af spredningen er på 2.83% af L . I de røde er der blevet brugt en værdi af 0.569, hvilket resulterer i 8.99% før 25 dage. Disse er resultatet af 20% ændring i begge retninger fra det forrige kriterie af 0.711. Som det ses på figur 3.14, forekommer der minimale ændringer i udviklingerne, selvom parameteren er ændret med 20% fra udgangspunktet i hver deres retning. Dette viser at der ikke forekommer nogle store afvigelser gennem denne fors-

pling. Modellen bliver samtidig matematisk simple i den forstand, at den er differentiabel over hele kurven og ligningen ikke skal deles op i to. På baggrund af disse to betragtninger kan forsimplingen umiddelbart anses som værende god. Det må derved konkluderes, at selvom indførelsen af en variation i modningstiden er en mere virkelighedsnær forklaring, er ændringen i modellens output meget lille. Da der altid må tages valg i modelleringsprocessen om forsimplinger af systemet, ser dette ud som en oplagt én af slagsen.

3.3.2 Stokastisk parameter

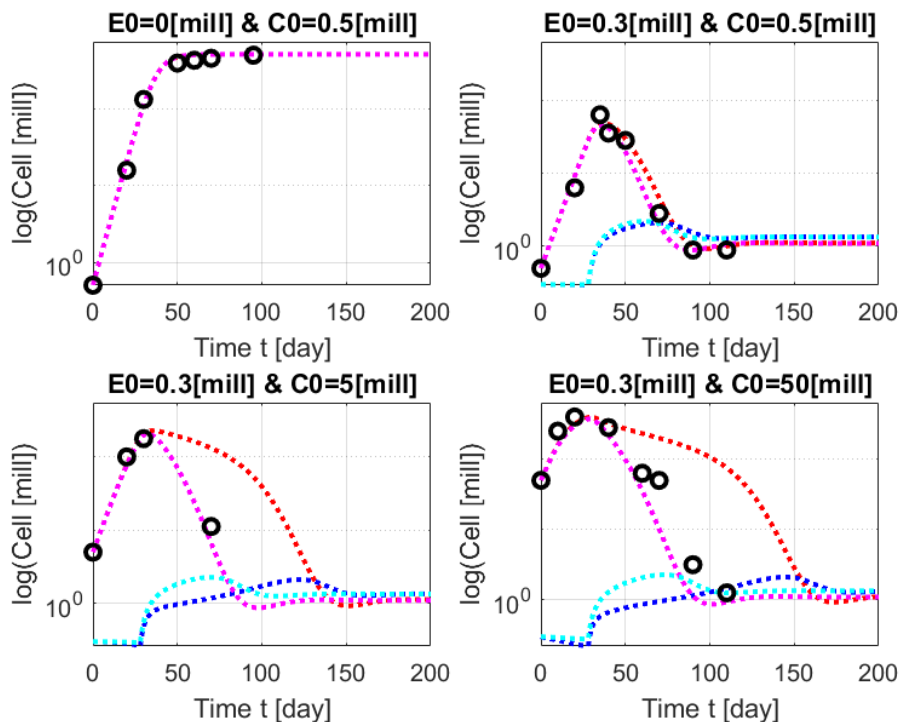
Først foretager vi en analyse af effekten af indførelsen af et stokastisk led i case-modellen. Efterfølgende bruger vi dette som argumentation for at parameteren p kan gøres stokastisk. Vi vil ikke foretage den egentlige udskiftning af p med en stokastisk parameter, men vil derimod vise dens variation i grænsetilfældene af et sandsynligheds spektrum. Dette bringer os videre til en analyse af valgets implikationer for modellens udvikling når dette er underlagt tilfældige variationer.

Parameteren p_e er den statisk procentvise sandsynlighed for resultatet af udfaldet mellem T-celler og cancercellers interaktion, resulterende i den enes død, og dermed en formindskelse i mængden af denne celle. Denne mekanisme synes oplagt at tillægges et stokastisk udfald, hvilket vil resultere i tilfældige ændringer af de kontinuerte variable. Begrundelsen for dette er, at det i et biologisk perspektiv antages troværdigt at udfaldet af mødet mellem cellerne ikke er statisk, men er præget af et tilfældigt element. Systemet er netop så komplekst at forfatterne af case studiet understreger at forståelsen af system dynamikkerne er relativt ukendte (Kuznetsov, Makalkin et al., 1994).

Der foretages ikke en egentligt udførelse af operationen med ændring i modellen og konvertering af p_e til en stokastisk variabel, derimod vil det blot undersøges, hvordan udviklingen ændrer sig hvis p_e varieres i intervallet $[0.001 : 0.004]$. Intervallet er valgt med begrundelse i at p_e er restriktiv ($0 \leq p_e \leq 1$). Samtidig er en p_e -værdi på 0 uinteressant at undersøge med henblik på tilfældig variation, da dette vil tilsvare et fuldstændigt perfekt immunforsvar. $[0.001 : 0.004]$ tilsvare henholdsvis en halvering og en fordobling af T-cellernes effektivitet, som den er præsenteret af KK. Det vælges kun at simulere hvordan modellen opfører sig i de to yderpunkter af intervallet for at vise hvordan modellen responderer i grænsetilfælde. Kurverne for disse funktioner vil dermed udgøre det maksimale spektrum, hvori data ville

kunne forventes at findes inden for disse grænser. De opnåede resultater, som vil blive vist i figurer, vil danne grundlag for en kvalitativ analyse af effekten af variationerne.

Figur 3.15 viser disse to scenarier for hver af de forskellige startbetingelser for e_o og c_o , sammenlignet med datasættet.



Figur 3.15: Eksemplificering af variation i parameteren p_e . Der er vist cancer og T-celler for to forskellige p_e -værdier: $p_e = 0.1\%$ og $p_e = 0.4\%$. Den stiplede røde linje viser cancervækst ved $p_e = 0.4\%$, mens den mørkeblå stiplede linje viser T-cellevækst ved samme p_e -værdi. Den stiplede lilla linje viser cancercellevækst ved $p_e = 0.1\%$, mens den stiplede cyan linje viser T-cellevækst ved samme p_e -værdi. Såfremt en variation i parameteren blev introduceret med disse grænser, ville cancercellernes vækst ligge i området som den røde og lilla linje. Det samme ville være gældende for T-cellerne, som ville ligge mellem den blå og cyan linje. Figuren er lavet gennem appendix E.

Det første der er værd at bemærke er at en ændring til $p_e = 0.1\%$ ikke får cancercellerne eller T-celler til at afvige i stor grad fra udviklingen set på figur 3.7 i parameterestimeringen. I de fire forskellige situationer, med

forskellig c_0 , er tendensen at toppunktet for cancercellerne er forskudt en smule bagud, mens T-cellerne ser ud til at stige over en længere periode, og med større hastighed. Dette medfører at ækvilibriumsstadiet hurtigere nås, for denne variation.

Når p_e -værdien varieres til $p_e = 0.4\%$, dominerer cancercellernes vækst betragteligt, især når $c_0 = 5$ og $c_0 = 50$. Effekten af dette er en forsinkelse af ækvilibriumpunktets indtræffen. Ækvilibriumpunktet indtræffer dog ved omtrent samme værdi, til trods forskydningen. Dette indikerer at dette stadie altid vil indtræffe når den stokastisk parameter varieres indenfor et realistisk spænd, og at modellen er deterministisk i denne henseende. Dette er med til at vise et naturligt systems skiftende udvikling, og giver et mere repræsentativt indtryk af det undersøgte fænomen. En stokastisk parameter, som varierer p_e -værdien i dette spektrum, ville derfor resultere i en model som afspejler den observerede udvikling nærmere.

For at få et virkeligt billede af implikationerne af et stokastisk led, ville det være nødvendigt med indførelsen af dette, samt at foretage i tusindvis af simulationer, hvorved en fordeling af disse resultater kunne give et mere helstøbt billede af systemets natur. Denne fremgangsmåde er derfor kun ment som et indledende redskab til at undersøge virkningen af et stokastisk led, ved yderpunkterne af en sandsynlighedsdistribution, som viser spændet af udfald når dette er aktualiseret.

For at indførelsen af en stokastisk parameter er interessant, kræver det dog at der er data tilgængeligt, som viser en variation i udfaldet, som kunne være resultatet af en tilfældighed. Da dataen brugt af KK, er baseret på gennemsnit, er den stokastiske parameter ikke brugbar til at beskrive fænomenet.

3.3.3 Kvasistabil tilstand

Vi vil i dette afsnit opstille en forsimplet model, funderet i metoden fra en ækvilibriumsanalyse af differentiallyigningssystemet. Dette gøres ved at isolere $e(t)$ i ligning (3.2) og indsætte det i ligning (3.1). Vi udfører derefter en parameterestimering med den nye enklere model, som bruges til at sammenligne dennes udvikling med KK's models. Resultatet af dette vil så ligge op til en diskussion om hvorvidt dette trin er meningsfuldt.

For at kunne opstille denne forsimplede model, foretages der en algebraisk manipulation af ligningen for $e'(t)$, når $e'(t) = 0$, således at denne udtrykkes

som kun en funktion af c .

Dette gøres ved at isolere $e(t)$ i ligning (3.2) og indsætte udtrykket for $e(t)$ i ligning (3.1). Dermed påvirker T-cellerne $c'(t)$ med en konstant effekt som nu kun er en funktion af mængden af cancerceller, hvilket fjerner nødvendigheden for at gøre brug af en differentiallyigning til at undersøges deres udvikling. Denne forsimpelse udføres på det grundlag, at ændringerne i $e'(t)$ er langt mindre end ændringerne i $c'(t)$. Handlingen vil primært være fordelagtig, såfremt at $e'(t)$'s relativt små afvigelser fra sit ækvilibrium har en tilsvarende lille effekt på udviklingen af $c'(t)$.

Der er flere fordele ved denne forsimpelse, blandt andet at en algebraisk ligning er nemmere at løse numerisk og analytisk end en differentiallyigning. Dette afføder dog et modelleringsdilemma omhandlende formål for model, da almen nytte, virkelig repræsentation, og præcision trækker i hver deres retning. Et mere simpelt og alment system svækker dermed modellens præcisionsevne (Haefner, 2005).

Argumentet skal findes ved at parameterestimere for ligning (3.1), hvor ligning (3.2) er konstant. Såfremt det skridt resulterer i en lille ændring af systemet, så kan det betragtes som en fornuftig forsimpelse.

Først isoleres $e(t)$. Ligning (3.2) sættes lige med nul, og e isoleres ved dette ækvilibriumspunkt for T-celler.

$$e'(t) = r + \frac{f(t) \cdot c(t)}{g + c(t)} \cdot e(t) - d \cdot e(t) - a \cdot (1 - p) \cdot c(t) \cdot e(t) = 0$$

Derefter isoleres $e(t)$ så det kan udtrykkes på følgende måde:

$$e(t) = \frac{-r}{-d - a \cdot (1 - p) \cdot c(t) + w} \quad (3.15)$$

Hvor

$$w = \frac{f \cdot c(t)}{g + c(t)} \quad (3.16)$$

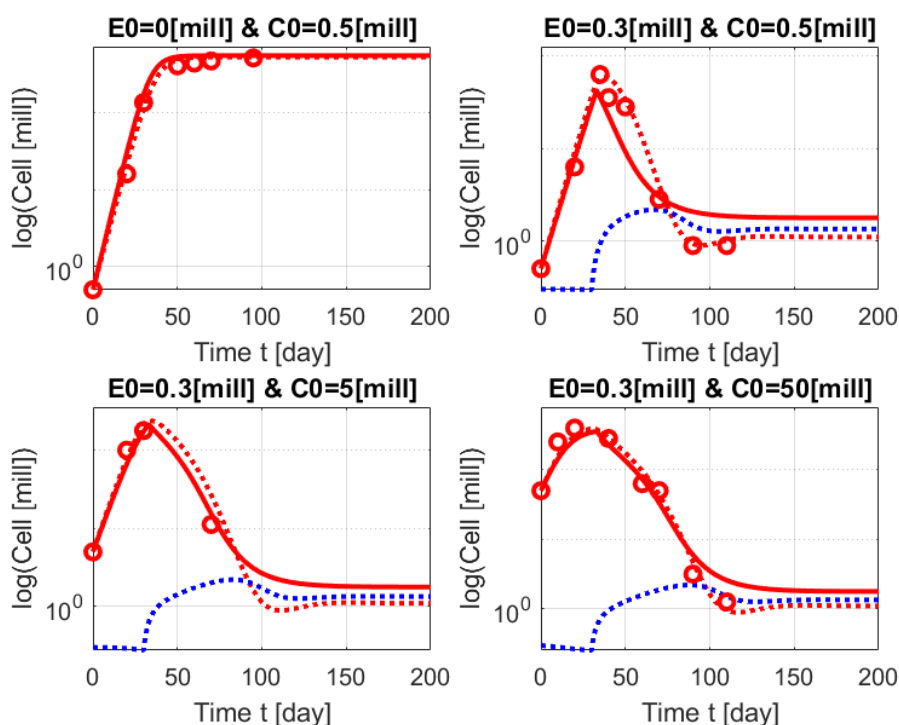
Afsluttende indsættes ligning (3.15) i ligning (3.1):

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot \frac{-r}{-d - a \cdot (1 - p) \cdot c(t) + w} \cdot c(t) \quad (3.17)$$

Derved er der opstillet en enkelt differentiallyigning, som kan beskrive udviklingen af cancerceller, uden at det nødvendiggøres, at den indgår i et

differentialligningssystem med T-cellerne.

Med denne nye model, kan der nu foretages en parameterestimering, således at modellen kan fittes til datasættet. Derved vil det være muligt at se konsekvenserne af denne forsimpning på modellen, således at en diskussion af disse konsekvenser kan have.



Figur 3.16: Parameterestimering for den forsimplede model, hvori T-cellernes differentialligning er fjernet. De røde linjer beskriver væksten af cancerceller, mens den blå beskriver T-celler. De to stiplede linjer er KK's fit, mens de solide linjer er den nye models simulerede udvikling. Datasættet er vist som punkter. Figuren er lavet gennem appendix F.

Som fremgår på figur 3.16, resulterer denne forsimpning i en langt mere spids top, omkring tidspunktet u . Da antallet af T-celler nu er en funktion af cancerceller, forekommer der ingen glidende overgang når den betingede erklæring træder i kraft, hvori T-cellerne skal vokse. T-cellernes påvirkning på cancercellerne ses mindre nuanceret, men stadig forekommer dette skift, hvori cancercellerne går fra at være stigende, til at være faldende. På figur 3.16 fremgår det, at den nye model beskriver de forsøg, hvori en stor mængde

af cancer er brugt, beskrives mest nøjagtigt. Dette kunne tænkes at være grundet i, at modellen har problematikker med at beskrive de situationer, hvori antallet af cancerceller og T-celler er tæt på ens. I figuren uden nogen T-celler er der selvfølgelig ingen forringelse af fittet, da ændringen ikke træder i spil i denne situation.

Fit	Original	Kvasi stabil
Bedste fit	0.107	0.331
a	0.139	0.126
p_e	0.180%	0.209%
u	28.1	30.85
v	0.525	0.527
g	0.161	0.144
d	0.591	0.553
k	0.188	0.207
b	0.00188	0.00193

Tabel 3.5: *Parameterverdier og bedste fit, fundet gennem parameterestimering af KK's model i en kvasistabil fase. Disse værdier sammenlignes med parameterverdier og bedste-fit-værdier fra KK's originale model.*

Som fremgår i tabel 3.5 og som kan ses på figur 3.16, opnår denne nye model ikke nær så godt et fit, som KK's originale model gjorde. Dette kunne dog forventes, da forsimplingen af at reducere modellen til blot én differentialligning ikke er en beskeden ændring. Grundet i den store ændring i modellen, er parameterestimering også nået frem til et sæt af parameterverdier, som ikke stemmer overens med de fundet til den forrige model. Disse parameterstørrelser kan dog ikke sammenlignes direkte, da modellen er undergået så store ændringer gennem denne reduktion.

Forskellen i kvalitet af fit mellem de 3 forskellige startpopulationer af cancerceller, kunne antyde at denne model er bedst egnet til at beskrive situationer med store mængder af cancerceller. Dette kan dog ikke verificeres til at være sandt, før en sammenligning med et nyt datasæt er foretaget. Da ingen sammenligning med andre datasæt er foretaget, kunne det lige så vel være en tilfældighed at det bedst fundne fit, beskriver de store populationer bedst. Derudover kunne problematikken med overfitting også komme ind i billedet, når denne forsimples brugbarhed skal tages under mikroskopet. Det kunne meget vel være tilfældet, at algoritmen blot har fundet nogle parametre som beskriver datasættet bedst, og at dette ved et tilfælde har resulteret i et bedre fit på de grafer med mange cancerceller.

Denne forsimplede model vil aldrig kunne bruges, som en erstatning af KK's model, grundet i dennes forværrede fit til det generelle fænomen. Såfremt interessen ligger inden for situationer med store mængder af cancer-celler, kunne denne model dog tages til overvejelse. Dette ville dog kræve at denne models gode fit til disse situationer, er globalt gældende, og ikke blot for dette datasæt.

3.3.4 Differensligning

Dette afsnit indledes med at opstille en differensligning som beskriver det samme system som KK's artikel undersøger. Herefter introduceres en tidsforskydning i f -leddet, således at det refererer tilbage til mængden af cancer-celler da processen blev startet. Dette gøres for bedre at kunne beskrive udviklingen ud fra de naturlige mekanismer. Til at opstille modellen, blev der foretaget en parameterestimering med den samme algoritme som tidligere er benyttet. Modellen blev derefter undersøgt for at vise hvordan den kunne beskrive det samme data. Efterfølgende foretages en sammenligning med KK's model, da resultaterne af dette vil lægge op til en diskussion af valg af modeltype.

Som tidligere nævnt i den kvalitative analyse af KK's differentiaalligningsmodel, så er både opsætningen og funktionen af f -leddet præget af en vis uklarhed. I den forbindelse blev det konkluderet, at det umiddelbart ville være mere virkelighedsnært, gennem den forklaring som KK gav for den bagvedliggende mekanisme, at cancercellepopulationen som leddet refererer til, var den mængde som var til stede i systemet for u dage siden.

I denne rapport er dette valgt at blive introduceret ved at tilpasse KK's model til et differensligningssystem, da differensligningers naturlige egenskaber resulterer i, at man har nem adgang til populationsstørrelser ved tidligere tidstrin. Derved kan man indføre et led i ligningen, hvori der refereres tilbage til populationsstørrelsen på et tidligere tidspunkt. Den første ændring som dette vil nødsage er, at u skal være et heltal, da differensligninger kun løses til diskrete tidstrin, og derved ikke har nogen løsninger ud over for tiden $t \in \mathbb{N}$.

For at opstille KK's model som differensligninger, skal hver af differentiaalligningerne opstilles således, at de ikke blot beskriver ændringen i systemet,

men den totale population til næste tidstrin. Derved bliver ligningerne:

$$c_{t+1} = c_t + k \cdot c_t \cdot (1 - b \cdot c_t) - a \cdot p \cdot e_t \cdot c_t \quad (3.18)$$

$$e_{t+1} = e_t + r + \frac{f \cdot c_t}{g + c_t} \cdot e_t - d \cdot e_t - a \cdot (1 - p) \cdot c_t \cdot e_t \quad (3.19)$$

Med $e_{t=0}$ som antallet af T-celler ved starten af sygdomsforløbet, og $c_{t=0}$ som antallet af cancerceller ved starten af sygdomsforløbet. Som beskrevet tidligere, er en differensligning en ligning, hvori mængden efter et diskret tidstrin er defineret som en rekursiv funktion af det forrige tidstrin. Derved er der tilføjet et led i hver ligning, hvor ændringen bliver lagt til, eller trukket fra, den mængde som var tilstede i det forrige tidstrin.

Der kan nu introduceres en tidsforskydning i differensligningen, således at f-leddet referer u dage tilbage:

$$c_{t+1} = c_t + k \cdot c_t \cdot (1 - b \cdot c_t) - a \cdot p \cdot e_t \cdot c_t \quad (3.20)$$

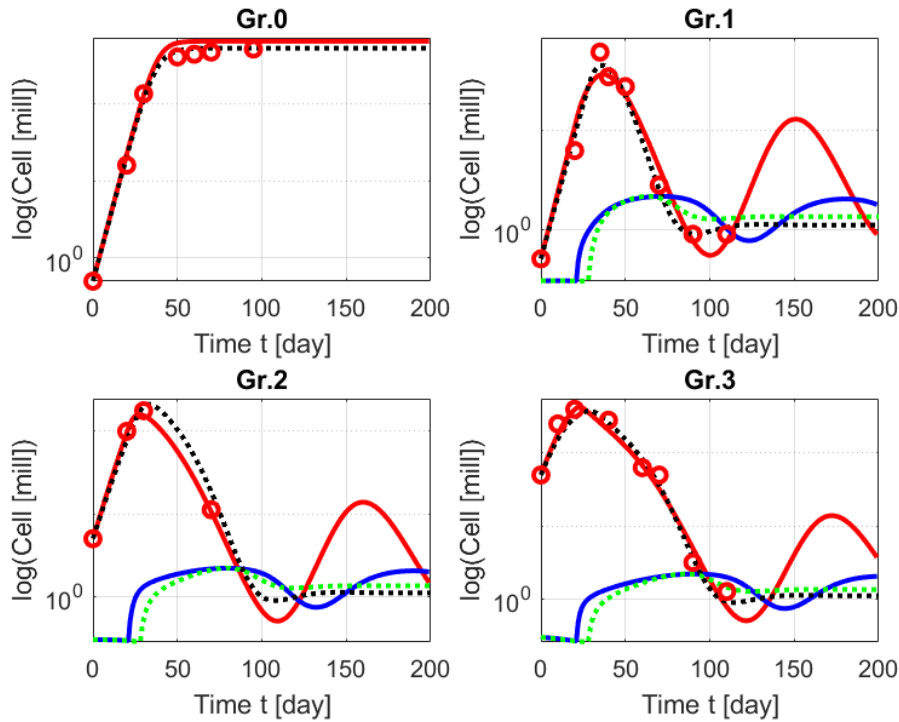
$$e_{t+1} = e_t + r + \frac{v \cdot c_{t-u}}{g + c_{t-u}} \cdot e_t - d \cdot e_t - a \cdot (1 - p) \cdot c_t \cdot e_t \quad (3.21)$$

Den eneste ændring af ligningsystemet, er i ligning (3.21), i dette led:

$$\frac{v \cdot c_{t-u}}{g + c_{t-u}} \cdot e_t \quad (3.22)$$

Hvori der nu refereres til mængden af cancerceller for u dage siden, således at der på dag u , produceres en mængde af t-celler, som afhænger af antallet af cancerceller ved starttidspunktet. Da processen stadig ikke skal starte før u dage er overskredet, og der ikke er data før $t = 0$, skal der tages stilling til tilfældet hvor $t < u$, hvor der refereres til et negativt tidstrin. Her defineres det, at $c_x = 0$, for $x < 0$, således at leddet med c_{t-u} bliver 0, indtil $t = u$ (ligesom det conditional statement i KK's model dikterer).

For at undersøge hvordan dette ændrer modellens dynamikker, kan en parameterestimering udføres, og den tilhørende udvikling ved disse parameterverdier kan findes. Dette er vist på figur 3.17.



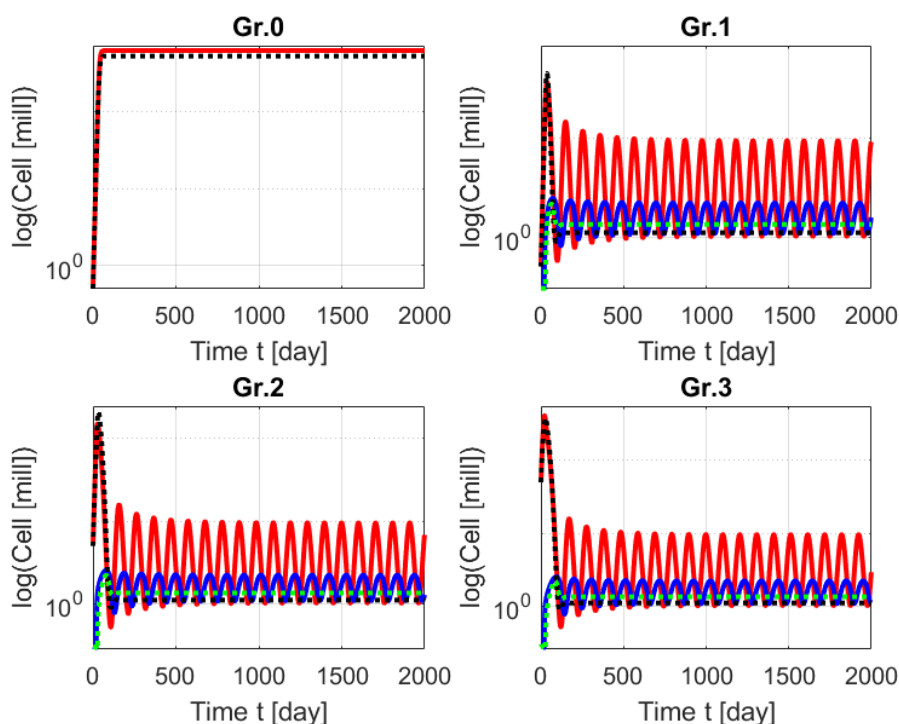
Figur 3.17: Parameterestimering for differensligning med tidsforsinkelse introduceret i modellen, over 200 dage. De stiplede linjer er fittet med KK's fundne parametre, mens de solide linjer er differensligningens bedst fundne fit til datapunkterne. Den røde linje viser cancerceller i differensmodellen, mens den blå viser T-celler i samme model. Cancercellerne i KK's model vises med en sort linje, mens T-cellerne vises med en grøn. Cirklerne viser de eksperimentelt bestemte antal af cancerceller. En bølgende mængde af cancerceller og T-celler ses gennem simuleringen. Figuren er lavet gennem appendix G.1.

Det fremgår af figur 3.17, at en introduktion af tidsforskydning resulterer i en bølgeform for begge populationer. Dette skyldes den store stigning af cancerceller i tidsrummet $t \in [0 : u]$. I den forbindelse er det relevante led:

$$\frac{v \cdot c_{t-u}}{g + c_{t-u}} \cdot e_t$$

Den numeriske størrelse af dette led vil stige i tidsrummet $t \in]u : 2u]$, hvorefter den vil falde med antallet af cancerceller. Som konsekvens af denne bølgende mængde af T-celler, vil antallet af cancerceller også antage denne form, da de påvirkes af antallet af T-celler.

For at undersøge hvorledes denne bølgeform påvirker modellens output i en generel forstand, undersøges udviklingen over en længere periode, således at udviklingen af bølgenes amplitude kan undersøges. Dette er vist på figur 3.18



Figur 3.18: Parameterestimering for differensligning med tidsforsinkelse introduceret i modellen, over 2000 dage. De stiplede linjer er fittet med KK's fundne parametre, mens de solide linjer er differensmodellens bedst fundne fit til datapunkterne. Den røde linje viser cancerceller i differensmodellen, mens den blå viser T-celler i samme model. Cancercellerne i KK's model vises med en sort linje, mens T-cellerne vises med en grøn. Det ses, at den bølgende form af begge differensligninger reduceres i amplitude ved starten af processen, men finder hen til et slags "bølgende ækvilibrium" med konstant amplitude og frekvens. Figuren er lavet gennem appendix G.2.

Som fremgår af figur 3.18, reduceres amplituden af bølgerne kun i den første del af forløbet, hvorefter de oscillerer i en fast periode med en fast amplitude.

Det kan nu undersøges hvorledes parameterværdierne blev ændret, efter diskretiseringen af differentialligningen. Det forventes at parameterværdierne

vil være vidt forskellige, på grund af de diskrete tidstrin i differensligning, kontra den kontinuerte opdatering udvikling af hældningen i differentialligningen. Disse parameterverdier og *bedste fit* fundet til differensligningen, sammenlignes med de parameterverdier tilhørende KK, samt *bedste fit* for disse er vist i tabel 3.6. Derudover er der også vist det bedste fit fundet under parameterestimering i tabellen.

Fit	Differens	Original	Optimeret
Bedste fit	0.1474	0.1070	0.0948
k	0.210	0.188	0.184
b	0.00154	0.00188	0.00188
a	0.139	0.139	0.148
p_e	0.215%	0.180%	0.228%
u	21.0	28.1	31.2
v	0.771	0.525	0.661
g	0.240	0.161	0.137
d	0.884	0.591	0.759

Tabel 3.6: Sammenligning af parameterverdier og bedste fit, for parameterestimering af differensligning med tidsforsinkelse, den optimerede differentialligningsmodel samt for de værdier præsenteret af KK.

Det ses her, at en differensligning med tidsforskydning opnår en dårligere *bedste fit* til beskrivelse af datasættet, end både den oprindelige model af KK, samt deres model med korrigerede parameterverdier. Det fremgår også af figur 3.17, at modellen har problemer med at opnå en tilstrækkeligt spids top til at ramme toppunkterne, og stadig kunne følge den nedadvendte kurve som efterfølger den.

Det opnåede fit er dog stadig godt nok til, at det er en plausibelt forklaring på datasættets udvikling. Derudover er situationen også således, at selvom et bedre fit er opnået i differentialligningerne uden tidsforskydning, er det en ringere forklaring af de bagvedliggende mekanismer som modellen bygges på, såfremt at KK's forklaring af disse er korrekt.

Derudover skal der også tages stilling til hvorvidt *bedste-fit*-værdien kan sammenlignes direkte mellem differentialligninger og differensligninger, da de to ligningsformer er af så forskellig natur. Da differensligningen foregår over diskrete tidstrin, har den mindre fleksibilitet når det kommer til at fitte parameterverdierne. Altså er der flere mulige minimale korrigeringer tilgængeligt i parameterestimeringen af en funktion over kontinuert tid, hvilket medfører

at et bedre *bedste fit* vil være muligt at opnå til en sådan funktion, i langt de fleste situationer. Dette er af samme ræsonnement, som i situationen hvor en parameter med minimal indflydelse på funktionens udvikling, stadig vil tillade et bedre fit når den varieres, sammenlignet med når den ses som en konstant. Denne sammenhæng mellem antallet af parametre som fittes og *bedste fit*, er noget der er taget højde for i beregningen, i modsætning til forskellen mellem kontinuert tid og diskret tid.

Hertil skal tilføjes, at forskellene i *bedste fit* for differens- og differentiallyigningerne er tilstrækkeligt store til, at den ikke nødvendigvis kun kan tilskrives denne fordel i *bedste fit* for kontinuerte modeller, men sandsynligvis også er et resultat af, at modellerne bedre beskriver datasættets udvikling. Dog er forskellen muligvis ikke helt så stor som *bedste-fit*-værdierne umiddelbart hentyder til. For entydigt at kunne vurdere hvilken udviklingsform der beskriver fænomenet bedst, ville et større datasæt som dækker en længere tidsperiode være nødvendig, da det er efter datasættets slutning den største forskel i udviklingerne forekommer.

Med hensyn til parameterværdierne ses det, at de to differentiallyigninger er langt tættere på hinanden i forhold til parametrene størrelse, end de er på differensligningen med tidsforsinkelse. Dette var en forventet konsekvens af diskretisering af modellen, og kan derved ikke kun tilskrives tidsforsinkelsen.

Som fremgår på tabel 3.6, er værdien af u ændret i modsat retning ved differensligningen, end ved den korrigerede parameterestimering af differentiallyigningen. Da der nu refereres tilbage i tiden gennem f -leddet, forekommer stigningen af e over en længere periode end i differentiallyigningen. Derved er det nødvendigt at have en mindre værdi af u , for at kunne opnå et fald i cancercellerne omkring samme tidspunkt, da det tager længere tid for dette at træde i kraft. Denne ændring kan tilskrives næsten udelukkende at være et resultat af tidsforskydningen og ikke blot diskretiseringen af modellen.

Derved kan det med sikkerhed konkluderes at den største ændring i parametrene størrelse, som kan tilskrives denne tidsforskydning, er en mindre værdi af u . Den største ændring i modellens opførsel, er den bølgede sammenhæng mellem cancerceller og T-celler. Såfremt en af disse situationer blev påvist at være tilfældet, ville det være oplagt at introducere en tidsforsinkelse i modellen, da det nu er påvist, at modellen derved udtrykker denne egenskab.

Det skal hertil tilføjes, at tidsforskydningen ikke nødsager en diskretisering af modellen, men blot er vist gennem et differensligningssystem for at

gøre implementeringen af tidsforsinkelsen mere simpel. Hvilken ligningsstype der benyttes til beskrivelse af systemet, er en beslutning som skal tages separat fra hvorvidt en tidsforsinkelse bør tages i brug.

I denne sektion har vi observeret effekten af hvad der sker med modellen ved at få visse led til at fungere på alternative måder. Vi undersøgte hvad der skete hvis vi udskiftede den betingede erklæring $\frac{f-c}{g+c}$, som fremgår i ligning (3.2), med et logistisk led. Dette gjorde at differentiallyigningen ikke længere er autonom, hvilket indebærer at modellen nu er tidsafhængig og gør den svære at analysere.

Vi undersøgte også hvordan parameteren p_e udviklede sig indenfor et spektrum for at simulere grænseværdier for en opsættelse af p_e som en stokastisk parameter. Det viste sig her, at den undersøgte variation i p_e kunne forsinke perioden for at cancercellepopulation nåede et ækvilibrium, men at selve ækvilibrumet ikke ændredes. Vi lavede også en analyse ved kvasistabil tilstand. Modellen der kom ud af dette var en simplere version af KK's model, men var er ikke nær lige så god som deres. Vi vurderede derfor, at forsimplingen i langt de fleste henseende vil være uegnet til modellen. Tilsidst opstillede vi en model baseret på differensligninger. Differensligningens bedste fit var dårligere end KK's bedste fit. Qua forskellene ved de to typer af modeller, var det usikkert om hvorvidt de kunne sammenlignes meningsfuldt ud fra bedste fit

3.4 Delkonklusion

Sensitivitetsanalysen af KK's model belyste at g , b og p_e er robuste parametre. g viste sig dog sensitiv senere i en parameterestimering, men dette sås kun ved slutningen af datasættet, hvilket betyder at det er udenfor det domæne som denne rapport er interesseret i. Derfor holdes g som konstant i parameterestimeringen. p_e blev ligeledes ikke fastholdt under estimeringen, da den har et lille spænd over antagelige værdier, hvilket indebærer en høj modstand mod ændringer. b viste at have følgende betydning. På trods af at være robust, har dette led en stor overordnet effekt på hele modellen, da den bestemmer bærekapaciteten af cancerceller.

Fittet som blev opnået af parameterestimeringen ændrede modellens output meget lidt. Det betyder at plottet lavet med de nye parametre og KK's plot ligger meget tæt. Dette underbygger at KK's model er opbygget med troværdige parametre. Modellens 'betinget erklæring' blev udskiftet med en logistisk funktion, hvor u kan varieres mellem 25 til 31, afhængigt af hvilken estimering der anvendes. Derfor blev det logistiske led udformet således at leddets vækstrate er størst i området [25:31]. At udskifte den betingede erklæring med et logistisk led ændrede kun minimalt i modellens udvikling, hvilket giver grundlag for at modellen, med dette led, var bedre egnet end KK's. Indførelsen af p_e som et stokastisk led belyste ikke noget nyt, da punkterne i datasættet var et gennemsnit af flere punkter. Det stokastiske led ville være aktuelt såfremt og bidrage til en mere nøjagtig model, hvis der kunne analyseres på det data som gennemsnits punkterne er baseret på.

Der blev forsøgt at foretage en simplificering af modellen ved at samle de to ligninger til én ved at sætte ligningen for T-cellers vækstrate i en kvasistabil tilstand. Udfaldet af dette var en model der bevægede sig på samme måde, men modellens toppunkt var mindre detaljeret. Denne nye model blev fundet uegnet til at beskrive data. Selvom modellen var nemmere at bruge end KK's var dens output dårligere end forventet. En anden metode der kan bruges til at gøre modellen simplere er ved at ændre den fra en differentiellmodel til en differensmodel. Differensmodellens største forskel var at den havde et bølgende ekvilibrium. Selvom parameterværdierne ligger tæt på hindanden, blev differensmodellens *bedste fit* dårligere end KK's *bedste fit*. Det opnåede *bedste fit* er dog ikke så dårligt, at modellen kan afvises alene af denne grund.

4. Diskussion

Vi vil i dette kapitel diskutere hvordan valg af modelleringsværktøjer kan påvirke modelleringsprocessen, samt hvordan både selve refleksionen over disse valg og formidlingen deraf påvirker processen. Konsekvenserne af de enkelte valg er i udgangspunktet blevet fortolket i direkte forlængelse af den enkelte analyse. Følgende diskussion vil derfor primært fokusere på overordnede overvejelser omkring modelleringsprocessen, fremfor fortolkninger af hvordan enkelte valg og fravalg har påvirket den undersøgte case.

4.1 Case-modellen

All models are wrong, but some are useful.
(George E. P. Box)

I denne sektion vil vi belyse valgene af parameterverdier, samt parameter- og modeltype i KK. For parameterverdierne vil det betragtes hvor stor refleksionsgraden har været i forbindelse med valgene. For valgene af parameter- og modeltype, vil vi antage et abstrakt perspektiv på formålet med modellen. Valgene foretaget undervejs i analysen, samt de opstillede alternativer til KK og KP's model, vil blive diskuteret. Dette gøres med intentionen at uddybe, om valgene kan vægtes som gode forsimplinger af modellen, og til at give et nuanceret indblik i stillingstagen af dilemmaer som forekommer under en modelleringsproces. Dog vil den indledes med et afsnit, hvor der reflekteres over nogle af KK og KP's beslutninger, som endnu ikke er belyst, samt over konsekvensen af de alternativer til KK's beslutninger, som er undersøgt i denne rapport.

4.1.1 Valg af parameterverdier

Som nævnt i afsnit 3.1.2, så er der stor forskel på de angivne parametre i KK og KP i forhold til antallet af betydende cifre. Antal betydende cifre kan ofte tolkes som en indikator på sikkerheden bag præcisionen af parameterverdierne. Selv i KP, som trods alt blot benytter fire betydende cifre, anerkendes dette,

”Parameters are estimated to four decimal places because, as we show below, the initial condition for the experimental data given in Fig. 1c lies close to the separatrix (see Fig. 3) and very small changes in the parameters give noticeable differences in the predicted tumor growth curve.”

(Kuznetsov, Makalkin et al., 1994)

Fra dette citat kan der udledes to pointer. Først og fremmest, at KP betragter fire betydende cifre for systemet, som værende et højt nok antal, til at det er relevant at inkludere en retfærdiggørelse i artiklen af valget om at beholde dem i modellen. Yderligere ses det, at deres retfærdiggørelse for valget binder i, at små afvigelser i parametrene medfører store afvigelser i kurverne.

Modellens følsomhed overfor parameterændringer uddybes i artiklen ved hjælp af ækvilibriums- og bifurcation analyser, hvor det illustreres, at ændringer i parameterverdier kan resultere i at systemet søger mod andre ækvilibriumpunkter. Dette argument er fænomenologisk funderet i den forstand, at der foreligger en grundig matematisk analyse, som sikrer at modellen skal være i stand til, funktionelt, at imitere det modellerede fænomen. Der er dog ikke mekanistisk belæg for denne tilgang. Langt de fleste parameterverdier er fundet alene via parameterestimering, altså rent fænomenologisk, og selv de bedst mekanistisk funderede værdier er fundet på baggrund af estimater med langt mindre nøjagtighed, end antallet af inkluderede betydende cifre insinuerer. Samtidig er de parametre, som KP argumenterer for har mest troværdige værdier, præcis de samme parametre, som blot har to betydende cifre; e_0 , k og b .

Den mest meningsfulde tilgang til udvælgelsen af antal betydende cifre ville umiddelbart være, at lade valget være dikteret af de mest troværdige parameterstørrelser og dermed begrænse parametrene til to betydende cifre. At sætte denne grænse ved tre, eller fire cifre som KP har gjort, er dog ikke direkte kritisabelt. Når grænsen, som i KK, sættes ved ti cifre, så fremstår afgrænsningen dog uovervejende. Der er absolut ingen biologisk baggrund for at antage at man kan estimere værdierne så præcist. Omvendt kan beslutning heller ikke tilskrives en rent fænomenologisk tilgang. Den parameterestimering, som opførtes i denne rapport blev udført med tre betydende cifre og opnåede et mindst lige så godt fit til data.

Ydermere er argumentet omkring at benytte mange betydende cifre på

grund af stor parameterfølsomhed en diskussion værd i sig selv. Ækvilibriums- og bifurcation analyser kan, for visse systemer, sagtens anvendes, som argument for netop at indsnævre spektret af realistiske parameterverdier. Det kræver dog at der betragtes et system, hvor der er meget stor tiltro til både selve den mekanistiske forståelse, dens systematisering og dens matematisering. Yderligere er det nødvendigt for systemet at de mekanismer, som parametrene beskriver kan antages at være konstante eller i hvert fald meget stabile. KP anerkender, at dette ikke er tilfældet for det modellerede system

It is clear that small fluctuations of parameter values take place in vivo. Neither the effector cell or tumor cell populations are homogeneous. Differing subpopulations will have different parameter values characterizing their behavior.

(Kuznetsov, Makalkin et al., 1994)

Hvis de bagvedliggende mekanismers opførsel er præget af stor tilfældighed eller er meget påvirkelige overfor udefrakommende faktorerer, så tilsvarende det matematisk set en situation, hvor parameterverdierne flukturerer over tid. Hvis dynamiske analyser af modellen viser stor parameterfølsomhed i en sådan situation, så vil den resulterende model være ustabil. I en situation, hvor data ikke viser store afvigelser i systemets opførsel, så vil den oplagte konklusion være, at modellen dermed er dårligt egnet til at beskrive systemet. I KP nævnes det dog faktisk, at datasæt over cancercellers vækst netop er præget af store afvigelser.

From our numerical experiments, shown in Fig. 6, we see that small changes of initial conditions in the sneaking through region lead to very large differences in the time period needed for the appearance of large tumors. These results correspond to observations on the emergence from the dormant state for BeL1 lymphoma (Krolick et al., 1979; Strober et al., 1979; Siu et al., 1986; Uhr et al., 1991) as well as for other experimental models of the tumor dormant state (Weinhold et al., 1979a,b; Uyttenhove et al., 1983; Hiernaux et al., 1986).

(Kuznetsov, Makalkin et al., 1994)

Dermed kan parameterfølsomheden i denne situation faktisk betragtes, som et argument til fordel for modellens validitet, da små korrigeringer af modellens parametre gør modellen i stand til potentielt at beskrive forskellige observerede opførsler af systemet. Det gør det også meningsfuldt at inkludere mange betydende cifre i selve parameterestimeringen for at sikre at modellen

er i stand til beskrive det specifikke datasæt. Dog er det ikke meningsfuldt at ekstrapolere argumentet, for at beholde betydende cifre under parameterestimeringen, til at beholde cifrene i beskrivelsen af det egentlige biologiske system. Mens dette ikke nævnes eksplicit i KP, så anerkendes det generelt at parameterverdierne skal betragtes som foreløbige estimater.

Given the limited data of Fig. 1, the numerical estimates of the model parameters for the response to BC£1 cells should only be regarded as preliminary.

(Kuznetsov, Makalkin et al., 1994)

Med denne anerkendelse i mente, samt den dybdegående tolkning af parameterverdiernes rolle for systemet generelt i KP, så virker det endnu mere uforståeligt at der i KK er anvendt ti betydende cifre; specielt taget i betragtning, at artiklen på intet tidspunkt italesætter denne beslutning. Dette kommer ikke blot til udtryk ved antallet af cifre for parametrene, men også ved deres enheder. KP præsenterer alle parameterverdier udførligt med enheder, hvorefter der foretages en dimensionsanalyse for at afdimensionalisere hele systemet, som derefter opsættes med nye parameterverdier; uden enheder. I KK derimod, angives der ingen enheder for parametrene, til trods for at systemet ikke præsenteres som værende dimensionsløst. Tværtimod antydes der i brødteksten, ved forklaringen af de forskellige parametres funktion, i flere tilfælde enheder relateret til antal celler eller tid.

the parameter u is the delay time of the immune response

(Kuznetsov, Makalkin et al., 1994)

Dette citat er symptomatisk for KK's beskrivelser af parametrene. Der antydes tydeligt en tidlig enhed for parameteren, men det klargøres ikke hvilken enhed der er tale om. For u er det ganske vist relativt oplagt at betragte dage, som et kvalificeret gæt på enheden, men for andre parametre er det væsentligt mindre intuitivt, hvilken enhed der passer til parameteren. Ved enkelte parametre er beskrivelsen en smule klarere,

the parameter b is the reciprocal of the maximal tumor cell population size; i.e., $1/b = (\text{the maximum number of tumor cells permitted to arise})/10^6$.

(Kuznetsov, Makalkin et al., 1994)

Her er det tydeligt antydnet at parameteren b har en enhed relateret til celler. Samtidig tydeliggør beskrivelsen også at b påvirkes af beslutninger om

at modellere populationerne, som antal millioner af celler, fremfor blot antal af celler. Det er i sig selv positivt, men problematiseres af, at det ikke nævnes for de øvre parametre hvorvidt de ligeledes er påvirket af denne faktor. Specielt fordi, at denne faktor rent faktisk påvirker nogle, men ikke alle, af de øvrige parametre. For eksempel indgår g i modellen som en addition til c . Addition kræver enhedsoverensstemmelse, hvormed g må have $\frac{\text{celler}}{10^6}$, som enhed. d derimod indgår blot som en tidlig enhed og er ikke påvirket af faktoren. Gennemgående er KK præget af mindre fortolkning og refleksion over de fundne parameterværdier, samt metoden bag deres estimering.

4.1.2 Valg af parameter- og modeltype

KK forsøger at beskrive et kompliceret biologisk fænomen i en matematisk model. Der er frembragt flere eksempler på hvordan KK er påvirket af en push-pull effekt, som konsekvens af at skulle vælge en tilgang der forøger modellens egnethed i ét aspekt, og mindsker dens potentiale i et andet, ud fra principperne om præcision, generel anvendelse og realisme. Kompromisser foretaget med tanke for dette, er udgangspunktet af opstillingen af alternative valg til deres fremgangsmåde. De foregående processer vil nu anskues i et bredere perspektiv centreret i overordnede formål for modelleringen.

Kvasistabil analyse

Det kvasistabile led, som blev indsat i ligning (3.1), forringede modellens præcision, vurderet ud fra dens *bedste fit*. Dette var en forventelig konsekvens som følge af simplificeringen.

Følgende tager derfor ikke udgangspunkt i resultatet af den kvasistabile analyse. Det forholder sig teoretisk til anvendelsen en model i kvasi stabil tilstand til at beskrive samme fænomen som i casen.

Animal models of tumor dormancy are essential for understanding fundamental aspects of cancer biology and for exploring therapeutic strategies that may reduce the risk of tumor relapse
(Kuznetsov, Makalkin et al., 1994)

I KK's model er formålet at beskrive cancer- og T-cellernes interaktion således at man i et bredere virke kan forbedre udviklingen af immunterapi. Systemet er modelleret med en mekanistisk tilgang, medvirkende til en mere realistisk beskrivelse af systemets dynamiker, og i en tolkning af deres

formål, at være præcis i dens forudsigelser.

The dynamics of the anti-tumor immune response in vivo are complicated and are not well understood

(Kuznetsov & Knott, 2001)

Da de ikke kender systemets opførelse på fyldestgørende vis, er det nødvendigt at stille sig kritisk overfor om hvorvidt deres model giver en realistisk indsigt i systemet. De har været nødsaget til at foretage valg baseret på antagelser om genstandsfænomenet, ud fra kliniske og eksperimentelle forsøg, der ikke ledte til definitive resultater.

Selve deres mekanistiske tilgang baseres på en, bemærk *plausibel*, antagelse om at *cancer dormancy* skyldes mindst to faktorer; enten iboende vækst hæmmende aktiviteter, eller et opnået ligevægtsstadie, forsat af interaktionen mellem immunsystemet og cancerceller (Kuznetsov & Knott, 2001). At fænomenet er modelleret på et tvivlsomt kendskab til mekanistiske årsagsammenhænge, forstærker mistanken om en forringelse af præcisions evne (Haefner, 2005). Deres model beskriver tre datasæt som har en vis uafhængighed. Dog kan denne formodning udfordres da miljøet er ens og at eksperimentet er udført med samme type mus som forsøgsobjekter, og under ens betingelser. Dette er selvsagt et vigtigt kriterie for reproducerbarheden, men svækker dog også en principiel stræben efter universalitet, og efterprøvelse af model med uafhængig data.

Vurderingerne af svagheder ved deres model, styrker en kvasistabil analyse på nogle områder. Her accepteres det at systemet bliver grovere sat op med en forøgelse i dens generelle anvendelsespotentialer. Det er mere sandsynligt at en model under indflydelse af en kvasistabil funktion, er mere bred anvendelig, og kunne genskabe samme resultater, med andre datasæt, dog med en grovere præcision. Deres model er derved teknisk set bedre til at beskrive data, dog underlagt antagelser om systemet foretaget i systematiseringsprocessen.

Betinget erklæring

I analysen af den betingede erklæring valgtes det at korrigerer deres f-led med en logistisk vækst model alternativt til deres valg.

The term $(\frac{f(t) \cdot c(t)}{g+c}) \cdot e(t)$ is the additional rate of both local proliferation and arrival of killer cells due to the chemokine and cytokine

stimulation signals induced by the tumor-cell population of size $c(t)$. The expression $\frac{f(t) \cdot c(t)}{(g+c)}$ is a logistic growth rate expression in which $f(t) = v$ if $t = 0$ then v else if $t < u$ then 0 else v

(Kuznetsov & Knott, 2001)

Argumentet for betingelsen begrundes i en antagelse om at det er biologisk meningsfuldt at inkorporere i modellen. Dette vægter KK yderligere med at fittet passede godt med data (Kuznetsov & Knott, 2001). Antagelsen understøttes ikke med teoretisk- eller biologisk indsigt, for så vidt det er undersøgt, og i deres artikel. Med et skøn om at det var mere biologisk plausibelt at indføre, udskiftedes deres led med et logistisk udtryk i modelkorrigeringssektionen. Tiltaget vurderes mere realistisk funderet, da modellen derved blev kontinuert, og mere fleksibelt, da de fleste biologiske fænomener er dynamiske- og normalfordelt, fremfor en udvikling med en konstant faktor (Haefner, 2005). Disse to fordele vægtes som vigtige for en mere realistisk beskrivende model, i forhold til den biologiske natur. Ved at indføre et logistisk led er det derfor argumenterbart at modellen er tættere knyttet til KK's egentlige formål.

En anden positiv konsekvens er at formidlingen af model bliver mere anskueliggjort, da den betingede erklæring fjernedes. Som modvægt til fordelene, konverterede dette valg, modellen til et ikke autonomt system, hvilket komplicerer processen med at foretage en dynamisk analyse.

Differensligningssystem

The assumption that, after the introduction of tumor cells (at time 0), no enhanced immune response occurred during the first 28 days after the mice were challenged with such injections was key to obtaining the excellent fits exhibited in Figure 1. (Kuznetsov & Knott, 2001)

KK formoder at tidsforsinkelsen kan forsvares med at fittet er godt til at beskrive data. Dette er selvsagt et matematisk stringent argument som dog er med til at validere deres model, og at modellen repræsenterer virkeligheden. Dette er tidligere ræsonneret i rapporten, med indskydelse til at tidsforsinkelsen rettidigt afspejler den faktiske immunionselle respons på cancercellernes vækst. I KK's identifikation af mekanismen under systematiseringsprocessen, fremstilles dette også således at det er matematisk nøjagtigt at imitere udviklingen således. Dette ledte til en opstilling af en differensligning hvis

modus operandi er diskretiseringen af tid.

The parameter v is the maximal rate of logistic growth of killer-cell population due to tumor growth, and the parameter u is the delay time of the immune response before “new” killer cells can be applied to attack the tumor; i.e., u is the time for precursor cells to mature into killer cells.

(Kuznetsov & Knott, 2001)

At vælge den ene model overfor den anden, er oftest bestemt af modellørens forudtagende vaner i modelleringspraksis. Der er derved ingen afklarende konklusion af hvad der er bedst egnet, med begge typer af model repræsenterende hver deres styrker.

4.2 Modelleringsværktøjernes indflydelse

”Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.”
(Bertrand Russell)

I denne sektion vil vi beskæftige os med, at drage generelle konklusioner omkring modellering, på baggrund af de resultater som blev præsenteret i analysen. For at opnå dette, vil vi for en række af analysepunkterne diskutere hvilke af de observerede konsekvenserne, som kunne siges at være unikke for lige netop implementeringen i denne model.

4.2.1 Analyseværktøjer

I rapporten ses det, at for modellen opstillet af KK, blev antallet af sensitive parametre i sidste ende reduceret fra 8 parametre til 6 (parametrene b og p_e kunne ses som robuste). Der er ingen tvivl om, at en sådan reduktion af parametre både forsimples modellen i stor stil, men også reducerer parameterrummet som skal undersøges i parameterestimeringen meget. Derved reduceres sandsynligheden for overfitting, hvilket er den største mulige fejl for modellen, såfremt mekanikkerne i systemet er beskrevet korrekt i den. At finde et lokalt minimum i parameterestimering, vil have samme effekt på modellens evne til at beskrive andre datasæt, men i et mindre omfang. Dette skyldes at det lokale minimum vil være tæt på et lige så godt fit som det globale,

såfremt undersøgelsen af parameterrummet har været nogenlunde dækkende.

Disse robuste parametre er noget som sensitivitetsanalysen tillod at finde, gennem en undersøgelse af den resulterende ændring i modellens output, når en parameter blev ændret. Problematikken med parameteren g 's sensitivitet, vil tages op til sidst. Der er ingen tvivl om, at en undersøgelse af sensitive parametre, som tillader en forsimpning af modellen, er ekstremt gavnligt. Processen af at finde disse robuste parametre, kræver ikke nødvendigvis en undersøgelse, som den der blev gjort brug af i denne rapport, men reducerer kompleksiteten af dette gevaldigt.

Er der med en meget simpelt model er et analyseværktøj ikke nødvendigt for at undersøge robustheden. Her vil en simpel algebraisk undersøgelse af parametrenes størrelsesordener, og hvordan de indgår i led, være nok. Er der derimod tale om en model med mange variable, så er det ikke muligt at bestemme alle parametrenes effekt på modellen uden mere sofistikerede analyseværktøjer.

I denne rapport blev der kun udført en sensitivitetsanalyse som undersøgt parametrenes isolerede sensitivitet. Mere dækkende programmer, som undersøger alle de mulige kombinatoriske ændringer i parametrene er lavet. Dette ville i casens tilfælde have givet et mere nøjagtigt udbytte, da det ikke senere havde været nødvendig at sammenligne parameterestimerings resulterende parameterværdier. Disse programmer er dog ikke altid at foretrække over dem som blev brugt i denne rapport, da der undersøges langt flere ting, og analysen derved bliver mere tidskrævende at udføre.

Som en afsluttende note skal det nævnes, at såfremt en sensitivitetsanalyse konkluderer at alle parametre er sensitive, kan det siges at analysen ikke havde nogen relevans for modellen. Dog ville denne viden ikke være muligt at opnå uden at foretage den faktiske sensitivitetsanalyse, efter hvilket punkt det bliver svært at definere hvor stor en relevans den har.

4.2.2 Parameterestimering

Som vist i afsnit 3.2.2 med parameteren u , er det muligt for en søgealgoritme at finde et lokalt minimum under optimering af parameterstørrelser, i stedet for det globale minimum. Som nævnt i forrige afsnit, kan sandsynligheden for dette reduceres, ved mindskning af antallet af parametre i modellen. En anden faktor i dette, er valg af algoritme til parameterestimeringen. Dette blev eksemplificeret i denne rapport, gennem den store forskel i kvaliteten af fit

til datasættene. Dette var et ekstremt tilfælde, da en moderne algoritme blev sammenlignet, med en fra 2001, men udpensler dog stadig den store mulige fejlkilde det kan være. Selvom KK's algoritme har haft langt mindre computerkraft til rådighed, har den været effektiv nok til at være danne grundlag for en model, som fortsat bruges i forskning, og kunne derved tænkes at være sammenlignelig med en mindre god algoritme, med mere computerkraft bag sig.

At finde et lokalt minimum i parameterestimeringen, er stadig en relevant risiko i dag. Jo større et parameterrum modellen har, jo vigtigere er algoritmen som skal søge det igennem. Altså vil vigtigheden af en god algoritme, for modellens evne til at beskrive systemet, være en funktion af parameterummets størrelse. Dette betyder at med mindre der er med en model af få parametre at gøre, har parameterestimeringens kvalitet, og derved algoritmen som bestemmer denne kvalitet, en meget stor indflydelse på modellens evne til at beskrive systemet.

4.2.3 Teoretisk viden

Når der foretages en modellering af et system, hvorom man kender noget teoretisk viden, kan dette ofte modelleres på forskellige måder. At vælge den måde som afspejler mekanismen bedst kan ofte være en udfordring, men er af meget stor betydning for modellens brugbarhed, såfremt en mekanistisk model ønskes opstillet. Dette afsnit funderes i en antagelse om, at det er en mekanistisk model som modelløren ønsker at opstille. Der tages udgangspunkt i implementeringen af tidsforsinkelse i modellen, men postuleres ikke at dette er den korrekte forståelse af systemet som KK modellerer. Dette tages blot op som eksempel, for at eksemplificere konsekvenserne af forkerte basisantagelser i en mekanistisk model, som skal ændres efter opstilling af modellen.

I alle mekanistiske modeller, er det vigtigt at mest muligt teoretisk viden om systemet kommer til udtryk i den opstillede model. I afsnittet med implementering af tidsforsinkelse i en differensligningsversion af KK's model, sås det at dette ændrede modellens karakter gevaldigt. Uden tidsforskydelsen kan modellen ikke imitere en repetitiv bølgende form, som ligningerne ser ud nu. På samme måde kan man med tidsforskydelsen ikke få et model output, som falder til ro ved et equilibrium, som de ser ud nu. Såfremt at KK's forklaring af den teoretiske viden er korrekt, og at forståelsen af at dette må nødsage en tidsforskydelse er korrekt, er det dårligt nyt for model-

len, medmindre den fortsatte udvikling af cancercellerne også følger denne bølgende form. Er dette ikke tilfældet, ville en anden struktur af differential-ligningerne være nødvendig. Differentialligningerne ville skulle ændres, således at modellen kunne beskrive den observerede udvikling, samtidigt med at den reflekterede den teoretiske viden om tidsforsinkelsen. Såfremt data viste at cancercellerne nåede en fast mængde, som modellen uden tidsforsinkelse viser, ville modellen med tidsforsinkelse aldrig kunne beskrive virkeligheden, som den ser ud nu.

Dette viser hvor vigtigt det kan være i en mekanistisk model, at alle led reflekterer den faktiske mekanisme, eller en forsimpning af dette. Selvom den generelle form af modellen ville forblive den samme, med implementering af tidsforsinkelse, ændres modellen så meget at det ikke er muligt for det, at opnå samme udvikling. Opstilling af en model baseret på forkerte præmisser kan have katastrofale konsekvenser for modellens evne til at beskrive systemet, når den korrekte forståelse indsættes. Derved kan det være nødvendigt at starte forfra med modellering når en sådan situation opstår.

Tilfældighed i modeller

Som tidligere nævnt i afsnit 3.3.2, om introduktion af en stokastisk parameter, kan en variation i en parameters størrelse introduceres i den matematiske model. Dette er relevant i modeller, hvor parameters størrelser ikke kan gøres universelle, men vil variere under forskellige forhold. Medmindre alle disse forhold forstås, kan der ikke opstilles en mekanistisk model af et sådant system. Dette kan enten håndteres ved at arbejde ud fra gennemsnit af de observerede fænomener, eller ved at introducere en tilfældighed i parameterstørrelsen. Der kan derudover også opstilles argumenter for, at den forsimpning som forekommer i sammenhæng med at beskrive en kompleks biologisk process som en tilfældighed, er at foretrække frem for en model som inkorporerer alle parametre. Dette er funderet i en forståelse af at en overfittet model aldrig vil have nogen nytte, og at forsimplinger derved er nødsaget. Derved skal indførelsen af den stokastiske parameter holdes op imod de andre forsimplinger som kan foretages, således at en system opnås som ikke er overfittet, og som stadig afspejler størstedelen af mekanismerne på en virkelighedstro måde.

Der er to situationer hvori en stokastisk parameter, kan bruges til at beskrive fænomenet. Enten introduceres den på baggrund af, at den teoretiske viden (eller den tilgængelige information) ikke er tilstrækkelig omkring denne parameter, eller også introduceres den på baggrund af en teoretisk viden om

tilfældighed i systemet. I den første af disse situationer, vil brug af en stokastisk parameter tillade modellen, at beskrive den observerede opførsel bedre. Dette vil dog ikke være gennem de samme mekanismer som i systemet, og vil derved være en fænomenologisk tilgang til modelleringen. I biologiske systemer er der dog ofte så mange forhold som kommer i spil, at en mekanistisk model som inkluderer dem alle, vil være for kompliceret til at gøre nogen gavn. I sådanne tilfælde kunne argumenter opstilles for, at en tilfældighed som afspejler disse forskellige forhold fra individ til individ, vil være tættere på den bagvedliggende mekanisme, end en model som er baseret på gennemsnit.

Såfremt det søges at få opstillet en mekanistisk model af systemet, vil det være fordelagtigt at tage en stokastisk parameter i brug, såfremt der findes teoretisk belæg for dens nødvendighed i systemet. Såfremt ingen andre alternativer til at beskrive spredningen i fænomenet kan findes, kan en stokastisk parameter muligvis være den mest virkelighedsnære forklaring, som ikke resulterer i en gigantisk model.

4.3 Procesorienteret modellering

”In mathematics the art of proposing a question must be held of higher value than solving it.”
(Georg Cantor)

I denne sektion vil problemformuleringen anskues fra en ny vinkel. Det vil her forklares, at en models kvalitet ikke blot kan repræsenteres som dens evne til at beskrive den observerede virkelighed. Vi vil først diskutere vigtigheden ved formidlingen af det teoretisk belæg, samt andre overvejelser herom, der er essentielle for indsigt i modelleringprocessen. Derefter vil vi reflektere over præcis formidling som et redskab til at lave en god model. Med udgangspunkt i KK's formidling fremhæver vi de områder hvor deres kommunikation er uskarp eller mangelfuld og diskutere betydningen af dette.

En models kvalitet kan ikke alene afgøres ud fra hvor præcist den passer til den modellerede data. Det er afgørende for modellen, at samspillet mellem matematikken og den modellerede virkelighed er meningsfuld og gennemskuelig. I denne forbindelse er det ikke blot vigtigt at modellen er opsat med virkelighedstro parametre og matematiske operationer. Det er også essentielt at de overvejelser og konklusioner der er draget under denne proces bliver videreformidlet i beskrivelsen af modellen. Betydningen af denne formidling

er direkte relateret til den bredere modelleringsproces, hvor en model betragtes som et redskab der kan tilpasses og optimeres fremadrettet (Blomhøj & Højgaard, 2003); eller endda ekstrapoleres til andre scenarier, som for eksempel ses ved Lotka-Volterra modeller. (Marino et al., 2013).

4.3.1 Formidling som middel til optimering

En velstruktureret redegørelse for den proces der ligger bag den præsenterede model vil både kunne forenkle videreudvikling af modellen, samt tydeliggøre dens potentielle virkeområde. Videreudvikling vil især kunne drage nytte af sensitivitetsanalyser, samt overvejelser om i hvor grad de enkelte parametres effekt på systemet kan vurderes ud fra den tilgængelige data. Disse overvejelser vil kunne vejlede fremtidig optimering i forhold til, hvilke parameterværdier der er mest relevante at optimere på. I forlængelse af dette, vil de samtidig kunne agere som en guide for hvilke datasæt der bør indsamles for bedre at kunne fastsætte disse parametre præcist. Som eksempel kan den case, som er blevet betragtet i denne rapport, anvendes med udgangspunkt i parameteren g :

“However, increasing the parameter g from 0.16 to 5 did not change approaching the tumor dormancy steady state.”

(Kuznetsov & Knott, 2001)

Som KK selv nævner i artiklen, så forårsager selv en meget stor relativ ændring i g nærmest ingen ændring i systemet før dets ækvilibria. Som det kunne ses i denne rapportes parameterestimering, så har g en reel indflydelse på ved hvilke værdi modellen rammer disse ækvilibria. Siden g ikke påvirker modellen for den periode som sammenlignes med data, så vil g ikke kunne estimeres meningsfuldt. En formidling af denne pointe i forbindelse med modellen ville både klargøre g 's rolle for modellen og samtidig give anledning til at foretage målinger, som enten foregår over en længere periode eller som hurtigere vil nå et ækvilibrium. På denne måde ville der være en oplagt fremgangsmåde for at opnå et mere solidt estimat for g

I forhold til at klargøre optimeringspotentialer for modellen har det også betydning at det beskrives hvilken metode der er benyttet til at estimere parametrene. Dette kan gøres relativt simpelt ved at tilgængeliggøre den anvendte kode i forbindelse med udgivelse af modellen. Såfremt dette ikke gøres, kan der i stedet inkluderes en beskrivelse af processen. Fra artiklen af KK betragtes deres beskrivelse af processen.

”This fitting required various exploratory computations and careful search of the parameter space for suitable initial guesses that lead to physiologically-plausible values.”

(Kuznetsov & Knott, 2001)

Denne beskrivelse af processen rummer i virkeligheden de vigtigste pointer at formidle fra en parameterestimering, men formår ikke at relatere dem til deres egen proces. Det hævdes at parameterrummet blev grundigt undersøgt, men ikke med hvilken metode, og det hævdes at der blev fundet fysiologisk plausible værdier, men ikke ved hvilke grænser de blev undersøgt.

4.3.2 Formidling som middel til anvendelse

I forhold til at definere modellens virkeområde er det især vigtigt at klargøre hvordan selve parameterværdierne er fundet. For den enkelte parameter bør det klargøres hvilke mekanistiske overvejelser der ligger bag estimeringen. På denne måde vil selve modellens struktur langt lettere kunne tilpasses til andre situationer via klarhed omkring hvilke parametre der eventuelt skal justeres, samt et udgangspunkt for hvilken baggrund denne justering skal udføres fra. For parametre der betragtes som mekanistisk velfunderede, i forhold til det pågældende felt, vil det i denne forbindelse være gavnligt både at angive hvilken værdi parameteren kan forventes at have, samt hvilken afvigelse herfra der kan betragtes som realistisk. Hvis baggrunden for, at parameteren betragtes som mekanistisk ikke bygger på et forhenværende kildemateriale, men i stedet et fagrelevant ræsonnement, så bør dette ræsonnement fremgå af opgaven. For den pågældende case ses også et eksempel på dette,

”Note particularly that 28 days is an appropriate estimate of the maturation time for CD8+ cytotoxic T lymphocytes.”

(Kuznetsov & Knott, 2001)

Mens det selvfølgelig er positivt at denne betragtning er formidlet, så fremstår sætningen nærmest udelukkende som en validering af den fundne værdi, men er af begrænset brugbarhed i forhold til videreudvikling. Udsagnet er ikke underbygget af kildemateriale og det klargøres hverken på hvilken baggrund, at de 28 dage vurderes som et passende estimat eller hvilket spænd af værdier der ligeledes kunne betragtes som realistiske. Yderligere præciseres det ikke, hvorvidt der er taget højde for et sådant spænd under selve parameterestimering; det vil sige hvorvidt de har begrænset det undersøgte parameterrum

tæt ved de 28 dage.

For parametre, som umiddelbart er fænomenologiske i forhold til den modellerede videnskab, men som via deres matematiske funktion i modellen alligevel vurderes at have en begrænset rækkevidde af værdier, så bør disse overvejelser ligeledes fremgå. Casen fra KK viser et eksempel på, at dette også kan være fordelagtigt i forhold til, som modellør at bevare overblikket over parametrenes rolle i modellen:

”Now, if we modify our model to have the parameter p change from 0.9982002827 to 90 percent of that value after 150 days, then tumor regrowth is exhibited as shown in Figure 2. This is done by replacing p in our seven differential equations by the expression $h(t)$, and defining $h(t) = \text{if } t < 150 \text{ then } p \text{ else } 0.9p$. This result indicates that a modest decrease in immune response effectiveness, corresponding to a small increase in the proportion $1 - p$ of killer lymphocytes being inactivated by tumor cells, dramatically changes the outcome of the disease.”

(Kuznetsov & Knott, 2001)

Denne sætning refererer egentlig til en anden del af artiklen, hvor modellen tilpasses til at beskrive systemet under andre forhold. Sætningens relevans for vores rapport vedrører derimod deres betragtninger omkring parameteren p . Som tidligere nævnt i denne rapport, så er p en procentsats, som altså vil have en værdi i mellem 0 og 100%. En ændring fra $p = 99,8\%$ til $p = 0,9 \cdot p = 89,8\%$ kan derfor betragtes som en meget markant ændring af modellen. Til trods for dette beskrives ændringen i rapporten som *”a modest increase”*, som *”dramatically changes the outcome”*. Det tyder umiddelbart på, at forfatterne i forbindelse med denne observation ikke havde et fuldt overblik over funktionen af parameteren.

Alt i alt er hele formålet med at opstille en model at skabe et værktøj, som gør det lettere at analysere forskellige aspekter af omverdenen. Hvilke aspekter der undersøges, hvordan de undersøges og hvilke værktøjer der benyttes til at undersøge dem kan variere. Fælles for dem alle er dog, at et værktøj er mest effektivt, når der foreligger en klar forståelse af præcis hvordan værktøjet anvendes. I forhold til modellering er det gennem formidling at denne forståelse skabes.

4.4 Foregrebet iværksætter

”Mathematics is the art of giving the same name to different things.”
(Henri Poincare)

I denne sektion vil et nyt perspektiv på problemformuleringen undersøges. Indtil nu har rapporten beskæftiget sig med de observerbare konsekvenser for modellen, af det konkrete modelleringsværktøj som undersøges. I denne sektion vil omdrejningspunktet, være valg af modelleringsværktøj på baggrund af tidligere opstillede modeller. Det vil herunder diskuteres, hvordan en undersøgelse af de tilgængelige værktøjer kan undgå denne faldgrube.

Enhver modelleringsprocess vil være præget af det tidligere modelleringsarbejde, som ligger før denne model. Dette kan både være i tidligere modelleringsarbejde udført af de, som modellerer systemet, men også i andre modelleringer af lignende systemer, som kendes på forhånd. I hvor stor grad dette vil påvirke modellen, vil kunne bestemmes ud fra mængden af modelleringsarbejde. Såfremt at et bestemt modelleringsværktøj er blevet brugt i alt tidligere arbejde, vil den nye model om systemet højst sandsynligt inkludere dette værktøj, uden at skænke alternativerne en tanke. Sandsynligheden for at alternativerne er undersøgt eftertrykkeligt, vil stige med antallet af andre modelleringer af lignende systemer. Dette er dog ikke tilfældet, når det kommer til modelleringsarbejde af andre systemer, udført af de som opstiller modellen.

Dette skaber en problematik, da valg taget på denne baggrund sjældent vil mistænkes for at være årsagen til eventuelle problematikker for modellen. Det kan derved munde ud i, at ændringer af korrekt opstillede matematiske termer foretages, hvilket vil bringe modellen længere væk fra den mekanistiske del af skalaen, som modelløren søger hen imod. Derved bliver de valg af modelleringsværktøjer, som foretages på baggrund af foregrebet iværksætter, langt mere problematiske for modellens evne til at beskrive systemet, end ville være synligt gennem undersøgelser som de, foretaget i denne rapport.

For at undgå denne problematik, vil det være nødvendigt at undersøge alle de mulige modelleringsværktøjer som kan bruges, såfremt der ikke findes teoretisk belæg for at bruge et specifikt af dem.

Derved kunne det være interessant at kunne opstille en indeksering af de forskellige værktøjers egenskaber og konsekvenser, således at valg af disse

kan tages med en større upartiskhed. Derved ville man kunne undgå alle problematikker om fejlagtige modeller, som er resultater af forudgående modelleringer. Dette ville dog kræve en meget stor mængde arbejde, da en sådan indeksering ville nødvendiggøre, at en langt mere dybdegående analyse af hvert af værktøjerne foretages, end den som er foretaget i denne rapport.

Den potentielle fordel ved en sådan indeksering, ville dog ikke stoppe ved modarbejdning af foregrebet iværksætter, men vil for eksempel også have stor nytte inden for undervisning. Derudover ville en sådan indeksering have samme effekt for matematisk modellering, som en matematisk model i almindelighed har for det område som modelleres; en større indsigt i dynamikkerne vil opnås.

5. Konklusion

Modelleringsværktøjerne som er undersøgt i denne rapport, er alle vigtige for modellens evne til at beskrive systemet som det modellerer. Dog er deres relevans unikt for den konkrete model, som der tales om.

Sensitivitetsanalysen er altid vigtig, for at sikre det mindst mulige parameterum som undersøges, men da det er et værktøj som forenkler processen, bliver det først nødvendigt ved systemer som er komplicerede. Konsekvensen af ikke at have udført en sensitivitetsanalyse, er tofoldig i et kompliceret system. Først og fremmest vil der blive arbejdet med et unødvendigt stort parameterum, i parameterestimeringen, og derigennem sandsynligheden for at opnå en overfittet model. Derudover er det muligt at modellen ikke vil være entydig, da det ikke vil være undersøgt hvorvidt nogle af parameterene er fuldstændigt robuste, hvilket kan give store problematikker for parameterestimeringen.

At gøre brug af en sofistikeret søgealgoritme i parameterestimeringen er vigtigst i modeller af mange variable, og har derved kun konsekvenser for modeller af mange parametre. I modeller af mange parametre, vil konsekvensen af en uegnet algoritme betyde at parameterestimeringen med stor sandsynlighed vil resultere i et lokalt minimum.

En stokastisk parameter vil kunne forbedre modellens evne til at beskrive fænomenet, men vil gøre modellen mere fænomenologisk såfremt der ikke findes noget belæg for, at selve mekanismen involverer tilfældighed. Med mindre den stokastiske parameter er underbygget af en teoretisk viden, vil en stokastisk parameter forbedre modellens evne til at beskrive en variation i data, men vil forværre modellens brugbarhed til at forstå systemet, da modellen vil blive mere fænomenologisk af denne tilgang.

At opstille sin model således at den teoretiske viden om systemet kommer til udtryk, eller en lettere forsimplet version af dette, er ikke blot fundamentalt vigtigt for en mekanistisk model, men er også på en mere funktionel plan. Beskrives en mekanisme en smule forkert, kan det resultere i en model som aldrig ville kunne opføre sig på en måde der ligner virkeligheden. Såfremt en

mekanistisk model skal opstilles, kan en enkelt parameter som er opstillet på en forkert måde, resultere i at modellen opnår en fuldstændigt anderledes udvikling. Dette blev set i afsnittet om differensligninger, da en tidsforsinkelse blev introduceret, hvilket resulterede i en bølgende udvikling i stedet for at finde hen til et ækvilibrium.

Ydermere er det fordelagtigt at videreformidle den proces der er gennemgået under modellens opsætning, da denne proces i sig selv rummer meget information. En strukturel redegørelse af processen kan både agere som en validering af den opnåede model, samt et udgangspunkt for optimering af modellen. Samtidig kan overvejelser der gøres i forbindelse med opstillingen af en model ofte være relaterede til forståelsen af det egentlige system eller til sammenhængen mellem systemet og modellen. Da mekanistiske modeller har større troværdighed og brugbarhed end fænomenologiske modeller, er det fordelagtigt at beskrive processen. Konsekvensen af en utilstrækkeligt forklaret modelleringsproces, kan derved ikke ses i modellens evne til at beskrive et system. Konsekvensen vil være begrænset til modellens troværdighed inden for det videnskabelige fælleskab, og hvordan den modtages inden for dette.

6. Litteraturliste

- Blomhøj, M. & Højgaard, T. (2003). Developing mathematical modelling competence: conceptual clarification and educational planning. Teaching Mathematics and Its Applications. *Teaching Mathematics and Its Applications*.
- Charles W. Misner, J. A. W., Kip S. Thorne. (1973). *Gravitation* (1st). W. H. Freeman.
- Ciancio, A. & Quartarone, A. (2013). A Hybrid Model For Tumor-immune Competition. *The Scientific Bulletin, Series A*.
- DiStefano, J. (2013). *Dynamic Systems Biology Modelling and Simulation* (1st edition). Elsevier Inc.
- Haefner, J. W. (2005). *Modeling Biological Systems - Principles and Applications* (2nd edition). Springer.
- Hirsch, M. W., Smale, S. & Devaney, R. L. (2013). *Differential equations, dynamical systems, and an introduction to chaos* (3rd). elsevier.
- Küppers, G., Lenhard, J. & Shinn, T. (2006). *Computer Simulation: Practice, Epistemology, and Social Dynamics*. Springer.
- Keesman, K. J. (2011). *System Identification an introduction* (1st edition). Springer.
- Kuznetsov, V. A. & Knott, G. D. (2001). Modeling Tumor Regrowth and Immunotherapy. *Mathematical Biosciences*.
- Kuznetsov, V. A., Makalkin, I. A., Taylor, M. A. & Perelson, A. S. (1994). Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. *Bulletin of Mathematical Biology*.
- Marino, S., Baxter, N. T., Huffnagle, G. B., Petrosino, J. F. & Schloss, P. D. (2013). Mathematical modeling of primary succession of murine intestinal microbiota. *Proceedings of the National Academy of Sciences*.
- Nannen, V. (2003). The Paradox of Overfitting. Hentet fra http://volker.nannen.com/pdf/the_paradox_of_overfitting.pdf
- Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica* (1st). The President, Council og Fellows of the Royal Society of London for Improving Natural Knowledge.
- Noble, R. D. (1982). Mathematical modelling in the context of problem solving.

Siu, H. & Vitetta, E. S. (1986). Tumor dormancy. I. Regression of BCL1 tumor and induction of a dormant tumor state in mice chimeric at the major histocompatibility complex. *The Journal of Immunology*.

Bilag

Formelsamling

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2} \quad (2.1)$$

$$F = 9.82 \frac{N}{kg} \cdot m \quad (2.2)$$

$$P_{t+1} = f(P_t) \quad (2.3)$$

$$S = \frac{n}{n_0} \quad (2.4)$$

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot e(t) \cdot c(t) \quad (3.1)$$

$$e'(t) = r + \frac{f(t) \cdot c(t)}{g+c(t)} \cdot e(t) - d \cdot e(t) - a \cdot (1 - p) \cdot c(t) \cdot e(t) \quad (3.2)$$

$$k \cdot c(t) \cdot (1 - b \cdot c(t)) \quad (3.3)$$

$$a \cdot p \cdot e(t) \cdot c(t) \quad (3.4)$$

$$\frac{f \cdot c}{g+c} \cdot e \quad (3.5)$$

$$\frac{dT}{dt} = a \cdot T(1 - b \cdot T) - n \cdot E \cdot T \quad (3.6)$$

$$\frac{dE}{dt} = s + \frac{p \cdot E \cdot T}{g+T} - m \cdot E \cdot T - d \cdot E \quad (3.7)$$

$$\frac{dC}{dt} = k \cdot c \cdot (1 - b \cdot c) - a \cdot p \cdot E \cdot C \quad (3.8)$$

$$\frac{dE}{dt} = r + \frac{v \cdot E \cdot C}{g+C} - a \cdot (1 - p) \cdot E \cdot C - d \cdot E \quad (3.9)$$

$$\text{levetid} = \text{precursor levetid} + \text{T-celle levetid} = u + \frac{1}{d} \quad (3.10)$$

$$\sum_{n=1}^i \sqrt{\frac{\log(x_i(\text{data}) - \log(x_i(\text{fit})))}{\log(\bar{x}(\text{data})) - N}} \quad (3.11)$$

$$f(t) = \frac{L}{1+e^{-h \cdot (t-t_0)}} \quad (3.12)$$

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot e(t) \cdot c(t) \quad (3.13)$$

$$e'(t) = r + \frac{\frac{v}{1+e^{-1000 \cdot (t-u)}} \cdot c(t)}{g+c(t)} \cdot e(t) - d \cdot e(t) - a \cdot (1-p) \cdot c(t) \cdot e(t) \quad (3.14)$$

$$e(t) = \frac{-r}{-d-a \cdot (1-p) \cdot c(t)+w} \quad (3.15)$$

$$w = \frac{f \cdot c(t)}{g+c(t)} \quad (3.16)$$

$$c'(t) = k \cdot c(t) \cdot (1 - b \cdot c(t)) - a \cdot p \cdot \frac{-r}{-d-a \cdot (1-p) \cdot c(t)+w} \cdot c(t) \quad (3.17)$$

$$c_{t+1} = c_t + k \cdot c_t \cdot (1 - b \cdot c_t) - a \cdot p \cdot e_t \cdot c_t \quad (3.18)$$

$$e_{t+1} = e_t + r + \frac{f \cdot c_t}{g+c_t} \cdot e_t - d \cdot e_t - a \cdot (1-p) \cdot c_t \cdot e_t \quad (3.19)$$

$$c_{t+1} = c_t + k \cdot c_t \cdot (1 - b \cdot c_t) - a \cdot p \cdot e_t \cdot c_t \quad (3.20)$$

$$e_{t+1} = e_t + r + \frac{v \cdot c_{t-u}}{g+c_{t-u}} \cdot e_t - d \cdot e_t - a \cdot (1-p) \cdot c_t \cdot e_t \quad (3.21)$$

$$\frac{v \cdot c_{t-u}}{g+c_{t-u}} \cdot e_t \quad (3.22)$$

6.1 MATLAB

Den anvendte matlabkode for rapporten kan findes i en separat zipfil. Denne zipfil vil indeholde koden, som er opdelt efter det afsnit som det bliver brugt i. Disse undermapper er nummererede og bliver refereret når de bliver brugt i rapporten til at producere en figur eller lignende. Tilhørende hvert sæt af filer, er en specifik ”README” tekstfil, som beskriver hvordan programmerne bruges.