

Faculteit Bio-ingenieurswetenschappen

Academiejaar 2010-2011

PARAMETRISATIE VAN EEN RUIMTELIJK EXPLICIET MODEL CASIR

Lien Eyerman Promotor: Prof. dr. B. De Baets Tutor: ir. Jan M. Baetens

> Masterproef voorgedragen tot het behalen van de graad van MASTER IN DE BIO-INGENIEURSWETENSCHAPPEN: MILIEUTECHNOLOGIE

De auteur en de promotor geven de toelating deze scriptie voor consultatie beschikbaar te stellen en delen ervan te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron te vermelden bij het aanhalen van resultaten uit deze scriptie.

The author and supervisor give the permission to use this thesis for consultation and to copy parts of it for personal use. Every other use is subject to the copyright laws, more specifically the source must be extensively specified when using results from this thesis.

Gent, juni 2011

Promotor

Tutor

Auteur

Woord vooraf

"A pessimist sees the difficulty in every opportunity; an optimist sees the opportunity in every difficulty."

Sir Winston Churchill

Vijf jaar geleden leek een thesis en zeker het woord vooraf van die thesis nog een verre droom. Vijf zware jaren zijn voorbij, maar het waren ook jaren vol plezier en vriendschap. Een dubbel gevoel overvalt me dus; ik ben klaar voor een nieuwe uitdaging, maar ik zal het studentenleventje toch wel missen.

Als milieumeisje wil ik vooreerst graag kwijt dat ik heel bewust gekozen heb om mijn thesis rectoverso te laten afdrukken. Ik leerde ooit dat één blad papier een watervoetafdruk heeft van ongeveer 10 liter; mijn uitgespaarde blaadjes zullen de wereld niet veranderen, maar vele kleintjes maken een groot!

Er zijn een aantal mensen die ik graag wil bedanken. Ten eerste prof. dr. Bernard De Baets om mij de kans te geven om zowel mijn bachelorproject als mijn masterproef te schrijven over een onderwerp dat mij blijft boeien. Ook ir. Jan Baetens die mij 3 jaar geleden de wondere wereld van cellulaire automaten leerde kennen, krijgt een grote merci. Je begeleiding was heel belangrijk in het slagen van deze thesis. Ik wil mijn thesisgenootjes bedanken om mijn gezaag en geklaag te aanhoren en mij af en toe af te leiden. Bedankt Wouter, om mij aan het lachen te brengen wanneer er 'No memory available' op mijn scherm kwam en voor de gezellige momenten ook buiten ons thesiskot. Bedankt Denis, om je enthousiasme en je straffe verhalen en proficiat met je overwinning! Ook Melanka en Steffie, bedankt voor de babbels en hulp! Tinne, jou kan ik natuurlijk niet vergeten te danken voor de gezellige babbels, de lekkere gebakjes en een knuffel op het juiste moment. Maar ook alle anderen van de vakgroep, bedankt voor het oprechte meeleven.

Tijdens deze laatste vijf jaar heb ik ook heel wat mensen leren kennen die mijn leven net iets leuker hebben gemaakt. Ik ga niet iedereen opnoemen want dan vergeet ik waarschijnlijk een paar mensen, maar allemaal bedankt! Een speciale dank je wel gaat naar Elke en Jolijn; met z'n drieën hebben we heel wat onvergetelijke momenten beleefd de laatste drie jaar. Ik zal jullie missen als jullie in het verre Canada zitten! Natuurlijk wil ik ook graag mijn ouders bedanken, niet alleen voor de financiële maar zeker ook voor de mentale steun. Als laatste wil ik mijn twee broers, schoonzus en twee schatten van nichtjes bedanken, ik zie jullie graag!

Tot slot, in deze laatste weken heb ik geleerd wat belangrijk is in het leven en dat is familie. Luc, ik zal je rake opmerkingen, grappige commentaren en de nachtelijke discussies met veel bier heel hard missen. Vaarwel!

Inhoudsopgave

In	Inleiding				
1	Lite	ratuurs	tudie	3	
	1.1	Cellula	aire automaten	3	
		1.1.1	Definitie en historiek	3	
		1.1.2	Elementaire CA	6	
		1.1.3	Evenwicht en stabiliteit van CA	8	
		1.1.4	Perspectieven van CA-gebaseerde modellen	12	
		1.1.5	Koppeling met geografische informatiesystemen	12	
	1.2	Model	len voor ziekteverspreiding	14	
		1.2.1	Belang	14	
		1.2.2	Continue modellen	15	
		1.2.3	Discrete modellen	17	
		1.2.4	CA-gebaseerd SIR	19	
	1.3	Parame	etrisatie van cellulaire automaten	21	
		1.3.1	Gevoeligheidsanalyse	21	
		1.3.2	Optimalisatietechnieken	22	
•				~-	
2	Invl	oed van	de topologie op de stabiliteit van CA	27	
	2.1	Soft- e	n hardware	27	
	2.2	Maxim	ale Lyapunov exponent van totalistische onregelmatige CA	27	
	2.3	Invloed	d van de topologie	29	
	2.4	Topolo	gische bifurcaties	32	
	2.5	Conclu	ısie	36	

INHOUDSOPGAVE

3	CA-	gebaseerd SI-model voor de verspreiding van een hantavirus	37
	3.1	Modelopbouw	37
	3.2	Modelimplementatie	38
		3.2.1 Problemen en onvolledigheden	38
		3.2.2 Simulatieresultaten	41
	3.3	Maximale Lyapunov exponent	43
	3.4	Conclusie	45
4	Para	ametrisatie van een CASIR	47
	4.1	Tesselatie	47
	4.2	Vergelijking met andere ziekteverspreidingsmodellen	48
	4.3	Gevoeligheidsanalyse	50
		4.3.1 Initiële conditie, evaluatiefunctie en modelparameters	50
		4.3.2 Perturbatie factor	52
		4.3.3 Analyses	54
	4.4	Parametrisatie	58
		4.4.1 Gesimuleerde gegevens	58
		4.4.2 Parameterschatting	59
	4.5	Conclusie	60
5	Alge	emene conclusies en verder werk	61
	5.1	Algemene conclusies	61
	5.2	Verder werk	62
Bi	bliogi	rafie	63

Lijst van figuren

1.1	Een onregelmatige tesselatie in \mathbb{R}^2	5
1.2	Graaf-voorstelling van een onregelmatige tesselatie	6
1.3	Ordening van de cellen in een 1D CA	6
1.4	Ordening van de cellen in een 2D CA	7
1.5	Von Neumann en Moore omgeving	8
1.6	Restrictieve von Neumann ongeving met restrictie	9
1.7	Verschillende indelingsklassen van elementaire CA	10
1.8	GIS objecten	13
1.9	Integratie van GIS gegevens in een CA	13
1.10	Richtingsveld in het fasevlak	17
1.11	Simulaties met constante verbindingsfactoren en een homogene populatieverdeling .	19
1.12	Simulaties met niet-constante verbindingsfactoren en een heterogene populatieverdeling	19
2.1	Tesselatie van onregelmatige cellen die gebruikt wordt in Hoofdstuk 2	28
2.2	Histogram van het aantal buren en de gedeelde omtrek	29
2.4	Gemiddelde connectiviteit bij von Neumann-omgeving met restricties	31
2.5	Genormaliseerde MLE in functie van \overline{C} van regel 98, 105, 250 en 132	34
2.6	Genormaliseerde MLE in functie van \overline{C} van regel 180, 136, 184 en 174	35
3.1	Graafvoorstelling van een regelmatige tesselatie met periodieke randvoorwaarden	39
3.2	Grafiek van het gesimuleerd aantal vatbare en geïnfecteerde knaagdieren bij verschil- lende omgevingsomstandigheden	40
3.3	De waarde van $y(0, t)$ die de richting aangeeft waarheen een knaagdier kan migreren indien zich daar een lege cel bevindt	41
3.4	Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie	

3.5	Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie van het aantal tijdstappen t voor $K = 40 \dots $	43
3.6	Tesselatie met in het donkergrijs de regio waar geen infectie kan plaatsvinden ($p_i = 0$) en in het lichtgrijs de regio waar p_i volgt uit Vgl. (3.1)	43
3.7	Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie van het aantal tijdstappen t voor $K = 40$ en een heterogene virulentie zoals weergegeven in Fig. 3.6	44
4.1	Landbedekking in het studiegebied	48
4.2	Histogram van het aantal buren en de gedeelde omtrek	49
4.3	Vergelijking tussen een model met DV en het CASIR model	50
4.4	Gesimuleerde spatio-temporele evolutie van het aantal geïnfecteerden gebruikmakend van (a) het model van White <i>et al.</i> en (b) het CASIR model waarbij een homogene populatie en een constante virulentie wordt beschouwd	51
4.5	Gesimuleerde spatio-temporele evolutie van het aantal geïnfecteerden gebruikmakend van (a) het model van White <i>et al.</i> en (b) het CASIR model waarbij een homogene populatie en een niet-constante virulentie wordt beschouwd	52
4.6	Fouten tussen voorwaartse en achterwaartse verschil bij verschillende perturbatiefactoren voor ρ en $v_{0,1}$	53
4.7	Relatieve gevoeligheid van het CASIR model voor ρ variërend van 0.1 tot 0.9 met verschillende evaluatiefuncties	55
4.8	Het gemiddelde van de relatieve gevoeligheden van de individuele cellen van het CASIR model voor de verschillende virulenties	56
4.9	Ruimtelijke verspreiding van de gevoeligheid in elke cel van de onregelmatige tesse- latie op tijdstap 8, voor de intra- en intercelvirulentie bij landbouwgronden	57

Lijst van tabellen

1.1	Spatio-temporele modelvormen	3
1.2	Regeltabel van Regel 30	7
2.1	Aantal keer dat de regels in Klasse 2 evolueren naar een superstabiel evenwicht	30
2.2	Invloed van de topologie	33
2.3	Bifurcatie	35
4.1	Het aandeel van de zes verschillende landbedekkingsklassen in de tesselatie	48
4.2	De waarden voor de modelparameters wanneer er een lokale gevoeligheidsanalyse wordt uitgevoerd op een andere modelparameter	53
4.3	De waarden voor alle modelparameters waarmee de gesimuleerde gegevens werden bekomen	58

Lijst met afkortingen en symbolen

Afkorting Betekenis

cellulaire automa(a)t(en)
CA-gebaseerd ziekteverspreidingsmodel
coupled map lattice
individu-gebaseerd model
geografisch informatiesysteem
graphical user interface
maximale Lyapunov exponent
SI model beschreven in Sectie 3.1
SI model beschreven in Sectie 3.2
(partiële) differentiaalvergelijking(en)
som van de absolute fouten
ziekteverspreidingsmodel met vatbaren en geïnfecteerden
ziekteverspreidingsmodel met vatbaren, geïnfecteerden en herstelden
som van de gekwadrateerde fouten

Symbool	Betekenis
A_t	ensemble dat alle replica's r bevat op tijdstap t
a	snelheid waarmee geïnfecteerden de ziekte verspreiden; de virulentie
α	reflectiecoëfficiënt
В	omgevingsfunctie van een cel in de tesselatie
B^*	omgevingsfunctie die de centrale cel c_i niet bevat
b	snelheid waarmee geïnfecteerden sterven
β	contractiecoëfficiënt
\mathcal{C}	cellulaire automaat
C	constante
\overline{C}	gemiddelde connectiviteit
\overline{C}^{\dagger}	bifurcatiepunt
C_i	gemiddelde connectiviteit van alle cellen die informatie kunnen ontvangen van c_i
c_i	i^{de} cel in de tesselatie
$ ilde{c_i}$	cel waarvan de toestand aangepast wordt tijdens een asynchrone tijdstap
$c_{i,j}$	cel in een regelmatige 2D tesselatie op rij i en kolom j
d_{ij}	afstand tussen knopen i en j
E	ensemble van initiële perturbaties
ε	perturbatiefactor
η	modelparameter van White et al. [1]
f	groeifunctie van X
Φ	familie van functies die de toestand van de cellen in de volgende tijdstap bepalen
ϕ_i	functie die de toestand van c_i in de volgende tijdstap bepaalt
G	graaf
g	groeifunctie van Y
γ	expansiecoëfficiënt
h	vector die de verschillen tussen twee modellen weergeeft
Ι	aantal of fractie geïnfecteerden
I^+	fractie geïnfecteerden bij een positieve perturbatie
I^-	fractie geïnfecteerden bij een negatieve perturbatie
\mathbf{I}_i	ruimtelijke informatie over c_i
I _{obs}	geobserveerde fractie geïnfecteerden
J	Jacobiaan
j	sterfteklok knaagdier
j^{max}	maximale waarde voor j

Symbool	Betekenis
K	draagkracht van het milieu
k	aantal mogelijke toestanden van een CA
κ	schalingsfactor
Λ	doelfunctie
λ	maximale Lyapunov exponent
m	sterftesnelheid ten gevolge van andere oorzaken
μ	gemiddelde fractie van cellen in $B(c_i)$ die invloed hebben op $s(c_i, t+1)$
N	populatiegrootte
ν	restrictie bij een restrictieve von Neumann omgeving
0	modelparameter van White et al. [1]
Ω	zoekruimte
Ω^*	deelruimte van de zoekruimte
P_i	punt in Ω
P_s	voorlopig punt in de zoekruimte Ω
P_u, P_v, P_w	willekeurige punten in de populatie
p	probabiliteit in Simulated Annealing
p_i	kans waarmee een vatbaar knaagdier in c_i geïnfecteerd wordt
p_r	kans waarmee geïnfecteerden herstellen
p_s	kans waarmee herstelden terug vatbaar worden
p_v	kans waarmee vatbaren geïnfecteerd worden
q	modelparameter van White et al. [1]
R	aantal of fractie herstelden
$R_{ au}^T$	regelnummer van een totalistische CA met maximumsom $ au$
r^{j}	replica van een tesselatie met enkel $r^j(c_j,t) \neq s(c_j,t)$
ρ	fractie geïnfecteerden die herstelt tijdens één tijdstap
S	eindige verzameling van k toestanden
s_0	initiële toestand van alle cellen binnen de tesselatie
s_0^*	initiële conditie waarbij $s(c_i, 0) \neq s^*(c_i, 0)$
$s(c_i, t)$	toestand van c_i op tijdstap t
σ	straal van een omgeving B
${\mathcal T}$	eindige aftelbare tesselatie
T	aantal tijdstappen
au	maximale som van de toestanden van de buurcellen die beschouwd wordt
θ	parameter die perturbeerd wordt

Symbool	Betekenis
θ^+	positief geperturbeerde parameter
θ^{-}	negatief geperturbeerde parameter
U	uniforme dichtheidsfunctie
V	aantal of fractie vatbaren
\mathbf{v}_0	overdrachtsnelheden tussen individuen levend in dezelfde cel
\mathbf{v}_1	overdrachtsnelheden tussen individuen levend in aaneengesloten cellen
v_c	willekeurig gekozen punt in de zoekruimte Ω
v_n	nieuw punt in de zoekruimte Ω
$\mathbf{w}_{i,j}$	contactkarakteristieken tussen c_i en c_j
x	toestand knaagdier
χ	probabiliteit bij Differential Evolution
ξ	fractie vatbaren op gehele populatie
y	migratierichting knaagdier
z	geboorteklok knaagdier
z^{max}	maximale waarde voor z

Samenvatting

Ziekteverspreiding wordt al eeuwenlang gemodelleerd met (partiële) differentiaalvergelijkingen. Deze modellen worden gebruikt om een efficiënt vaccinatieprogramma op te stellen. Echter, partiële differentiaalvergelijkingen zijn vaak moeilijk op te lossen en vereisen veel randvoorwaarden. Cellulaire automaten (CA) discretiseren het tijds-, ruimte- en toestandsdomein waardoor met eenvoudige transitieregels complexe interacties kunnen gemodelleerd worden. In deze scriptie werd onderzocht of voor een CA-gebaseerd ziekteverspreidingsmodel (CASIR) parametrisatie mogelijk is en welke zaken hiervoor in acht dienen genomen te worden.

In het eerste hoofdstuk worden de definiëring en enkele voorbeelden van cellulaire automaten besproken. Verder wordt er dieper ingegaan op enkele ziekteverspreidingsmodellen en het CASIR model dat geparametriseerd zal worden. Als laatste worden de parametrisatie van CA en de optimalisatietechnieken die daarvoor nodig zijn bestudeerd. De volgende twee hoofdstukken bouwen op naar Hoofdstuk 4 waarin een antwoord zal geven op de onderzoeksvraag.

De stabiliteit van een dynamisch systeem is van groot belang voor de parametrisatie van een dynamisch systeem. Bij continue systemen wordt onder andere de maximale Lyapunov exponent (MLE) gebruikt om de stabiliteit te bepalen. Deze maat voor stabiliteit werd in Hoofdstuk 2 toegepast op discrete systemen. Voor alle 256 totalistische onregelmatige regels met maximum som zeven en twee mogelijke toestanden werd de MLE berekend bij verschillende topologie. Hieruit bleek dat deze regels in verschillende categorieën kunnen ingedeeld worden naargelang hun gedrag bij wijzigingen in de topologie. Bij verschillende regels werden één of meerdere bifurcatiepunten vastgesteld, waaruit volgt dat een kleine wijziging in de topologie een grote invloed kan hebben op het gedrag van CA.

In Hoofdstuk 3 wordt de stabiliteit van een CA-gebaseerd SI model onderzocht. Echter, tijdens de implementatie van dit model werden veel inconsistenties en onvolledigheden in het artikel gevonden. Om deze reden werd het model aangepast. Vervolgens werden enkele simulaties uit het artikel uitgevoerd met het aangepaste model. Tot slot werd de MLE voor het aangepaste model berekend om de stabiliteit te bepalen.

Ten slotte wordt de parametrisatie van het CASIR model bestudeerd. Eerst worden enkele simulaties met het CASIR model vergeleken met simulaties van zowel een continu als een discreet ziekteverspreidingsmodel. Daarna werd een lokale gevoeligheidsanalyse uitgevoerd op de belangrijkste modelparameters, waaruit de invloed van een kleine wijziging in de modelparameters op de simulatieresultaten van model kan afgeleid worden. Als laatste werd onderzocht of het CASIR model kan geparametriseerd worden en welke optimalisatietechnieken daarvoor gebruikt kunnen worden. Er worden vier verschillende optimalisatietechnieken toegepast en de rekentijd voor elk van deze technieken wordt onderzocht.

Inleiding

Verscheidene besmettelijke ziekten hebben door de jaren heen een groot aantal dodelijke slachtoffers gemaakt. Een goed georganiseerd vaccinatieprogramma zou duizenden levens kunnen redden. Echter, om een efficiënt vaccinatieprogramma op te zetten dient de dynamiek van de ziekte goed gekend zijn. Wiskundige modellen trachten aan de hand van observaties de verspreiding van ziekten accuraat te simuleren en zo het vaccinatiebeleid bij te sturen.

Het modelleren van de verspreiding van ziekten gebeurt al enkele eeuwen. Het aantal vatbare en geïnfecteerde individuen en eventueel het aantal herstelde individuen worden hierbij gesimuleerd doorheen de tijd. Echter, deze modellen houden veelal geen rekening met een ruimtelijke component, waardoor ze enkel bij een perfect gemengde populatie kunnen gebruikt worden. Bij partiële differentiaalvergelijkingen wordt deze ruimtelijke component wel beschouwd, maar vaak zijn deze niet of moeilijk op te lossen. Cellulaire automaten (CA), daarentegen, zijn discreet in het ruimte-, tijdsen toestandsdomein, terwijl hun dynamiek wordt gestuurd door eenvoudige transistieregels. Ze kunnen complexe systemen modelleren en worden reeds toegepast binnen verscheidene domeinen van de wetenschap. CA-gebaseerde ziekteverspreidingsmodellen zijn talrijk aanwezig in de literatuur. Deze hebben het grote voordeel dat er makkelijk kan gewerkt worden met een heterogene verdeling van de virulenties. Bijvoorbeeld, binnen een dichtbevolkte regio zullen ziekten zich sneller verspreiden dan in een dunbevolkte regio.

In deze scriptie wordt nagegaan of het identificeren van de parameters van een bestaand CA-gebaseerd ziekteverspreidingsmodel (CASIR) mogelijk is op basis van observaties. Bij dit model is de virulentie afhankelijk van de landbedekkingsklasse. Er worden zes landbedekkingsklassen gedefinieerd die elk worden getypeerd door een virulentie binnen de cel en tussen de cellen. Daarnaast bevat het model één globale modelparameter, namelijk de fractie geïnfecteerden die in één tijdstap herstelt.

Eerst wordt de bestaande literatuur over CA, ziekteverspreidingsmodellen en de parametrisatie van CA-gebaseerde modellen beschreven. Voor de volgende hoofdstukken wijkt de opbouw van deze scriptie af van de vooropgestelde indeling van de masterproef voor de afstudeerrrichting Milieutechnologie, aangezien een opbouwend onderzoek gevoerd werd in drie delen. Aangezien de stabiliteit van een dynamisch systeem belangrijk is voor de parametrisatie van de vergelijkingen die dat systeem beschrijven, wordt in Hoofdstuk 2 een maat voor de stabiliteit van CA, namelijk de maximale Lyapunov exponent, bekeken. Het effect van een wijziging in de topologie op de stabiliteit van zulke dynamische systemen wordt onderzocht voor 256 verschillende totalistische CA. Het volgende hoofdstuk bespreekt de stabiliteit van een CA-gebaseerd ziekteverspreidingsmodel dat twee discrete toestanden bevat, namelijk vatbare en geïnfecteerde knaagdieren. Om hiervan de stabiliteit te onderzoeken wordt de maximale Lyapunov exponent bepaald. In Hoofdstuk 4 wordt een CA-gebaseerd ziekteverspreidingsmodel met drie continue toestanden (CASIR) onderzocht. Er wordt hierbij een antwoord gegeven op de onderzoeksvraag, namelijk of het CASIR model kan geparametriseerd worden en welke aandachtspunten hiervoor in acht dienen genomen te worden. Na een vergelijking met enkele andere ziekteverspreidingsmodellen en een uitgebreide gevoeligheidsanalyse van de belangrijkste modelparameters worden optimale parameterwaarden gezocht met vier verschillende optimalisatietechnieken wat tevens de mogelijkheid biedt om te verifiëren welke techniek de voorkeur verdient voor de parametrisatie van CA-gebaseerde modellen. Het laatste hoofdstuk vat alle resultaten en de conclusies samen en belicht het onderzoek dat in het verlengde van deze scriptie kan uitgevoerd worden.

Literatuurstudie

Om biologische processen te simuleren zijn zowel tijd als ruimte belangrijke componenten. Aangezien partiële differentiaalvergelijkingen (PDV) meerdere onafhankelijke variabelen bevatten, worden deze vaak gebruikt om biologische processen te modelleren [2, 3]. Bij PDV zijn zowel tijd, ruimte als toestand continu, maar omdat zij veelal niet of moeilijk op te lossen zijn, worden soms één of meerdere domeinen gediscretiseerd. In Tab. 1.1 wordt een overzicht gegeven van de acht mogelijke spatio-temporele modelvormen, waarbij de keuze voor één van deze modelvormen afhankelijk is van de toepassing [4].

Tabel 1.1: Spatio-temporele modelvormen	[4]
---	-----

Toestand	Ruimte	Tijd	Benaming	Toepassingen
discreet	discreet	discreet	cellular automaton of	[5, 6]
			individual-based model (IBM)	[6, 7]
continu	discreet	discreet	coupled map lattice (CML)	[8, 9, 10, 11]
discreet	discreet	continu	interacting particle system	[12, 13]
continu	discreet	continu	reaction-dispersal network	[14]
discreet	continu	discreet	neighborhood models	[4, 15]
discreet	continu	continu	spatial point processes	[16, 17, 18]
continu	continu	discreet	reaction-dispersal network	[4]
continu	continu	continu	partial differential equations	[19, 20, 21]

1.1 Cellulaire automaten

1.1.1 Definitie en historiek

In 1948 zocht von Neumann een vereenvoudigd model voor biologische evolutie [22]. Hij volgde hiervoor de suggesties van Ulam om het vlak op te delen in cellen, waaruit de eerste CA ontstond. In de jaren zeventig introduceerde Conway de 2D CA 'Game of life' [23], maar het duurde nog tot 1983

vooraleer Wolfram een referentiewerk publiceerde over de eigenschappen van CA [24], wat geleid heeft tot een revival van CA binnen de axacte en toegepaste wetenschappen [25].

De wiskundige formulering van een CA wordt gegeven in Definitie 1 [26].

Definitie 1 (Cellulaire automaat) Een cellulaire automaat (CA) kan voorgesteld worden als een sextet

$$\mathcal{C} = \langle \mathcal{T}, S, s, s_0, B, \Phi \rangle,$$

met

- (1) \mathcal{T} een aftelbare, oneindige tesselatie van een n-dimensionale Euclidische ruimte \mathbb{R}^n , met cellen $c_i, i \in \mathbb{N}$.
- (2) S een eindige verzameling van k toestanden, dikwijls $S \subset \mathbb{N}$.
- (3) De output functie $s : \mathcal{T} \times \mathbb{N} \to S$ beschrijft de toestand van cel c_i op de discrete tijdstap t, d. i. $s(c_i, t)$.
- (4) De functie $s_0 : \mathcal{T} \to S$ wijst aan elke cel c_i haar initiële toestand toe.
- (5) De omgevingsfunctie B: $\mathcal{T} \to \bigcup_{p=1}^{\infty} \mathcal{T}^p$ bakent voor elke cel c_i een eindige sequentie $B(c_i) = (c_{i_j})_{j=1}^{|B(c_i)|}$ af bestaande uit $|B(c_i)|$ verschillende cellen c_{i_j} .

(6) $\Phi = (\phi_i)_{i \in \mathbb{N}}$ is een familie van functies

$$\phi_i: S^{|B(c_i)|} \to S,$$

waarbij elke ϕ_i de dynamica van cel c_i stuurt, d. i.

$$s(c_i, t+1) = \phi_i(\tilde{s}(B(c_i), t)),$$

met $\tilde{s}(B(c_i), t) = (s(c_{i_i}, t))_{i=1}^{|B(c_i)|}$.

Met een **tesselatie** \mathcal{T} wordt de indeling van de Euclidische ruimte in cellen bedoeld, waarbij overlap noch lege ruimten toegestaan zijn. Indien iedere cel c_i in \mathcal{T} hetzelfde aantal buren heeft, ongeacht de gebruikte omgeving, spreekt men van een regelmatige tesselatie [27]. In het vlak komt dit veelal neer op vierkante cellen, maar ook zes- en driehoekige cellen komen voor. Een tesselatie met onregelmatige cellen zoals afgebeeld in Fig. 1.1, kan in sommige gevallen een betere weergave leveren van natuurlijke systemen. Een tesselatie kan ook als een graaf voorgesteld worden, waarbij cellen die tot elkaars omgeving behoren met elkaar worden verbonden (Fig. 1.2).

Een onregelmatige tesselatie van \mathbb{R}^n kan gemakkelijk gecreëerd worden met behulp van een Voronoitesselatie. Hiertoe worden initieel een aantal willekeurige punten gekozen in \mathbb{R}^n waarrond vervolgens polygonen worden afgebakend zodanig dat alle punten binnen een zeker polygoon dichter bij het centrum van dat polygoon liggen, dan bij enig ander punt [28]. De keuze voor een bepaalde tesselatie hangt af van de eigenschappen van het systeem [4].



Figuur 1.1: Een onregelmatige tesselatie in \mathbb{R}^2

Zoals aangegeven in Definitie 1 is \mathbb{R}^n oneindig uitgestrekt, maar voor de uitvoering van computersimulaties is er uiteraard een begrenzing nodig. Daarom moeten **randvoorwaarden** vastgelegd worden. Fig. 1.3 toont een ordening van de cellen in een 1D CA, waarbij c_0 en c_{n+1} twee imaginaire cellen zijn die gebruikt worden om de randvoorwaarden te implementeren. Drie types randvoorwaarden worden vaak gebruikt [4]:

- 1. vast: $s(c_0, t) = 0$ en $s(c_{n+1}, t) = 0, \forall t$
- reflecterend: s(c₀, t) = s(c₁, t) en s(c_{n+1}, t) = s(c_n, t), ∀ t
 Dit wordt bijvoorbeeld gebruikt in modellen voor waterdieren in een vijver, aangezien de dieren aan de rand worden teruggekaatst.
- periodiek: s(c₀, t) = s(c_n, t) en s(c_{n+1}, t) = s(c₁, t), ∀ t
 De tegenovergestelde randen van de tesselatie worden met elkaar verbonden, waardoor een torus gevormd wordt. Dit wordt gebruikt om een oneindige ruimte te simuleren.

De **omgeving** B wordt op voorhand vastgelegd voor elke cel en geeft de buurcellen weer waarmee c_i interageert tijdens de opeenvolgende tijdstappen. De straal van B wordt gedefinieerd met behulp van σ , zodat voor een 1D CA waarvoor geldt dat σ gelijk is aan twee $B(c_i) = \{c_{i-2}, c_{i-1}, c_i, c_{i+1}, c_{i+2}\}$ zal zijn. Een veel gebruikte ordening van de cellen in een regelmatige 2D CA wordt weergegeven in Fig. 1.4. In het vlak worden verschillende omgevingen gedefinieerd. De von Neumann omgeving van een cel c_i (Fig. 1.5(a)) omvat enkel de cellen die één zijde gemeenschappelijk hebben met c_i , terwijl in de Moore omgeving van een cel c_i (Fig. 1.5(b)) ook cellen die slechts één punt gemeenschappelijk hebben worden opgenomen. Voor onregelmatige tesselaties wordt soms een von Neumann omgeving



Figuur 1.2: Graaf-voorstelling van de onregelmatige tesselatie afgebeeld in Fig. 1.1



Figuur 1.3: Ordening van de cellen in een 1D CA met aanduiding van de imaginaire cellen c_0 en c_{n+1}

met restricties gebruikt (Fig. 1.6), die enkel de cellen bevat die minstens een vooropgestelde proportie $\nu \in]0, 1[$ van de totale omtrek van c_i uitmaken. Deze omgevingen kunnen intuïtief uitgebreid worden naar tesselaties van een 3- of meerdimensionale ruimte [25]. In het vervolg van deze scriptie wordt de term topologie gebruikt om het geheel van cellen en hun interconnecties aan te duiden.

Tijdens elke tijdstap wordt de toestand van elke cel geüpdatet met behulp van een **transitiefunctie** ϕ_i . Deze beschrijft de toestand van cel c_i op de volgende tijdstap d. i. $s(c_i, t+1)$, op basis van de toestand van c_i en van de cellen in de omgeving van c_i in de huidige tijdstap, gegeven door respectievelijk $s(c_i, t)$ en $\tilde{s}(B(c_i), t)$. Bij homogene CA wordt de toestand van elke cel met eenzelfde transitiefunctie geüpdatet. Totalistische CA zijn een familie binnen de homogene CA en worden omschreven in Definitie 2 [29].

Definitie 2 (Totalistische CA)

Een totalistische CA is een CA die voldoet aan voorwaarden 1 t. e. m. 5 van Definitie 1, terwijl voor voorwaarde 6 geldt: $\exists \Omega : \mathbb{N} \to S$ waarvoor geldt dat $s(c_i, t+1) = \phi_i(\tilde{s}(B(c_i), t)) = \Omega(\sigma_i)$ met $\sigma_i = \sum_{c_i \in B(c_i)} s(c_j, t).$

1.1.2 Elementaire CA

De omgeving van een 1D CA omvat $2\sigma + 1$ cellen, wanneer elke cel k verschillende toestanden kan aannemen, zijn er $k^{k^{(2\sigma+1)}}$ mogelijke transitiefuncties die elk een regelnummer krijgen. Er bestaan bijgevolg 256 verschillende elementaire CA: dit zijn 1D CA met $\sigma = 1$ en k = 2 ($S = \{0, 1\}$), welke worden benoemd met een decimaal getal [30]. Dit getal wordt bekomen door de transitiefuncties voor



Figuur 1.4: Ordening van de cellen in een 2D CA

te stellen in een tabel en dan via het binaire telsysteem dit getal te berekenen. Zo wordt de regeltabel van Regel 30 gegeven in Tab. 1.2, waaruit het regelnummer als volgt kan afgeleid worden [31]:

$$\begin{split} \Phi(1,1,1) \times 2^7 + \Phi(1,1,0) \times 2^6 + \Phi(1,0,1) \times 2^5 + \Phi(1,0,0) \times 2^4 + \Phi(0,1,1) \times 2^3 \\ + \Phi(0,1,0) \times 2^2 + \Phi(0,1,1) \times 2^1 + \Phi(0,0,0) \times 2^0 &= 30 \end{split}$$

Tabel 1.2: Regeltabel van Regel 30 voor een elementaire CA ($\sigma = 1, k = 2$), waarbij de cellen geordend
worden zoals aangegeven in Fig. 1.3 [31]

$s(c_{i-1},t)$	$s(c_i, t)$	$s(c_{i+1},t)$	$s(c_i, t+1)$
1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	1
0	0	1	1
0	0	0	0



Figuur 1.5: Een voorbeeld van de von Neumann (a) en de Moore (b) omgeving van cel c_1 een in onregelmatige tesselatie

De elementaire CA worden ingedeeld in vier verschillende gedragsklassen, naargelang hun evolutie in de tijd [24]. In klasse I bevinden zich de CA die onafhankelijk van de beginconditie evolueren naar een homogene toestand, zoals Regel 249 (Fig. 1.7(a)). Klasse II omvat CA die steeds naar een periodieke, geneste structuur evolueren, bijvoorbeeld Regel 4 (Fig. 1.7(b)). Klasse III omvat de CA waarbij een willekeurig patroon ontstaat, waarin meestal driehoekige structuren verweven zijn, zoals voor Regel 126 wordt waargenomen (Fig. 1.7(c)). De laatste klasse, klasse IV, is niet zo nauwkeurig omlijnd en omvat CA waarbij complexe interagerende patronen worden voortgebracht zonder voorspelbare uitkomst, zoals Regel 110 (Fig. 1.7(d)) [30].

1.1.3 Evenwicht en stabiliteit van CA

Het indelen van CA in gedragsklassen is veelal gebaseerd op hun evenwicht en stabiliteit. Er kan gebruik gemaakt worden van de Hammingafstand [32, 33], de Langton parameter [34] of Lyapunov exponenten [35, 36, 37] om de dynamiek te kwantificeren en de CA in de verschillende gedragsklassen onder te brengen.

De stabiliteit van dynamische systemen kan bepaald worden aan de hand van Lyapunov exponenten, welke de mate beschrijven waarin twee fasebanen, vanuit initiële condities die oneindig dicht bij elkaar liggen, zich ten opzichte van elkaar bewegen in de faseruimte. De Lyapunov exponent (λ) van een continu systeem wordt als volgt gedefinieerd [38]:

$$\Delta(t) \approx \Delta(0) \ e^{\lambda t} \tag{1.1}$$



Figuur 1.6: Een voorbeeld van de restrictieve von Neumann omgeving van c_1 met $\nu = 0.20$, zodat deze enkel de centrale cel en de cellen die meer dan 20% van de totale omtrek van c_1 delen bevat

waarbij $\Delta(t)$ de afstand tussen de fasebanen op tijdstip t voorstelt en $\Delta(0)$ de initiële afstand tussen de fasebanen is. Er zijn voor een dynamisch systeem evenveel Lyapunov exponenten als er dimensies zijn in de faseruimte. Echter, om de stabiliteit te bepalen wordt meestal de maximale Lyapunov exponent berekend [38]:

$$\lambda_{\max} = \lim_{t \to \infty} \lim_{\Delta(0) \to 0} \frac{1}{t} \ln\left(\frac{\Delta(t)}{\Delta(0)}\right) \,. \tag{1.2}$$

Om de dynamiek van een continu systeem, $\mathbf{y}' = \mathbf{f}(\mathbf{y}, \mathbf{x})$, te onderzoeken kan tevens gebruik gemaakt worden van Jacobianen, die de partiële afgeleiden naar elk van de variabelen bevat:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$
(1.3)

Analoog aan een continu dynamisch systeem kan voor een CA de maximale Lyapunov exponent berekend worden [39]. Hiervoor wordt gestart met twee initiële condities s_0 en s_0^* , die slechts in één cel c_j van mekaar verschillen zodat $s(c_j, 0) \neq s^*(c_j, 0)$. Dit is bij CA het kleinst mogelijke verschil dat kan verkregen worden en c_j wordt hierbij de geperturbeerde cel genoemd. De vector die de verschillen tussen de toestanden van twee modellen weergeeft, van een CA met $S = \{0, 1\}$ wordt gedefinieerd als

$$h(\cdot, t) = s(\cdot, t) \oplus s^*(\cdot, t), \tag{1.4}$$



Figuur 1.7: Elementaire CA die vanuit een willekeurige initiële conditie evolueren naar (a) een homogene toestand (Regel 249), (b) een periodieke structuur (Regel 4), (c) een willekeurig patroon met driehoekige structuren (Regel 126) en (d) een complex interagerend patroon (Regel 110)

met \oplus de mod 2 operator die elementsgewijs wordt toegepast. Vermits s_0 en s_0^* slects in één cel van mekaar verschillen, geldt er dat $\sum_{c_i} h(c_i, 0)$ gelijk is aan 1 [40]. Wanneer $t \to \infty$ zal bij een onstabiel systeem $\sum_{c_i} h(c_i, t)$ steeds toenemen, maar door het discreet gedrag moet rekening gehouden worden met het feit dat de verstoringen die te wijten zijn aan éénzelfde originele verstoring elkaar kunnen opheffen. Daarom dienen de verstoringen in elke stap bijgehouden te worden en wordt met volgende procedure gewerkt [40]:

- (i) Beschouw twee begincondities $s_0 = s(\cdot, 0)$ en $s_0^* = s^*(\cdot, 0)$ waarvoor geldt dat $\sum_{c_i} h(c_i, 0) = 1$. s_0^* wordt bewaard in A_0 .
- (ii) Bereken $s(\cdot, 1)$ en $s^*(\cdot, 1)$, bereken ook de vector $h(\cdot, 1)$ als volgt:

$$h(c_j, 1) = \begin{cases} 1, & \text{als } s^*(c_j, 1) \neq s(c_j, 1), \\ 0, & \text{anders.} \end{cases}$$
(1.5)

- (iii) Creëer voor elke c_j waarvoor geldt dat $h(c_j, 1) = 1$ een replica r^j zodanig dat $r^j(c_j, 1) = s^*(c_j, 1) = \bar{s}(c_j, 1)$, waarbij $\bar{s}(c_j, 1)$ is het Boolese complement van $s(c_j, 1)$ en $r^j(c_k, 1) = s(c_k, 1)$, voor elke $c_k \neq c_j$. Bewaar deze replica's in A_1 .
- (iv) Bereken $s(\cdot, 2)$ en $r(\cdot, 2)$ voor elke r in A_1 . Vector h wordt dan als volgt berekend:

$$h(c_j, 2) = |\{r \in A_1 | r(c_j, 2) \neq s(c_j, 2)\}|.$$
(1.6)

- (v) Creëer voor elke c_j en $r \in A_1$ waarvoor geldt dat $r(c_j, 2) \neq s(c_j, 2)$ een replica r^j zodanig dat $r^j(c_j, 2) = r(c_j, 2) = \bar{s}(c_j, 2)$ en $r^j(c_k, 2) = s(c_k, 2)$ voor elke $c_k \neq c_j$. Bewaar deze replica's in $A_2 = \{(r^j, h(c_j, 2)) | r \in A_1 \land r(c_j, 2) \neq s(c_j, 2)\}$
- (vi) Herhaal stappen (iv) en (v) voor elke volgende tijdstap t + 1. Bereken telkens $h(\cdot, t + 1)$ en construeer A_{t+1} op basis van de replica's in A_t .

Gebruikmakend van dit algoritme, kan de maximum Lyapunov exponent (MLE) van een CA berekend worden als [40]:

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \ln \left(\frac{\sum_{c_i} h(c_i, t)}{\sum_{c_i} h(c_i, 0)} \right).$$
(1.7)

In de praktijk kan deze berekening niet uitgevoerd worden voor $t \to \infty$ en dient een eindig aantal tijdstappen (T) gebruikt te worden. Bovendien zal ook een eindige tesselatie gekozen worden. Bij CA die evolueren naar een homogene configuratie waarin alle cellen dezelfde waarde aannemen, is $\lambda = -\infty$. Wanneer $\lambda < 0$ is de CA niet gevoelig aan perturbaties van s_0 en dus stabiel. Wanneer $\lambda > 0$ daarentegen zullen de fasebanen divergeren en is de CA onstabiel [40].

Voor een CA waarbij $S = \{0, 1\}$ kan een Boolese afgeleide als volgt berekend worden [40],

$$\frac{\partial \phi_i}{\partial s(c_{i_k}, t)} = \phi_i(s(c_{i_1}, t), \dots, \bar{s}(c_{i_k}, t), \dots, s(c_{i_{|B(c_i)|}}, t)) \\ \oplus \phi_i(s(c_{i_1}, t), \dots, s(c_{i_k}, t), \dots, s(c_{i_{|B(c_i)|}}, t))$$
(1.8)

De elementen van de Jacobiaan worden dan, in overeenstemming met de Jacobiaan van een continue systeem (cfr. Vgl. (1.3)), gegeven door

$$J_{ij} = \begin{cases} \frac{\partial s(c_i, t+1)}{\partial s(c_j, t)}, & \text{als } c_j \in B(c_i), \\ 0, & \text{anders,} \end{cases}$$
(1.9)

en zijn één wanneer het aanpassen van $s(c_j, t)$ een verandering in $s(c_i, t + 1)$ veroorzaakt en nul wanneer dit niet gebeurt. Het gemiddelde aandeel van cellen in de omgeving van c_i die invloed hebben op $s(c_i, t + 1)$ is [40]:

$$\mu(t) = \frac{1}{|\mathcal{T}^*|} \sum_{c_i} \frac{1}{|B(c_i)|} \sum_{j=1}^{|B(c_i)|} J_{ii_j}.$$
(1.10)

Het geometrisch gemiddelde van μ over een groot aantal tijdstappen (T) wordt dan:

$$\bar{\mu} = \left(\prod_{t=1}^{T} \mu(t)\right)^{1/T},\tag{1.11}$$

waaruit volgt dat hoe hoger $\bar{\mu}$ is, hoe gevoeliger het systeem zal zijn aan de initiële condities. Het systeem zal dus onstabieler worden naarmate $\bar{\mu}$ toeneemt [40].

1.1.4 Perspectieven van CA-gebaseerde modellen

Microscopisch CA-gebaseerde modellen worden in heel wat wetenschappelijke disciplines gebruikt, maar ze worden vooral toegepast binnen de biologische wetenschappen. Binnen de microbiologie bijvoorbeeld kunnen ze celgroei en biofilmvorming relatief eenvoudig modelleren [41, 42, 43, 44, 45]. Complexe PDV die heel moeilijk te interpreteren zijn, worden hierbij vervangen door eenvoudige transitiefuncties. Er zijn tevens medische toepassingen; zo werden tumorgroei [46, 47, 48, 49] en immuunsystemen [50, 51] reeds CA-gebaseerd gemodelleerd. In deze disciplines zijn veel experimentele gegevens verzameld, hetgeen een kwantitatieve validatie van deze modellen mogelijk maakt.

In hydrodynamica worden CA-gebaseerde modellen gebruikt om het systeem op microscopisch niveau te doorgronden. Er werd aangetoond dat deze microscopische benadering van vloeistofstroming de Navier-Stokes vergelijking benadert [31, 52, 53]. Ook in de thermodynamica worden CAgebaseerde modellen gebruikt omdat deze modellen eenvoudig met complexe, anisotrope media kunnen omgaan [54, 55]. Tot op heden worden CA-gebaseerde modellen vooral gebruikt om diffusieprocessen te modelleren, vermits de ruimtelijke component daar van heel groot belang is.

Macroscopisch Landdynamische processen, zoals verstedelijking, worden gemodelleerd met CAgebaseerde modellen, die als beleidssturend middel kunnen gebruikt worden en belangrijke beslissingen kunnen ondersteunen. Deze modellen kunnen tevens toegepast worden om toekomstig landgebruik te voorspellen [56, 57, 58, 59, 60] en effecten van de klimaatverandering te voorspellen [61, 62]. Ze zijn belangrijk voor het bekijken van de impact van de mens op het milieu, die leidt tot vele problemen zoals hongersnood, verlies van biodiversiteit, bodemerosie ...

Ook verspreidingsprocessen, zoals bosbranden [5, 63, 64, 65, 66] en infectieziekten [67, 68, 69], kunnen met CA-gebaseerde modellen beschreven worden. Immers, discrete modellen laten toe om ruimtelijke informatie te implementeren zodanig dat een realistische simulatie verkregen worden.

Naast talrijke voorbeelden van biologische processen, zijn er CA-gebaseerde modellen binnen de economie [70, 71], verkeer [72, 73] en politiek [74, 75].

1.1.5 Koppeling met geografische informatiesystemen

Een geografisch informatiesysteem (GIS) wordt op vele verschillende manieren gedefinieerd. Enerzijds wordt een GIS gezien als een toolbox die alle ruimtelijke gegevens verzamelt, bijhoudt en visualiseert. Anderzijds wordt het gebruikt als een informatiesysteem, een databank waaruit informatie kan gehaald worden om te analyseren, modelleren en voorspellen [76].

De integratie van teledetectie, GPS en luchtfoto's zorgt ervoor dat steeds meer onderzoek met en over GIS wordt uitgevoerd, waardoor het als een aparte tak binnen de wetenschap wordt gezien. GIS is op een korte tijd uitgegroeid tot een miljardenbusiness door de invoering van 'graphical user interfaces' (GUI), de verhoogde beschikbaarheid van meer performante computers en het eenvoudige gebruik van data [76].

Kenmerken van het aardoppervlak worden bewaard en in kaart gebracht als punten, lijnen of polygonen (Fig. 1.8). GIS kan ook menselijke activiteiten in kaart brengen, zoals een populatiekaart en verspreiding van ziektegevallen, en gebeurtenissen. Er wordt een tijdsdimensie toegevoegd aan GIS, om het moment en de duur van zulke gebeurtenissen weer te geven [76].



Figuur 1.8: Voorbeelden van geometrische objecten die in een GIS worden onderscheiden [76]

Een vectorgebaseerde GIS omgeving kan gemakkelijk gekoppeld worden aan een CA-gebaseerd model, aangezien bij beide het vlak wordt ingedeeld in onregelmatige polygonen [77]. Aldus kan een model voor de verspreiding van bosbranden gebruik maken van de hoogteligging, de helling en de brandbaarheid van de vegetatie die in een GIS bewaard wordt om een realistisch scenario op te stellen (Fig. 1.9). Op deze manier kunnen ook voorspellingen gemaakt worden over bosbranden in andere gebieden [65].



Figuur 1.9: Integratie van GIS gegevens in een CA [65]

1.2 Modellen voor ziekteverspreiding

1.2.1 Belang

Besmettelijke ziekten hebben door de eeuwen heen al meer dodelijke slachtoffers gemaakt dan oorlogen. Epidemieën hebben de loop van de geschiedenis bepaald, populaties zijn gedecimeerd en bloedlijnen zijn uitgestorven. Wanneer de ziekte slechts in enkele regio's op de wereld voorkomt, spreekt men van een endemie, maar als ze een invloed heeft op een groot gebied en een groot aantal mensen, wordt daarentegen gesproken van een pandemie. Epidemieën onder planten veroorzaken misoogsten, die op hun beurt leiden tot hongersnood. Hieronder worden tien ziektes beschreven die van belang zijn (geweest) voor de recente menselijke ontwikkeling [78].

- Polio verlamt en doodt al gedurende verschillende millenia kinderen. Er is geen geneesmiddel voor polio, maar door vaccinatie is het aantal gevallen de laatste decennia sterk afgenomen, van 350000 geïnfecteerden in 1988 tot slechts 1604 in 2009. Tegenwoordig wordt er gesproken van een endemie, omdat het slechts in vier landen voorkomt [79].
- **Tyfus** komt voor wanneer veel mensen dicht bij elkaar leven. Zo werd het Spaanse leger in 1489 op één maand gereduceerd van 25000 naar 8000 man. Ook tijdens de Eerste Wereldoorlog stierven in Rusland, Polen en Roemenië enkele miljoenen mensen aan tyfus [80].
- Gele koorts wordt verspreid door geïnfecteerde muggen en heeft in het begin van de negentiende eeuw 29000 soldaten van het Franse leger gedood. In Zuid-Amerika en Afrika maakt de ziekte nog steeds doden aangezien er geen werkend geneesmiddel gekend is [81].
- HIV/AIDS is een recente pandemie die sinds 1981 al ongeveer 25 miljoen slachtoffers maakte. Op dit moment zijn 33 miljoen mensen HIV-positief, ze zijn dus drager van het virus dat AIDS veroorzaakt. Het virus wordt overgedragen via seksueel contact of het gebruik van besmette naalden [82].
- Cholera wordt veroorzaakt door *Vibrio cholerae* en wordt vooral verspreid via gecontamineerd water en eten. De voorbije twee eeuwen zijn ten gevolge van zes cholerapandemieën al miljoenen mensen gestorven in China, Japan, Noord-Afrika, het Midden Oosten en Europa [81].
- **Tuberculose** verspreidt zich via de lucht en doodt nog steeds ongeveer twee miljoen mensen per jaar. Deze ziekte werd al beschreven in heel oude teksten en bewijzen ervan zijn teruggevonden in Egyptische mummies [83].
- Malaria is al meer dan 4000 jaar bekend; ook deze ziekte wordt verspreid via muggen. Elk jaar zijn er tussen 300 en 500 miljoen nieuwe gevallen in de Sub-Sahara en meer dan één miljoen mensen hiervan sterven van de ziekte [84].
- **De Zwarte Dood** of de pest is een middeleeuwse pandemie die tussen 1347 en 1351, 25% tot 50% van de Europese bevolking doodde. Ze werd verspreid via vlooien op ratten of via de lucht. Tegenwoordig kan de geneeskunde deze ziekte gemakkelijk behandelen in de vroege stadia [85].

- De Spaanse griep veroorzaakte in 1918 meer dan 40 miljoen doden wereldwijd. Het immuunsysteem van de bevolking was uitermate verzwakt door de Eerste Wereldoorlog en de verspreiding ging snel door massaal transport van troepen. Het was een H1N1 influenza A virus, hetzelfde dat in gewijzigde vorm de recente Mexicaanse grieppandemie veroorzaakte [86].
- De pokken werden ingevoerd door de Spaanse veroveraars in de Nieuwe Wereld. De inheemse bevolking bezat geen immuniteit voor deze ziekte en stierven in groten getale. Er wordt geschat dat de bevolking van 100 miljoen naar vijf tot tien miljoen mensen daalde. Het virus werd zelfs moedwillig verspreid om de vijand uit te moorden. In 1977 is voor het laatst een geval van pokken ontdekt, nu bestaat het virus enkel nog in laboratoria [81].

Gelet op de grote invloed die epidemieën hebben op de samenleving, is het belangrijk om de dynamiek ervan te kennen en zo het beleid voor bijvoorbeeld de uitbouw van een vaccinatieprogramma te ondersteunen [1].

1.2.2 Continue modellen

Het eerste wiskundige model ter beschrijving van ziekteverspreiding dateert van de achttiende eeuw. Bernoulli wou de verspreiding van de pokken modelleren en de kracht van vaccinatie bewijzen. Het model geeft het aantal vatbaren (V) en de totale populatie (N) in functie van de tijd t [87]:

$$\begin{cases} \frac{dV}{dt} = -[a+m(t)]V\\ \frac{dN}{dt} = -abV - m(t)N \end{cases}$$
(1.12)

met $V(0) = N(0) = N_0$, *a* de virulentie, *b* de sterftesnelhied van geïnfecteerde individuen en m(t) de snelheid van sterfgevallen door andere oorzaken voorstellen.

Deze vergelijkingen kunnen in één vergelijking geschreven worden door de fractie vatbaren binnen de volledige populatie te berekenen ($\xi = \frac{V}{N}$) [87]. Immers, er geldt

$$\frac{d\xi}{dt} = \frac{dV}{dt} \frac{N}{N^2} - \frac{dN}{dt} \frac{V}{N^2},$$

waaruit door substitutie van Vgl. (1.12) volgt dat

$$\frac{d\xi}{dt} = \frac{-a\,V\,N - m(t)\,V\,N + a\,b\,V^2 + m(t)\,N\,V}{N^2},$$

of na vereenvoudiging:

$$\frac{d\xi}{dt} = a h \left(b \xi - 1 \right). \tag{1.13}$$

Wanneer deze DV wordt uitgerekend via scheiding van de veranderlijken en splitsen in partieelbreuken, krijgt men:

$$dt = \frac{d\xi}{a\,\xi\,(b\,\xi-1)}$$

$$\Leftrightarrow \quad \int dt = \int \frac{-1}{a\,\xi}\,d\xi + \int \frac{b}{a\,(b\,\xi-1)}\,d\xi$$

$$\Leftrightarrow \quad t + C = \frac{-1}{a}\,\ln(\xi) + \frac{1}{a}\,\ln(b\,\xi-1)$$

$$\Leftrightarrow \quad t + C = \frac{1}{a}\,\ln\left(\frac{b\,\xi-1}{\xi}\right)$$

$$\Leftrightarrow \quad \xi = \frac{1}{b - e^{a\,t}\,e^{C}} \tag{1.14}$$

Met de randvoorwaarde $\xi(0) = 1$ kan C als volgt berekend worden:

$$1 = \frac{1}{b - e^C}$$

$$\Leftrightarrow C = \ln(b - 1)$$
(1.15)

zodat

$$\xi(t) = \frac{1}{b + (1 - b) e^{a t}} \tag{1.16}$$

Kermack en McKendrick [88] beschreven in het begin van de twintigste eeuw een algemeen model voor de verspreiding van ziekten te simuleren. Dit model deelt de populatie op in drie groepen, zijnde de vatbare (V), de geïnfecteerde (I) en de herstelde individuen (R). Het model houdt geen rekening met de aangroei of afname van de populatie en veronderstelt verder een perfecte menging van de populatie, waardoor de temporele dynamiek kan gevat worden door [88]:

$$\begin{cases} \frac{dV}{dt} = -a V I \\ \frac{dI}{dt} = a V I - b I \\ \frac{dR}{dt} = b I \end{cases}$$
(1.17)

waarbij a de virulentie en b de snelheid van verwijdering is. Uit deze vergelijkingen kan besloten worden dat een epidemie zich slechts kan ontwikkelen indien $\frac{dI(0)}{dt} > 0$ of $V(0) > \frac{a}{b}$ (Fig. 1.10).

Wanneer V + I + R = N voor alle t, kan de laatste vergelijking in Stelstel 1.17 herschreven worden als:

$$\frac{dR}{dt} = b\left(N - V - R\right). \tag{1.18}$$

Wanneer V in functie van R berekend wordt, krijgt men als vergelijking voor de fasebaan:

$$\frac{dV}{dR} = -\frac{a}{b} V,$$

1		-	~		æ	λ.	<u> </u>	~	~	N	~	<u></u>	~	-	12	~	<u></u>	~	~	- T	~	~	~		-
		_	۰.		Ť															1					
0.9 0.8			-	*	¥	*	*	*			~	~			-	-	-	-	-	~	~	-	-	~	
	-		-	4	↓	7	*	*	~	-	~	~	>	-	-		~	~	~	~	~	~	1	~ -	-
			-	*	↓	7	$\mathbf{\tilde{x}}$	7	~	-	~	~	>	1	~	~	~	~	1	~	~	1	1	~	
			-	*	↓	74	$\mathbf{\tilde{x}}$	$\mathbf{\tilde{x}}$	~	-	~	~	1	1	~	~	~	~	~	~	~	1	1	~	
			-	*	\downarrow	λ.	× .	$\mathbf{\tilde{x}}$	~	-	~	\sim	\mathbf{i}	~	~	$\overline{}$	~	~	~	~	$\overline{}$	1	~	\sim	1
			-	4	↓	λ.	\mathbf{x}	$\mathbf{\tilde{x}}$	\mathbf{x}	~	~		$\overline{}$	~	~	~	~	~	~	~	~	1	~	~	
0.7			~	*	\downarrow	λ.	~	$\mathbf{\tilde{x}}$	~	-	~	~	~	~	~	~	~	~	~	~	~	~	~	~	
			~	4	Ŷ	λ.	~	\mathbf{x}	~	-	~	~	~	~	~	~	~	~	~	~	~	~	~	~	
0.6			~	4	↓	ν.	~	\mathbf{x}	~	-	>	~	~	~	~	~	~	~	~	~	~	~	~	~	
			~	4	1	2	~	~	~	-	~	~	~	-	~	~	-	~	~	~	~	~	-	~	+
			~	1	j.	2	1	~		-	~	~	~	~	~	~	~	~	~	~	~	~	~	~	
- 0.5			~	4	.1.	х. Х.	~	2	<u> </u>	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	
			2	2		1	<u> </u>	<u>_</u>	1	2	_	~	~	~	~	~	2	~	~	~	~	~	~		1
			2	5	Ť			2	2	2	1		2	_	2	_	2	2	2	2	_	_	_	_	
Π4	L		<u>_</u>	5	¥.			0				<u>_</u>	<u> </u>					_	_	_	_	_	_	<u>.</u>	
0.4			<u></u>	×.	*			<u>`</u>	<u>`</u>	<u></u>	<u>`</u>		<u>_</u>	<u> </u>			2	_	_	_	_	_	_		
			<u> </u>	×.	4	*					~			. `	_						. ``			. *	
0.3	-		-	*	*	*		×			~	~	2	-	-	-	-	-	-	~	~	-	-	<u> </u>	1
			-	*	↓	*	*	*		24	~	~	1					-	-	~	~	-	~	~	
0.2 0.1			-	*	↓	7	~	*	~	-	~	~	~	-	-	~	~	-	-	~	~	-	-	~	
			-	*	↓	74	~	7	~	-	1	~	-	-	-	-	-	-	-	~	~	-	-	-	1
			-	*	↓	4	*	7	1	-	1	~	>	~	~	~	~	1	~	1	~	1	1	~	
			-	*	↓	4	$\mathbf{\tilde{x}}$	7	>	-	1	~	>	1	~	~	~	~	~	~	1	1	1	~ .	
			-	*	\downarrow	λ.	\mathbf{x}	$\mathbf{\tilde{x}}$	~	-	~	~	$\overline{}$	1	~	~	~	~	~	~	~	1	1	~	
			-	4	Ŷ	х,	×.,	$\mathbf{\tilde{x}}$	~	-	~	~	~	~	\sim	~	~	~	~	~	$\overline{}$	~	~	~	
0			1.05		0.1		0.1	E		0.0					0.2		0.24	-		04		0.44	-		1
U			0.00		0.1		0.1	9		U.Z		0.2	5		0.0		0.3			0.4		0.45		U	0

Figuur 1.10: Het richtingsveld in het fasevlak van Vergelijking 1.17 met $a = \frac{1}{10}$ en b = 1. Wanneer $V(0) < \frac{a}{b} = \frac{1}{10}$ zal zich geen epidemie ontwikkelen.

zodat het proces kan beschreven worden aan de hand van een 2D stelsel:

$$\begin{cases} \frac{dV}{dt} = -aV(N - V - R)\\ \frac{dR}{dt} = b(N - V - R) \end{cases}$$
(1.19)

waaruit volgt:

$$V = V(0) e^{\frac{a}{b}R}.$$
 (1.20)

Bovenstaande modellen verwaarlozen ruimtelijke karakteristieken. Immers, ze houden geen rekening met individueel gedrag van de mens en veronderstellen een perfecte menging van de populatie [1]. Om deze tekortkomingen in te vullen, worden discrete spatio-temporele modellen gebruikt [1, 7].

1.2.3 Discrete modellen

IBM Het individuele gedrag kan gemodelleerd worden met een individu-gebaseerd model (cfr. Tab. 1.1), waarbij zich slechts één individu in één cel van de tesselatie kan bevinden. Elk individu heeft een epidemiologische status, namelijk vatbaar (V), geïnfecteerd (I) of hersteld (R). Volgende regels bepalen het gedrag van de individuen doorheen de tijd [7]:

- (i) Wanneer een vatbaar individu zich in de omgeving van een geïnfecteerd individu bevindt, zal deze geïnfecteerd worden met een kans p_v .
- (ii) Een geïnfecteerd individu herstelt met een kans p_r .
- (iii) Herstelde individuen worden terug vatbaar met een kans p_s .

CML Daarnaast kan tevens gekeken worden naar het gedrag van een groep in plaats van het individuele gedrag [1]. Hierbij wordt verondersteld dat verschillende individuen in één cel $(c_{i,j})$ leven en de geografische ruimte bedekt wordt door een regelmatige, vierkantige tesselatie. Net als in het model van Kermack en McKendrick [88] wordt de populatie in elke cel $(c_{i,j})$ onderverdeeld in drie groepen, namelijk de fractie vatbare $(V(c_{i,j},t))$, geïnfecteerde $(I(c_{i,j},t))$ en herstelde $(R(c_{i,j},t))$ individuen. Het aantal individuen in elke cel $(N(c_{i,j},t))$ kan verschillend zijn, maar de som van de fracties vatbaren, geïnfecteerden en herstelden is steeds gelijk aan één. Vermits met fracties wordt gewerkt is de toestand van de cel continu, dus is dit een coupled map lattice (CML) (cfr. Tab. 1.1). Er wordt verondersteld dat de epidemie niet dodelijk is, er wordt niemand geboren en er vindt geen migratie plaats. Herstelde individuen zijn voor altijd immuun en er vindt transport tussen de cellen plaats, maar individuen komen binnen éénzelfde tijdstap steeds terug naar hun eigen cel. Geïnfecteerden kunnen dus anderen infecteren via contact en het aantal individuen dat geïnfecteerd wordt is onafhankelijk van de totale populatie in de cel. De transitiefuncties die gebruikt worden, zijn [1]:

$$\begin{cases} V(c_{i,j},t+1) = V(c_{i,j},t) - b V(c_{i,j},t) I(c_{i,j},t) - \\ V(c_{i,j},t) \sum_{(c_{k,l} \in B^*)} \frac{N(c_{k,l},t)}{N(c_{i,j},t)} \eta_{(k,l)}^{c(i,j)} I(c_{k,l},t) \\ I(c_{i,j},t+1) = (1-\rho) I(c_{i,j},t) + b V(c_{i,j},t) I(c_{i,j},t) + \\ V(c_{i,j},t) \sum_{(c_{k,l} \in B^*)} \frac{N(c_{k,l},t)}{N(c_{i,j},t)} \eta_{(k,l)}^{(i,j)} I(c_{k,l},t) \\ R(c_{i,j},t+1) = R(c_{i,j},t) + \rho I(c_{i,j},t) \end{cases}$$
(1.21)

waarbij B^* de omgeving van $c_{i,j}$ zonder $c_{i,j}$ voorstelt, de parameter $\eta_{(k,l)}^{(i,j)}$ het product is van drie factoren, namelijk $a, q_{(k,l)}^{(i,j)}$ en $o_{(k,l)}^{(i,j)}$, a is de virulentie van de epidemie en ρ is de fractie geïnfecteerden die herstellen in één tijdstap, welke epidemieafhankelijk zijn en dezelfde voor alle $c_{i,j}$. De bewegingsfactor $o_{(k,l)}^{(i,j)}$ geeft de kans dat een geïnfecteerd individu zich van $c_{k,l}$ naar $c_{i,j}$ verplaatst en $q_{(k,l)}^{(i,j)}$ is de verbindingsfactor, die afhangt van het aantal transportmogelijkheden tussen $c_{i,j}$ en $c_{k,l}$ [1]:

 $q_{(k,l)}^{(i,j)} = \begin{cases} 1, & \text{wanneer er drie transportmogelijkheden zijn,} \\ 0.6, & \text{wanneer er twee transportmogelijkheden zijn,} \\ 0.3, & \text{wanneer er slechts één transportmogelijkheid is,} \\ 0, & \text{wanneer er geen transportmogelijkheden zijn.} \end{cases}$ (1.22)

De epidemie zal pas verspreiden wanneer in de cel waar de eerste geïnfecteerden zich bevinden, de fractie vatbaren kleiner is dan $\frac{\rho}{a}$. De verspreiding van de epidemie zal in dit model sneller gebeuren in gebieden met (meerdere) transportmogelijkheden tussen naburige cellen. Fig. 1.11 toont het aantal geïnfecteerden per cel, met een grijsschaal van wit ($I(c_{i,j},t) = 0$) naar zwart ($I(c_{i,j},t) = 1$), op verschillende tijdstappen bij een simulatie met constante verbindingsfactoren en een homogene populatieverdeling en een ziekteuitbraak in de centrale cel. Hierbij is te merken dat de verspreiding van de epidemie in alle richtingen gelijkmatig gebeurt en er dus een cirkelvormig front is. Wanneer de verbindingsfactoren niet constant worden verondersteld en de populatie niet homogeen verdeeld
wordt, zoals dit in werkelijkheid ook is, wordt Fig. 1.12 bekomen. De verspreiding gebeurt in de verschillende richtingen op een verschillend tempo [1].



Figuur 1.11: Gesimuleerd aantal geïnfecteerden met het model vooropgesteld in [1], waarbij de populatie homogeen verdeeld is en de verbindingsfactoren constant worden verondersteld



Figuur 1.12: Gesimuleerd aantal geïnfecteerden met het model vooropgesteld in [1], waarbij de populatie heterogeen verdeeld is en de verbindingsfactoren niet-constant worden verondersteld

1.2.4 CA-gebaseerd SIR

Het CA-gebaseerd SIR model dat in deze scriptie zal gebruikt worden, werd opgesteld door Baetens en De Baets [11]. Hierbij wordt gestart vanuit een continu model ter beschrijving van spatiotemporele biologische processen om een CML af te leiden dat ziekteverspreiding beschrijft (CASIR). Het continue model wordt gegeven door:

$$\begin{cases} \frac{\partial X(x,y,t)}{\partial t} = f\left(X(x,y,t)\right) - X(x,y,t) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G\left((x,y),(x^*,y^*)\right) Y(x^*,y^*,t) \, \mathrm{d}x^* \mathrm{d}y^* \\ \frac{\partial Y(x,y,t)}{\partial t} = g\left(Y(x,y,t)\right) + X(x,y,t) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G\left((x,y),(x^*,y^*)\right) Y(x^*,y^*,t) \, \mathrm{d}x^* \mathrm{d}y^* \end{cases}$$
(1.23)

waarbij f en g respectievelijk de groeifuncties van X en Y voorstellen. G beschrijft de invloed die terreinkarakteristieken uitoefenen op de mate waarin besmette individuen in (x^*, y^*) nog vatbare individuen in (x, y) beïnvloeden. Discretisatie van dit model, gebruikmakend van een tesselatie met vierkante cellen zoals gegeven in Fig. 1.4, leidt tot:

$$\begin{cases} X(c_{i,j},t+1) = X(c_{i,j},t) + \tilde{f}(X(c_{i,j},t)) - X(c_{i,j},t) \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \tilde{G}((i,j),(i^*,j^*)) Y(c_{i^*,j^*},t) \\ Y(c_{i,j},t+1) = Y(c_{i,j},t) + \tilde{g}(Y(c_{i,j},t)) + X(c_{i,j},t) \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \tilde{G}((i,j),(i^*,j^*)) Y(c_{i^*,j^*},t) \end{cases}$$

$$(1.24)$$

Indien de omgeving van $c_{i,j}$ beperkt wordt tot een eindig aantal cellen, herleidt Vgl. (1.24) zich tot:

$$\begin{cases} X(c_{i,j},t+1) = X(c_{i,j},t) + \tilde{f}(X(c_{i,j},t)) - \\ X(c_{i,j},t+1) \sum_{-\sigma}^{\sigma} \sum_{-\sigma}^{\sigma} \tilde{G}((i,j),(i^*,j^*))Y(c_{i^*,j^*},t) \\ Y(c_{i,j},t+1) = Y(c_{i,j},t) + \tilde{g}(Y(c_{i,j},t)) + \\ X(c_{i,j},t+1) \sum_{-\sigma}^{\sigma} \sum_{-\sigma}^{\sigma} \tilde{G}((i,j),(i^*,j^*))Y(c_{i^*,j^*},t) \end{cases}$$
(1.25)

waarbij σ de straal van de omgeving voorstelt.

Wanneer een tesselatie \mathcal{T} opgebouwd uit onregelmatige cellen beschouwd wordt, kan een graaf G geconstrueerd worden waarin het knooppunt i correspondeert met cel c_i en waarin er een boog zal bestaan tussen twee knopen i, j indien c_i en c_j tenminste één punt gemeenschappelijk hebben (cfr. Fig. 1.2). Aldus wordt Vergelijking (1.25):

$$\begin{cases} X(c_i, t+1) = X(c_i, t) + \tilde{\tilde{f}}(X(c_i, t)) - X(c_i, t) \sum_{c_j \in B(c_i)} \tilde{\tilde{G}}(c_i, c_j) Y(c_j, t) \\ Y(c_i, t+1) = Y(c_i, t) + \tilde{\tilde{g}}(X(c_i, t)) + X(c_i, t) \sum_{c_j \in B(c_i)} \tilde{\tilde{G}}(c_i, c_j) Y(c_j, t) \end{cases}$$
(1.26)

Dit model is onafhankelijk van het onderliggend grid wanneer de tesselatie oneindig groot is.

Dit algemeen model kan eenvoudig worden omgevormd voor de spatio-temporele beschrijving van ziekteverspreiding. In elke c_i wordt het relatief aantal vatbare $(V(c_i, t))$, geïnfecteerde $(I(c_i, t))$ en herstelde $(R(c_i, t))$ individuen getraceerd doorheen de tijd. Verder wordt aangenomen dat het aantal inwoners per cel (N) constant en identiek is, zodat $N = I(c_i, t) + S(c_i, t) + R(c_i, t), \forall t, i$. Het aantal herstelde individuen op tijdstip t + 1 wordt bekomen door het aantal herstelde individuen op tijdstip t. Zodoende dient bij het aantal geïnfecteerde individuen op tijdstip t + 1 de fractie aan herstelde individuen op tijdstip t te worden afgetrokken. Rekening houdend met deze assumpties wordt het CASIR model

gegeven door:

$$\begin{cases} V(c_i, t+1) = V(c_i, t) - V(c_i, t) \sum_{c_j \in B(c_i)} G(d_{ij}, \mathbf{I}_i, \mathbf{w}_{ij}) \\ I(c_i, t+1) = I(c_i, t) + V(c_i, t) \sum_{c_j \in B(c_i)} G(d_{ij}, \mathbf{I}_i, \mathbf{w}_{ij}) - \rho I(c_i, t) \\ R(c_i, t+1) = R(c_i, t) + \rho I(c_i, t) \end{cases}$$
(1.27)

waarbij \mathbf{w}_{ij} een vector is die bijkomende informatie bevat over de contactkarakteristieken tussen c_i en c_j , \mathbf{I}_i een vector is die de ruimtelijke informatie bevat over c_i en d_{ij} geeft de afstand tussen de knopen i en j die worden gemeten langsheen de graaf G. We beschouwen de omgeving van een polygoon c_i :

$$B(c_i) = \{c_j \in \mathcal{T} \mid d_{ij} \le \sigma\},\tag{1.28}$$

die een uitgebreide Moore-omgeving voorstelt. Door dit model te koppelen aan een GIS dat informatie bevat over de landbedekking kan een onderscheid worden gemaakt tussen de virulentie in regio's met een verschillende landbedekking, zoals stedelijk en bosrijk gebied. Immers, het is te verwachten dat humane ziekten sneller zullen verspreiden naarmate de bevolkingsdichtheid toeneemt. De vector \mathbf{w}_{ij} bevat wegingsfactoren w_{ij} en \mathbf{I}_i omvat de informatie over de landbedekking. In het kader van deze scriptie wordt $G(d_{ij}, \mathbf{I}_i, \mathbf{w}_{ij}) = w_{ij} H(d_{ij}, \mathbf{I}_i)$ beschouwd en wordt $\sigma = 1$ gekozen in Vgl. (1.25). De functie H omvat de virulenties $v_{p,q}$ waarbij p duidt op de afstand d_{ij} en q op de landbedekkingsklasse van c_i . De intracelvirulenties, de snelheid waarmee vatbare individuen in een cel c_i geïnfecteerd raken door andere geïnfecteerde individuen binnen dezelfde c_i worden voor de verschillende landbedekkingsklasses dus bewaard in \mathbf{v}_0 . De intercelvirulenties, de snelheid waarmee vatbaren in een naburige cel een infectie oplopen door geïnfecteerden in c_i , wordt dan bewaard in de vector \mathbf{v}_1 .

We krijgen:

$$H = \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} v_{0,1} & v_{1,1} \\ v_{0,2} & v_{1,2} \\ \vdots & \vdots \\ v_{0,n_k} & v_{1,n_k} \end{bmatrix}$$
(1.29)

1.3 Parametrisatie van cellulaire automaten

1.3.1 Gevoeligheidsanalyse

Gevoeligheids- of sensitiviteitsanalyse is de studie over hoe de onzekerheid in de output van een model kan toegewezen worden aan verschillende bronnen van onzekerheid in de modelinput [89]. Bij een kleine verandering van één of meerdere modelparameters wordt de grootte van de veranderingen in de output van het model gemeten. Gevoeligheidsanalyse is een noodzakelijke stap bij de bouw van een model omdat hiermee de betrouwbaarheid van het model en de onzekerheden van de modelresulaten weergegeven worden. Verder kan een gevoeligheidsanalyse een idee geven over eventuele correlaties tussen parameters. Met behulp van een univariate gevoeligheidsanalyse kunnen de parameters die het meeste invloed hebben op de modeloutput bepaald worden [90].

Lokale sensitiviteitsanalyse wordt uitgevoerd rond een punt in de parameterruimte, bij een globale sensitiviteitsanalyse daarentegen wordt de sensitiviteit geëvalueerd over de volledige parameterruimte en is bijgevolg zeer rekenintensief. De absolute sensitiviteit van de modeloutput y(t) ten opzichte van een parameter θ wordt voorgesteld als volgt:

$$\frac{\partial y(t)}{\partial \theta}.$$
(1.30)

Veelal moet deze partiële afgeleide numeriek bepaald worden en krijgt men voor het voorwaartse verschil (met $\Delta \theta_j > 0$):

$$\left. \frac{\partial y(t)}{\partial \theta_j} \right|_+ = \lim_{\Delta \theta_j \to 0} \frac{y(t, \theta_j + \Delta \theta_j) - y(t, \theta_j)}{\Delta \theta_j}.$$
(1.31)

Analoog kan het achterwaartse en het centrale verschil gevonden worden uit:

$$\frac{\partial y(t)}{\partial \theta_j}\Big|_{-} = \lim_{\Delta \theta_j \to 0} \frac{y(t,\theta_j) - y(t,\theta_j - \Delta \theta_j)}{\Delta \theta_j},$$
(1.32)

$$\frac{\partial y(t)}{\partial \theta_j} = \lim_{\Delta \theta_j \to 0} \frac{y(t, \theta_j + \Delta \theta_j) - y(t, \theta_j - \Delta \theta_j)}{2 \,\Delta \theta_j}.$$
(1.33)

De perturbatie $\Delta \theta_j$ wordt gedefinieerd als het product van θ_j en de perturbatiefactor ε , d. i. $\Delta \theta_j = \theta_j \varepsilon$. Het voorwaartse en achterwaartse verschil moeten aanleiding geven tot eenzelfde gevoeligheidsfunctie, zodat ε bepaald kan worden door de fout tussen $\frac{\partial y(t)}{\partial \theta_j}\Big|_+$ en $\frac{\partial y(t)}{\partial \theta_j}\Big|_-$ te minimaliseren. De relatieve sensitiviteit ten opzichte van θ_j wordt gebruikt om de sensitiviteit van eenzelfde variabele ten opzichte van verschillende parameters te vergelijken. Om de relatieve sensitiviteit te berekenen wordt de absolute sensitiviteit vermenigvuldigd met θ_j [89]:

$$\frac{\partial y(t)}{\partial \theta_j} \theta_j. \tag{1.34}$$

1.3.2 Optimalisatietechnieken

Het zoeken van de optimale waarden van modelparameters gebeurt door een doelfunctie Λ op te stellen die de afwijking tussen de geobserveerde en de gesimuleerde gegevens beschrijft en ze vervolgens te minimaliseren. Het optimum van de doelfunctie kan opgespoord worden met verschillende optimalisatietechnieken.

Hill-climbing Hill-climbing is een lokaal zoekalgoritme wat met zich meebrengt dat de kans groot is dat enkel een lokaal optimum (lokaal) gevonden wordt en niet het globale optimum. Algoritme 1 geeft de pseudocode weer voor deze techniek. Vooreerst wordt er in de zoekruimte Ω een willekeurig punt v_c gekozen, voor hetwelke Λ geëvalueerd wordt. Waarna er in de omgeving Ω^* van v_c een aantal punten gekozen worden, die op hun beurt worden geëvalueerd aan de hand van doelfunctie Λ . Van deze punten wordt het beste punt v_n gekozen, waarna $\Lambda(v_n)$ wordt vergeleken met $\Lambda(v_c)$. Indien $\Lambda(v_n)$ een betere waarde geeft dan $\Lambda(v_c)$, zal een nieuwe iteratie starten waarbij $v_c = v_n$. Om in een globaal optimum terecht te komen moet dit algoritme verscheidene keren uitgevoerd worden met een ander beginpunt in Ω [91].

Algoritme 1: Pseudocode Hill-climbing [91]
$t \leftarrow 0;$
initialiseer beste;
repeat
$lokaal \leftarrow VALS;$
kies willekeurig punt $v_c \in \Omega$;
evalueer $\Lambda(v_c)$;
repeat
selecteer een verzameling nieuwe punten in de omgeving Ω^* van v_c ;
selecteer het punt v_n uit de verzameling van nieuwe punten dat de beste waarde heeft
voor doelfunctie Λ ;
if $\Lambda(v_n)$ is beter dan $\Lambda(v_c)$ then
$v_c \leftarrow v_n;$
else
$lokaal \leftarrow WAAR;$
until lokaal;
$t \leftarrow t+1;$
if v_c is beter dan beste then $_$ beste $\leftarrow v_c$
until $t=MAX$;

Simulated Annealing Simulated Annealing is een hill-climbing techniek waarvan de pseudocode staat beschreven in algoritme 2. De eerste stappen zijn gelijkaardig aan deze van Algoritme 1, maar er wordt telkens slechts één punt geëvalueerd in de omgeving van v_c . De probabiliteit dat v_n gebruikt wordt als v_c in de volgende tijdstap:

$$p = \frac{1}{1 + e^{\frac{\Lambda(v_c) - \Lambda(v_n)}{T}}},$$

die dicht bij nul zal liggen als $\Lambda(v_n)$ veel minder goed is dan $\Lambda(v_c)$. Wanneer $\Lambda(v_n)$ daarentegen veel beter is dan $\Lambda(v_c)$ ligt deze probabiliteit dicht bij één. Dit zoekalgoritme is gebaseerd op een analoog thermodynamisch proces, waarbij om kristallen te laten groeien metalen worden gesmolten

en vervolgens terug afgekoeld. Deze afkoeling mag niet te snel gebeuren zodat er perfecte kristallen gevormd worden [91].

Algoritme 2: Pseudocode Simulated Annealing [91]

```
t \leftarrow 0;

initialiseer T;

kies willekeurig punt v_c \in \Omega;

evalueer \Lambda(v_c);

repeat

\begin{vmatrix} \mathbf{repeat} \\ \mathbf{selecteer een nieuw punt } v_n \text{ in de omgeving } \Omega^* \text{ van } v_c;

if \Lambda(v_c) < \Lambda(v_n) then

+ v_c \leftarrow v_n

else if random(0, 1) < e^{\frac{\Lambda(v_c) - \Lambda(v_n)}{T}} then

\perp v_c \leftarrow v_n

until (stopcriterium);

T \leftarrow g(T, t);

t \leftarrow t + 1;

until (eindconditie);
```

Nelder-Mead Nelder en Mead [92] beschrijven een simplex methode om een functie te minimaliseren. Ze baseerden zich op een optimalisatiemethode van Spendley *et al.* [93], die simplexen in Ω aanpast door steeds één punt te vervangen en zo tot een optimum te komen. Het vervangen van dit punt gebeurt met gekende stappen, waardoor deze methode niet algemeen toepasbaar is. De Nelder-Mead methode, gegeven in Algoritme 3 gebruikt stappen die zichzelf steeds aanpassen. Hierin is α de reflectie-, β de contractie- en γ de expansiecoëfficiënt en zijn P_i de (n + 1) punten in de zoekruimte Ω .

Differential Evolution Differential Evolution is een stochastische optimalisatietechniek waarbij een populatie wordt gebruikt die bestaat uit een vast aantal willekeurige punten P_i in Ω . Tijdens elke iteratie wordt voor alle punten binnen de populatie een nieuw punt P_i^* gecreëerd, maar het vorige punt wordt enkel vervangen door het nieuwe als de evaluatie van de doelfunctie Λ hiervoor kleiner is dan deze bekomen met het vorige punt. Het nieuw punt wordt gecreëerd door eerst een voorlopig punt (P_s) te berekenen met drie willekeurige punten $(P_w, P_u \text{ en } P_v)$ uit de populatie:

$$P_s = P_w + \kappa \left(P_u - P_v \right),$$

met κ de schalingsfactor. Het nieuwe punt wordt dan samengesteld op basis van de coördinaten van het voorlopige en het vorige punt, zodanig dat elke coördinaat in het nieuwe punt met een kans χ een coördinaat van het voorlopige punt is en de andere coördinaten zijn die van het vorige punt [94].

Algoritme 3: Pseudocode Nelder-Mead [92]

initialiseer P_i ; repeat Bepaal P_h waarvoor geldt $\Lambda(P_h) = \max(\Lambda(P_i));$ Bepaal P_l waarvoor geldt $\Lambda(P_l) = \min(\Lambda(P_i))$; $\bar{P} \leftarrow$ centrum van de punten P_i zonder P_h ; $P^* \leftarrow (1+\alpha) \bar{P} - \alpha P_h;$ if $\Lambda(P^*) < \Lambda(P_l)$ then $P^{**} \leftarrow (1+\gamma) P^* - \gamma \bar{P};$ $\begin{array}{l} \mbox{if } \Lambda(P^{**}) < \Lambda(P_l) \mbox{ then } \\ \mid \ P_h \leftarrow P^{**} \end{array}$ else $P_h \leftarrow P^*$ else if $\Lambda(P^*) > \Lambda(P_i), i \neq h$ then if $\Lambda(P^*) > \Lambda(P_h)$ then $P^{**} \leftarrow \beta P_h + (1 - \beta) \bar{P};$ if $\Lambda(P^{**}) > \Lambda(P_h)$ then $P_i \leftarrow \frac{(P_i + P_l)}{2}$ else else $P_h \leftarrow P^*;$ $P^{**} \leftarrow \beta P_h + (1 - \beta) \bar{P};$ if $\Lambda(P^{**}) > \Lambda(P_h)$ then $| P_i \leftarrow \frac{(P_i + P_l)}{2}$ else $P_h \leftarrow P^{**}$ else $\ \ P_h \leftarrow P^*$ until (eindconditie);

Random Search Bij Random Search worden een groot aantal willekeurige punten in Ω gekozen en vanuit deze punten wordt een lokaal minimum gezocht met lokale optimalisatietechnieken. Het beste lokale minimum wordt dan het globale minimum [95].

2

Invloed van de topologie op de stabiliteit van CA

De invloed van een verstoring van de initiële conditie op de stabiliteit van een bepaalde totalistische CA-familie werd reeds onderzocht (Sectie 2.2). In dit hoofdstuk zal de invloed van de topologie op de stabiliteit van de regels binnen dezelfde familie bekeken worden. Sectie 2.3 beschrijft de verschillende categorieën waarin de regels kunnen worden opgedeeld en in Sectie 2.4 wordt van enkele regels het bifurcatiepunt berekend.

2.1 Soft- en hardware

De resultaten voorgesteld in dit en de volgende hoofdstukken zijn bekomen met een Dell Optiplex 745 met 2 Gb RAM-geheugen en een rekencluster (Maggie) met 14 nodes en 76 processoren. De software die zowel op de computer als op de cluster gebruikt werd, was Mathematica 7.0 (Wolfram Research Inc., Champaign, VSA). De rekencluster maakt gebruik van GNU/Linux SuSE en als gridmanager Sun Grid Engine 6.2.

2.2 Maximale Lyapunov exponent van totalistische onregelmatige CA

De regels van totalistische onregelmatige CA kunnen als volgt genummerd worden [40]:

$$R_{\tau}^{T} = z_{\tau} k^{\tau} + z_{\tau-1} k^{\tau-1} + \ldots + z_{2} k^{2} + z_{1} k^{1} + z_{0}, \qquad (2.1)$$

met τ de maximum som, wanneer $\sum_{c_j \in B(c_i)} s(c_j, t) \ge \tau$ is $f = \tau$, k het aantal mogelijke toestanden en z_f de toestand die wordt toegekend aan c_i indien $f = \sum_{c_j \in B(c_i)} s(c_j, t)$. Voor totalistische CA die gebaseerd zijn op een onregelmatige tesselatie, een maximumsom van zeven en k = 2 ((2,7) totalistische onregelmatige CA) wordt dit:

$$R_7^T = z_7 \ 2^7 + z_6 \ 2^6 + z_5 \ 2^5 + z_4 \ 2^4 + z_3 \ 2^3 + z_2 \ 2^2 + z_1 \ 2 + z_0.$$
(2.2)

Een onderzoek naar de maximale Lyapunov exponenten (MLE) van (2,7) totalistische onregelmatige CA werd reeds uitgevoerd [40]. De procedure die hierbij gevolgd werd staat beschreven in Sectie 1.1.3. Baetens en De Baets [40] berekenen de MLE van alle 256 mogelijke regels binnen deze CA-familie bij acht verschillende initieel geperturbeerde cellen, d. i. over een ensemble van initiële perturbaties $E = \{es_0^* \mid e = 1, 2..., 8\}$. De tesselatie werd bekomen met een Voronoi tesselatie [28] uitgaande van 675 willekeurig gekozen punten (Fig. 2.1) en er werden periodieke randvoorwaarden opgelegd om randeffecten te vermijden. Fig. 2.2 toont het aantal buren en de proportie van de totale omtrek die de cellen delen met hun buren. Voor alle regels binnen de beschouwde CA-familie werd de MLE bepaald na 500 tijdsstappen, aangezien dan steeds convergentie van de berekende MLE bekomen werd.



Figuur 2.1: Tesselatie van onregelmatige cellen die gebruikt wordt in Hoofdstuk 2

Op basis van de MLE kunnen (2,7) totalistische onregelmatige CA ingedeeld worden in drie klassen. Klasse 1 omvat de regels binnen deze CA-familie die naar een superstabiel (homogeen) evenwicht evolueren en bijgevolg gekenmerkt worden door $\lambda = -\infty$. Deze klasse bevat 32 regels die in hoogstens 126 tijdstappen evolueren naar een homogene toestand. Klasse 2 omvat CA die slechts bij enkele initiële perturbaties van het ensemble E naar $\lambda = -\infty$ evolueren. Tab. 2.1 geeft de 57 regels die zich in deze klasse bevinden, waarbij wordt aangegeven voor hoeveel $_es_0^*$ in E ze evolueren naar een superstabiel evenwicht. Klasse 3, ten slotte, bevat 167 regels die bij geen enkele initiële conditie van het ensemble E naar een superstabiel evenwicht evolueren.



Figuur 2.2: Een histogram van (a) $|B(c_i)|$ en (b) het aandeel van c_i 's omtrek dat gevormd wordt door elk van zijn buren in $B(c_i)$

2.3 Invloed van de topologie

In het artikel van Baetens en De Baets [40] wordt enkel gekeken welke invloed de keuze van de initieel geperturbeerde cel heeft op de stabiliteit van de (2,7) totalistische onregelmatige CA. Er wordt hierbij steeds eenzelfde onregelmatige tesselatie gebruikt (Fig. 2.1). Om de invloed van de topologie, dit is het geheel van cellen en hun interconnecties, op de MLE te onderzoeken, werd in het kader van deze scriptie gebruikt gemaakt van dezelfde tesselatie, maar de omgeving van de cellen werd aangepast. Er werd een restrictieve von Neumann omgeving gebruikt (Fig. 1.6) wat ervoor zorgt dat de cellen minder buren hebben naarmate ν stijgt. Wanneer de restrictie ν toeneemt, zullen de cellen een steeds kleinere omgeving krijgen waardoor de gemiddelde connectiviteit van de tesselatie (\overline{C}) afneemt (Fig. 2.4(a)) aangezien

$$\overline{C} = \frac{1}{|\mathcal{T}^*|} \sum_{c_i} |B(c_i)|.$$
(2.3)

Aantal keer	Regelnummer
1	11, 14, 40, 49, 65, 78, 79, 113,
	142, 143, 159, 177, 190, 206, 207, 223, 234-239
2	95, 112, 120, 121, 122, 123, 130-134, 191
3	2, 4, 6, 30, 62, 72, 193
4	5, 124, 135
5	125, 200, 240, 241
6	136
7	7, 15, 31, 126, 127, 129, 242, 245

Tabel 2.1: Aantal keer dat de regels in Klasse 2 evolueren naar een superstabiel evenwicht ($\lambda = -\infty$) bij de verschillende initiële perturbaties ($E = \{e_s_0^* \mid e = 1, 2, 3, 4, 5, 6, 7, 8\}$) [40]

Wanneer $\nu > 0.22$ is de gemiddelde connectiviteit zo laag, dat er enkele regio's in de tesselatie niet meer met elkaar verbonden zijn. Dit is duidelijk te zien op de graafvoorstelling van de tesselatie bij $\nu = 0.23$ (Fig. 2.3(a)). De MLE wordt dan afhankelijk van de ligging van de geperturbeerde cel en zal dus niet langer representatief zijn. Om deze reden werden de berekeningen slechts tot $\nu = 0.22$ uitgevoerd waarvan de graafvoorstelling wordt weergegeven in Fig. 2.3(b). De gemiddelde connectiviteit van alle cellen die informatie kunnen ontvangen van de initieel geperturbeerde cel (C_i) kan ook berekend worden. Fig 2.4(b) toont duidelijk dat C_{156} verschillend wordt van \overline{C} als $\nu > 0.22$ wat aantoont dat slechts tot $\nu = 0.22$.



Figuur 2.3: De graafvoorstelling van de tesselatie gebruikt in Hoofdstuk 2 met een restrictieve von Neumann-omgeving waarbij (a) $\nu = 0.23$ en (b) $\nu = 0.22$

De MLE werd berekend voor alle 256 (2,7) totalistische onregelmatige CA met $\nu = 0.01$ tot $\nu = 0.22$ in stappen van 0.01, met dezelfde initiële geperturbeerde cel (nl. c_{156}) en voor 500 tijdstappen omdat in [40] reeds na 500 tijdstappen convergentie van de MLE bereikt wordt. De pseudocode die gebruikt



Figuur 2.4: (a) De gemiddelde connectiviteit (\overline{C}) van de cellen in de tesselatie afgebeeld in Fig. 2.1 en (b) de gemiddelde connectiviteit van de cellen die informatie kunnen ontvangen van de initieel geperturbeerde cel c_{156} (C_{156}) bij gebruik van een restrictieve von Neumann omgeving (cfr. Fig. 1.6)

werd om de MLE van een CA met k = 2 te berekenen wordt gegeven in Algoritme 4. Om de MLE te vergelijken overheen de simulaties met een verschillende topologie werden ze genormaliseerd met het logaritme van \overline{C} (cfr. Vgl. (2.3)). De aldus bekomen genormaliseerde MLE werden voor elke regel binnen deze familie uitgezet in functie van \overline{C} zodat de regels in categorieën kunnen opgedeeld worden naargelang de invloed die de topologie heeft op hun MLE (Fig. 2.5 en 2.6). De rekentijd voor het berekenen van de MLE hangt sterk af van de stabiliteit van de CA. Bij een (super)stabiele CA treden weinig verschillen tussen s en s^* op waardoor er weinig replica's moeten bijgehouden worden. Onstabiele CA daarentegen worden gekenmerkt door veel verschillen tussen s en s^* en vergen dus meer rekentijd.

Er zijn 24 (2,7) totalistische onregelmatige CA die superstabiel blijven, ongeacht de veranderingen van de topologie (Tab. 2.2). Zeventien van deze 24 regels bevinden zich in Klasse 1 (cfr. Sectie 2.2) en zijn dus ook superstabiel ongeacht de initieel geperturbeerde cel. Vijftien regels, die zich volgens Baetens en De Baets [40] in Klasse 1 bevinden, zijn toch niet superstabiel bij een andere topologie. Dit kan voortkomen uit het feit dat er bij die topologie nog geen convergentie bereikt wordt na 500 tijdstappen of door de gewijzigde invloed van de initieel geperturbeerde cel.

Algoritme 4: Pseudocode voor het berekenen van de MLE van CA met twee toestanden [40]

Initialiseer twee begincondities $s_0 = s(\cdot, 0)$ en $s_0^* = s^*(\cdot, 0)$ waarvoor geldt dat $\Sigma_{c_i} h(c_i, 0) = 1 \text{ met}$ $h(c_i, 0) = \begin{cases} 1 & \text{als } s^*(c_i, 0) \neq s(c_i, 0) \\ 0 & \text{anders.} \end{cases}$ Bewaar s^* in A_0 ; foreach tijdstap t do Bereken $s(\cdot, t+1)$; Bereken $r(\cdot, t+1)$ voor alle $r \in A_t$; Bereken $h(c_j, t+1)$ met $h(c_j, t+1) = \#\{r \in A_t \mid r(c_j, t+1) \neq s(c_j, t+1)\};$ for each c_j waarvoor geldt $h(c_j, t+1) = 1$ do Creëer een replica r^j met $r^j(c_j, t+1) = s^*(c_j, t+1) = \bar{s}(c_j, t+1)$ en $r^{j}(c_{k},t+1) = s(c_{k},t+1)$ voor elke $c_{k} \neq c_{j}$; Bewaar in A_t alle replica's; Bepaal de MLE met behulp van Vgl. (1.7);

Voor 68 regels binnen de beschouwde CA familie geldt dat ze bij veranderingen van de topologie steeds onstabiel blijven (Fig. 2.5(a)). Tweeënzestig van deze regels bevinden zich in Klasse 3, terwijl voor zes van deze 68 regels geldt dat ze wel superstabiel zijn indien een andere initieel geperturbeerde cel gekozen wordt. De overige 164 regels kunnen ingedeeld worden in drie categorieën. De eerste categorie bevat 93 regels die bij een hoge connectiviteit onstabiel zijn, maar bij een lage connectiviteit superstabiel worden (Fig. 2.5(b)). Aangezien bij een lage connectiviteit $|B(c_i)|$ lager is, wordt de invloed van een topologische verandering sterk verminderd. De invloed van een perturbatie zal dus kleiner zijn en meer regels zullen superstabiel worden. Een tweede categorie omvat vier regels die bij een hoge connectiviteit superstabiel zijn, maar onstabiel worden wanneer de connectiviteit daalt (Fig. 2.5(c)). De laatste categorie, ten slotte, bevat 67 regels die bij de ene topologie superstabiel zijn, maar bij een andere dan weer onstabiel (Fig. 2.5(d)). Deze laatste categorie omvat 35 regels die bij lage connectiviteit superstabiel worden, wat waarschijnlijk te wijten is aan de kleiner wordende invloed van de initieel geperturbeerde cel. Een verandering van de topologie kan dus het gedrag van een CA wijzigen. Echter, de meeste studies over het gedrag van CA gebruiken slechts één topologie om conclusies te nemen.

2.4 **Topologische bifurcaties**

Bifurcatie is het verschijnen van een discontinuïteit in het fasevlak bij het variëren van de parameters die in de vergelijking(en) van het dynamisch systeem aanwezig kunnen zijn [96]. Het bifurcatiepunt is de parameterwaarde waarbij de bifurcatie optreedt. Bij de regels die bij een hoge connectiviteit onstabiel zijn, maar bij lage connectiviteit superstabiel worden, zullen er één of meerdere bifurcatiepunten \overline{C}^{\dagger} te zien zijn in het (\overline{C}, λ) -vlak. Het punt waarop de MLE gelijk wordt aan $-\infty$ is duidelijk het bifurcatiepunt. Sommige regels vertonen daarnaast een bifurcatiepunt bij een bepaalde connectiviteit; zo heeft regel 180 een bifurcatiepunt tussen $\overline{C} = 4.6$ en $\overline{C} = 4.8$ (Fig. 2.6(a)).

 Tabel 2.2: Gedrag van de (2,7) totalistische onregelmatige CA bij een verandering van de topologie waarbij

 initieel dezelfde cel geperturbeerd was

Categorie	Regelnummer
Superstabiel,	0, 1, 7, 8, 15, 16, 31, 32, 63, 64, 80, 96, 126, 127, 128, 129, 144,
ongeacht topologie	160, 192, 208, 224, 240, 254, 255
Onstabiel,	2, 10, 13, 18, 21, 26, 27, 28, 34, 37, 42, 43, 45, 50, 53, 58, 59,
ongeacht topologie	66, 69, 74, 75, 77, 82, 85, 90, 91, 92, 98, 101, 106, 107, 109, 114,
	117, 130, 133, 138, 139, 141, 146, 149, 154, 155, 156, 162, 165,
	170, 171, 173, 178, 181, 186, 187, 194, 197, 202, 203, 205, 210,
	213, 218, 219, 220, 226, 229, 234, 235, 237
Onstabiel bij hoge \overline{C} ,	6, 9, 14, 17, 20, 22, 24, 25, 30, 33, 38, 39, 41, 46, 47, 49, 52, 54,
superstabiel bij lage \overline{C}	55, 57, 62, 65, 70, 71, 73, 78, 79, 81, 84, 86, 87, 88, 89, 94, 95,
	97, 102, 103, 105, 110, 111, 112, 113, 118, 119, 121, 134, 136,
	137, 142, 145, 148, 150, 151, 152, 153, 158, 159, 161, 166, 167,
	169, 174, 175, 177, 180, 182, 183, 184, 185, 190, 191, 193, 198,
	199, 201, 206, 207, 209, 212, 214, 215, 216, 217, 222, 223, 225,
	230, 231, 232, 233, 238, 239
Superstabiel bij hoge \overline{C} ,	250, 251, 252, 253
onstabiel bij lage \overline{C}	
Soms superstabiel,	3, 4, 5, 11, 12, 19, 23, 29, 35, 36, 40, 44, 48, 51, 56, 60, 61, 67,
soms onstabiel	68, 72, 76, 83, 93, 99, 100, 104, 108, 115, 116, 120, 122, 123,
	124, 125, 131, 132, 135, 140, 143, 147, 157, 163, 164, 168, 172,
	176, 179, 188, 189, 195, 196, 200, 204, 211, 221, 227, 228, 236,
	241, 242, 243, 244, 245, 246, 247, 248, 249



Figuur 2.5: De genormaliseerde MLE in functie van de gemiddelde connectiviteit (\overline{C}) die wordt gestuurd door een aanpassing van de restrictieve von Neumann omgeving voor regel (a) 98, (b) 105, (c) 250 en (d) 132

De 93 regels, die bij een hoge connectiviteit onstabiel zijn en bij een lage connectiviteit superstabiel worden, kunnen aan de hand van de waarde van \overline{C} waarop een transitie plaatsvindt van een positieve MLE naar $\lambda = -\infty$ ingedeeld worden in vijf categorieën (Tab. 2.4). Zes regels worden reeds superstabiel bij $\overline{C} = 6.8$, terwijl eveneens zes regels pas superstabiel worden bij $\overline{C} = 3.6$ (Fig. 2.6(a)). Er zijn 40 regels die bij $\overline{C} = 4.4$ superstabiel worden en deze vormen dan ook de omvangrijkste categorie (Fig. 2.5(b)). Acht regels hebben een bifurcatiepunt tussen $\overline{C} = 5.2$ en $\overline{C} = 5.5$ (Fig. 2.6(c)) en 33 regels worden superstabiel tussen $\overline{C} = 5.9$ en $\overline{C} = 6.3$ (Fig. 2.6(d)).

\overline{C}^{\dagger}	ν	Regelnummer
6.8	0.01	62, 65, 136, 190, 191, 193
6.3 - 5.9	0.04 - 0.06	14, 17, 30, 33, 46, 47, 49, 78, 79, 81, 94, 95, 97,
		110, 111, 112, 113, 142, 145, 158, 159, 161, 174,
		175, 177, 206, 207, 209, 222, 223, 225, 238, 239
5.5 - 5.2	0.08 - 0.10	24, 71, 88, 152, 184, 199, 216, 232
4.4	0.14	6, 9, 22, 25, 38, 39, 41, 54, 55, 57, 70, 73, 86, 87,
		89, 102, 103, 105, 118, 119, 121, 134, 137, 150,
		151, 153, 166, 167, 169, 182, 183, 185, 198, 201,
		214, 215, 217, 230, 231, 233
3.6	0.18	20, 52, 84, 148, 180, 212

Tabel 2.3: Het bifurcatiepunt \overline{C}^{\dagger} en de bijbehorende restrictie ν van (2,7) totalistische onregelmatige CA die bij een hoge \overline{C} onstabiel en bij een lage \overline{C} superstabiel zijn



Figuur 2.6: De genormaliseerde MLE in functie van de gemiddelde connectiviteit (\overline{C}) die wordt gestuurd door een aanpassing van de restrictieve von Neumann omgeving voor regel (a) 180, (b) 136, (c) 184 en (d) 174

2.5 Conclusie

In dit hoofdstuk werd de invloed van de topologie op de MLE onderzocht van alle (2,7) totalistische onregelmatige CA. De topologie werd aangepast door een restrictieve von Neumann omgeving te gebruiken. Er kan geconcludeerd worden dat enkele regels steeds superstabiel blijven en enkele regels steeds onstabiel blijven. Echter, het grootste deel van de regels zijn bij hoge connectiviteit onstabiel en worden bij lage connectiviteit superstabiel, dit was te verwachten aangezien de invloed van de initieel geperturbeerde cel kleiner wordt naarmate de connectiviteit verlaagt. Desalniettemin zijn er toch enkele regels die bij hoge connectiviteit superstabiel zijn, maar bij lage connectiviteit onstabiel worden. Dit komt waarschijnlijk doordat er nog geen convergentie bereikt is na 500 tijdstappen. Om een duidelijker beeld te krijgen over de invloed van de topologie op de MLE moet gekeken worden bij verschillende initieel geperturbeerde cellen, aangezien de lokalisatie van de initiële perturbatie ook een invloed heeft op de MLE.

Een kleine wijziging in de topologie kan een grote verandering in het gedrag van CA teweeg brengen. Het merendeel van de studies over het gedrag van CA gebruiken echter slechts één topologie. De conclusies van deze studies zijn hierdoor niet éénduidig. Topologische Lyapunov exponenten [97] kunnen tevens gebruikt worden om de gevoeligheid bij veranderingen van de topologie te onderzoeken, maar deze kwantificeren slechts de invloed op de stabiliteit bij een minimale topologische verstoring.

3

CA-gebaseerd SI-model voor de verspreiding van een hantavirus

In dit hoofdstuk wordt een bestaand model voor de verspreiding van een hantavirus bij knaagdieren uitgewerkt en kritisch geëvalueerd. Het model wordt beschreven in de eerste sectie. In de tweede sectie worden de tekortkomingen van het model besproken, hoe die werden opgelost en worden enkele simulaties met het verbeterde model geïllustreerd. De voorlaatste sectie omschrijft hoe de MLE van dit asynchroon model werd berekend.

3.1 Modelopbouw

Karim *et al.* [98] beschrijven een CA-gebaseerd model voor de verspreiding van een hantavirus onder knaagdieren. De mens kan via aanraking met een geïnfecteerd knaagdier of de uitwerpselen ervan een dodelijke hantavirusinfectie krijgen. Hantavirussen tasten de knaagdieren niet aan, er is geen verticale overdracht en geïnfecteerde knaagdieren kunnen de infectie niet verliezen. Aangezien er geen herstel kan optreden, wordt dit fenomeen gemodelleerd met een SI-model dat enkel de vatbare (V) en geïnfecteerde (I) individuen doorheen tijd en ruimte traceert.

Het volgende model is gebaseerd op het model van Karim *et al.* [98] en zal doorheen dit hoofdstuk als Model 1 aangeduid worden. Dit model steunt op een regelmatige tesselatie \mathcal{T} die bestaat uit vierkante cellen. Om randeffect en tot een minimum te beperken worden periodieke randvoorwaarden beschouwd, zodat er een torus wordt gevormd (Fig. 3.1). Verder is de omgeving een Moore omgeving (cfr. Fig. 1.5(b)).

- Een cel kan leeg zijn, d. i. $s(c_i, t) = 0$, of één knaagdier bevatten ($s(c_i, t) = 1$) waarvan de toestand beschreven wordt door middel van een sextet, $\langle x, y, z, j, K, i \rangle$, waarin:
 - x = 1 wanneer het knaagdier vatbaar is en x = 2 wanneer het knaagdier besmet is met een hantavirus;
 - -y = 1, 2, 3 of 4 indien het individu in de volgende tijdstap naar respectievelijk een lege cel boven, rechts van, onder of links van de huidige cel migreert;

- *i* de index van de cel is waarin het knaagdier zich bevindt;
- z de geboorteklok is, die aangeeft hoeveel tijdstappen het nog duurt voor een nieuw vatbaar individu zal geboren worden. In elke tijdstap vermindert deze met één, zodat z(i, t + 1) = z(i, t) 1;
- j de sterfteklok is, die bepaalt na hoeveel tijdstappen het individu zal sterven. Net als de geboorteklok vermindert deze in elke tijdstap met één, j(i, t + 1) = j(i, t) 1;
- K de milieuomstandigheden beschrijft.
- Wanneer z(i,t) = 0 en de cel in de vooropgestelde richting i* leeg is, zal het moederdier migreren naar de lege cel c_{i*} zodat x(i*, t + 1) = x(i,t), y(i*, t + 1) = y(i,t), z(i*, t + 1) ~ U[0, z^{max}] en j(i*, t + 1) = j(i,t) en er in cel i een nieuw individu geboren wordt waarvoor geldt dat x(i, t + 1) = 1, y(i, t + 1) = 1, z(i, t + 1) ~ U[0, z^{max}] en j(i, t + 1) ~ U[0, j^{max}] waarin z^{max} en j^{max} respectievelijk de maximale waarden voor de geboorte- en sterfteklok voorstellen.
- Vatbare en geïnfecteerde individuen sterven wanneer hun sterfteklok op nul staat of wiskundig: $s(c_i, t+1) = 0$ als j(i, t) = 0.
- Indien de cel naar dewelke een knaagdier zou migreren zoals aangegeven door *y*, leeg is, zal het knaagdier migreren naar deze cel.
- Een vatbaar individu met ten minste één geïnfecteerd knaagdier in zijn omgeving wordt geïnfecteerd met een kans $\frac{1}{8}$, dus dit is een stochastisch model.
- Een lege cel zonder knaagdieren in zijn omgeving zal leeg blijven, zodat $s(c_i, t+1) = 0$ als $s(c_i, t) = 0 \land \tilde{s}(B(c_i), t) = 0.$

Uit het onderzoek van Karim *et al.* [98] blijkt dat er steeds een asymptotisch stabiel evenwicht wordt bereikt dat afhankelijk is van K. Bij relatief slecht milieuomstandigheden (K = 10), zullen de geïnfecteerden, zelfs indien er initieel 70% van de knaagdieren geïnfecteerd zijn, na tien tijdstappen uitsterven (Fig. 3.2(a)). Wanneer K verhoogt zal de verspreiding van de infectie vergroten, aangezien de omstandigheden voor de verspreiding dan beter worden (Fig. 3.2(b)).

3.2 Modelimplementatie

3.2.1 Problemen en onvolledigheden

Ondanks publicatie in het internationale tijdschrift Chaos, Solitons and Fractals dat - gelet op zijn relatief hoge impactfactor - hoogaangeschreven is, zijn tijdens het implementeren van Model 1 vele inconsistenties en onvolledigheden in het artikel van Karim *et al.* [98] naar boven gekomen. Verschillende berichten naar de betrokken auteurs bleven onbeantwoord, zodat het model diende aangevuld en aangepast te worden voor het bekomen van simulatieresulaten. Het aangepaste model wordt Model 2 genoemd.



Figuur 3.1: Graafvoorstelling van een regelmatige tesselatie met vierkante cellen en periodieke randvoorwaarden

De milieuomstandigheden worden beschreven door de parameter K in het sextet van elk knaagdier, maar in het originele artikel wordt nergens vermeld hoe dit in het model geïmplementeerd moet worden. Naar analogie met de notaties in andere SI-modellen zou deze parameter kunnen verwijzen naar de draagkracht van het milieu die aangeeft hoeveel individuen er maximaal op een gegeven oppervlakte kunnen leven. Wanneer K toeneemt, zijn de milieuomstandigheden beter en kunnen er bijgevolg meer individuen overleven per oppervlakte-eenheid. Er werd geopteerd om dit als volgt in Model 2 te implementeren: indien $\sum_{c_i} s(c_i, t) > |\mathcal{T}| (1 - \frac{1}{K})$ werd een aantal willekeurig gekozen knaagdieren 'gedood' totdat $\sum_{c_i} s(c_i, t) = |\mathcal{T}| (1 - \frac{1}{K})$. Deze aanpak heeft tot gevolg dat K kan beschouwd worden als een globale modelparameter en niet langer fungeert als een individu-specifieke parameter. Dit laatste lijkt logischer aangezien een knaagdier kan migreren naar een omgeving met minder goede of betere omstandigheden.

Het model van Karim *et al.* [98] wordt voorgesteld als een synchroon model, waarbij tijdens elke tijdstap alle cellen worden geëvalueerd. Echter, bij de migratie kan dit enkele problemen opleveren zoals twee knaagdieren die migreren naar éénzelfde lege cel, waardoor knaagdieren 'verdwijnen' zonder te sterven. Door het model asynchroon te updaten kan dit probleem opgelost worden want hierbij wordt in elke tijdstap de toestand van slechts één willekeurig gekozen cel geüpdatet. Wanneer $t \rightarrow \infty$ zal elke cel ongeveer hetzelfde aantal keer geëvalueerd zijn. Om een equivalent aantal tijdstappen te bekomen als bij een synchroon model, moeten er in het asynchrone $|\mathcal{T}|$ keer zoveel tijdstappen beschouwd worden. Wanneer de willekeurig gekozen cel geen knaagdier bevat, zal er niets gebeuren en wordt overgegaan naar een volgende tijdstap.

Een bijkomende eigenaardigheid in Model 1 heeft betrekking op de omgeving. Immers, er wordt gewerkt met een Moore omgeving, maar om de migratie te modelleren wordt een von Neumann



Figuur 3.2: Grafiek van het gesimuleerd aantal vatbare en geïnfecteerde knaagdieren met een initiële conditie waarbij 70% geïnfecteerden en 10% vatbaren aanwezig zijn bij (a) K = 10 en (b) K = 30 [98]

omgeving gebruikt (cfr. Fig. 1.5(a)) zodat knaagdieren enkel naar de cellen links, rechts, boven of onder hen kunnen migreren. Om deze inconsistentie te verhelpen werd ervoor gekozen om in Model 2 een Moore omgeving te gebruiken om zowel de migratie als de infectie te simuleren. In dit geval kan y acht verschillende getallen aannemen. Fig. 3.3 toont de cellen waarheen een knaagdier kan migreren vanuit c_0 indien er zich in deze cellen geen knaagdieren bevinden.

De transitiefunctie die in Model 1 werd vooropgesteld en die aangeeft dat een vatbaar individu met ten minste één geïnfecteerd knaagdier in zijn omgeving zal geïnfecteerd worden met een kans $\frac{1}{8}$, lijkt niet realistisch aangezien de kans niet toeneemt als het aantal geïnfecteerden in de omgeving stijgt. Fig. 3.2(a) lijkt heel onlogisch, aangezien de geïnfecteerden geen vatbaren lijken te infecteren en er bijgevolg geen sprake is van de verspreiding van het hantavirus. Dit is te wijten aan de lage probabiliteit waarmee geïnfecteerde knaagdieren vatbare individuen besmetten, wat er toe leidt dat wanneer er slechts één geïnfecteerd individu in de omgeving van een vatbaar individu leeft, de kans even groot is dat het zal geïnfecteerd worden, dan wanneer er acht geïnfecteerde knaagdieren in zijn omgeving leven. Om deze reden werd de transitiefunctie zodanig aangepast dat de probabiliteit afhankelijk wordt van het aantal geïnfecteerde buren:



Figuur 3.3: De waarde van y(0,t) die de richting aangeeft waarheen een knaagdier kan migreren indien zich daar een lege cel bevindt

$$p_i = \frac{|\{k|x(i_k, t) = 2\}|}{|B(i)|},\tag{3.1}$$

met $B(i) = (i_k)_{k=1}^{|B(i)|}$ en p_i een maat voor de virulentie van het hantavirus.

Als laatste tekortkoming van Model 1 dient meegegeven te worden dat de maximale waarden voor de geboorte- en de sterfteklok, respectievelijk z^{max} en j^{max} , niet gekwantificeerd worden in het artikel. Om de simulaties uit te voeren werd z^{max} gelijkgesteld aan 15 en $j^{max} = 50$.

3.2.2 Simulatieresultaten

De simulaties uitgevoerd in [98] werden opnieuw uitgevoerd met Model 2. Hiervoor werd een tesselatie van 100×100 vierkante cellen gebruikt zodat de 50 synchrone tijdstappen die beschouwd worden in Model 1 overeen komen met 500000 asynchrone tijdstappen in Model 2.

In [98] wordt de invloed van de draagkracht K op de ontwikkeling van een hantavirusuitbraak onderzocht, maar over de implementatie van deze parameter werd geen informatie gegeven. Er werd voor geopteerd om indien $\sum_{c_i} s(c_i, t) > |\mathcal{T}| \left(1 - \frac{1}{K}\right)$ een aantal willekeurig gekozen knaagdieren te 'doden'. De simulaties met Model 2 werden met dezelfde initiële conditie uitgevoerd zoals deze in [98] vermeld wordt, namelijk 70% geïnfecteerden, 10% vatbaren en 20% lege cellen die allen willekeurig verdeeld zijn over \mathcal{T} . Het gesimuleerd aantal vatbare en geïnfecteerde individuen bekomen door Kgelijk te stellen aan 10, 20 en 30 worden weergegeven in Fig. 3.4. Vooral het aantal geïnfecteerden wordt beïnvloed door een verandering in K. Echter, op t = 500000 is het aantal geïnfecteerden in de drie simulaties ongeveer gelijk waaruit blijkt dat de draagkracht K slechts een zeer beperkte invloed heeft op het verloop van de infectie, hetgeen contrasteert met de beweringen van Karim *et al.* [98]. Zoals reeds aangestipt is de kans op overdracht van het virus tussen geïnfecteerde en vatbare individuen in Model 1 heel laag waardoor de draagkracht het meest invloedrijke proces is. In Model 2 wordt de virulentie belangrijker. De mogelijk andere implementatie van K kan tevens de verschillende simulatieresultaten verklaren.

Om een duidelijker beeld te krijgen van de dynamiek van het model, werden de initiële condities aangepast naar 20% geïnfecteerde, 70% vatbare en 10% lege cellen die willekeurig verdeeld zijn over \mathcal{T} . De draagkracht werd gelijkgesteld aan 40 en er werden 500000 asynchrone tijdstappen



Figuur 3.4: Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie van het aantal tijdstappen t voor K = 10, K = 20 en K = 30

beschouwd. Aangezien Model 2, net als Model 1 stochastisch van aard is, werden er tien simulaties met dezelfde parameterwaarden uitgevoerd. Fig. 3.5 toont de resultaten van deze simulaties en geeft duidelijk aan dat de dynamiek slechts weing verschilt tussen de simulaties. Na 500000 tijdstappen zijn er 4262.6 ± 59.5 geïnfecteerde knaagdieren en 660.6 ± 33.7 vatbare knaagdieren. Het aantal geïnfecteerden stijgt eerst sterk, maar vertoont vervolgens een lichte daling die veroorzaakt wordt doordat zowel vatbare als geïnfecteerde knaagdieren vatbare nakomelingen voortbrengen en er knaagdieren sterven.

Voor de tot dusver voorgestelde simulaties hebben discrete modellen echter geen belangrijke meerwaarde vermits deze dynamiek evenzeer kan gemodelleerd worden met (P)DV. Bij het invoeren van een heterogene verdeling van de virulentie heeft dit discreet model wel een belangrijke meerwaarde. Om het effect van een heterogeen verdeelde virulentie te onderzoeken werd in enkele cellen van de tesselatie de kans op overdracht van het hantavirus tussen een geïnfecteerd en een vatbaar individu gelijkgesteld aan nul (Fig. 3.6). De resultaten van 10 simulaties met als initiële conditie ongeveer 70% vatbare en 20% geïnfecteerde knaagdieren en K = 40 worden weergegeven in Fig. 3.7. De infectie zal zich, zoals te verwachten, niet zo sterk verspreiden hetgeen bevestigd wordt door de simulatieresultaten: op t = 500000 zijn er 2863.4 ± 41.8 geïnfecteerde knaagdieren en 2025.7 ± 60.6 vatbare. In vergelijking met het voorgaande model zijn er significant minder geïnfecteerde en meer vatbare knaagdieren, wat een gevolg is van de kleinere oppervlakte waar het hantavirus kan overgedragen worden. De standaardafwijking op de bekomen resultaten na 500000 asynchrone tijdstappen is voor de vatbare knaagdieren groter dan bij de gesimuleerde aantallen die werden bekomen bij een homogeen veronderstelde virulentie, terwijl voor geïnfecteerde knaagdieren het omgekeerde werd vastgesteld.



Figuur 3.5: Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie van het aantal tijdstappen t voor K = 40



Figuur 3.6: Tesselatie met in het donkergrijs de regio waar geen infectie kan plaatsvinden ($p_i = 0$) en in het lichtgrijs de regio waar p_i volgt uit Vgl. (3.1)

3.3 Maximale Lyapunov exponent

De MLE van een asynchroon geüpdatet model moet op een andere manier berekend worden dan de procedure die werd gevolgd in Hoofdstuk 2 doordat er per tijdstap slechts één cel geëvalueerd wordt, namelijk \tilde{c}_i . Dit algoritme vergt veel minder rekentijd aangezien enkel naar verschillen tussen $s(\tilde{c}_i, t)$ en $s^*(\tilde{c}_i, t)$ moet gezocht worden in de omgeving van de geëvalueerde cel. Er kunnen enkel verschillen optreden als deze zich in een omgeving bevindt waar in één van de vorige tijdstappen reeds een perturbatie aanwezig was.



Figuur 3.7: Het met Model 2 gesimuleerd aantal vatbare en geïnfecteerde individuen in functie van het aantal tijdstappen t voor K = 40 en een heterogene virulentie zoals weergegeven in Fig. 3.6

Ten gevolge van de stochasticiteit van Model 2 kunnen er ook verschillen tussen $s(\tilde{c}_i, t)$ en $s^*(\tilde{c}_i, t)$ ontstaan wanneer een cel geëvalueerd wordt in een omgeving zonder eerdere perturbatie. De initiële perturbatie gebeurt door de toestand van één knaagdier uit de tesselatie van vatbaar naar geïnfecteerd te veranderen of omgekeerd. Aangezien enkel het eerste element x in het sextet wordt aangepast, zullen de knaagdieren in elk van de replica's migreren naar dezelfde cellen, sterven op hetzelfde moment en reproduceren op hetzelfde moment. De MLE zal dus enkel de eventuele verschillen tussen het aantal geïnfecteerde en vatbare knaagdieren capteren.

Een probleem dat zich dan stelt bij de berekening van de MLE betreft de implementatie van de draagkracht K. Immers, deze werd geïmplementeerd door een aantal knaagdieren willekeurig te verwijderen wanneer teveel knaagdieren op de oppervlakte leven, maar door het willekeurig doden van knaagdieren kunnen er verschillen ontstaan tussen x(i, t) en $x^*(i, t)$ die niet terug te voeren zijn tot de initiële perturbatie. Om deze reden werd K gelijkgesteld aan $+\infty$.

Het berekenen van de MLE is voor een asynchroon, stochastisch model zeer reken- en tijdsintensief. Het systeem kan nooit superstabiel worden, dus treden er veel verschillen op die in elke tijdstap moeten bijgehouden worden. Om deze reden werd de MLE berekend voor t = 100000. De aldus bekomen MLE is 0.0069, hetgeen wil zeggen dat er geen superstabiel evenwicht bereikt wordt. Dit was te verwachten, doordat het model stochastisch is waardoor een knaagdier in het oorspronkelijke model geïnfecteerd kan raken, maar in één van de replica's niet. Hierdoor blijven er steeds verschillen tussen x en x^* optreden die niet te wijten zijn aan de initële perturbatie. Echter, omdat de MLE zo dicht bij nul ligt is het evenwicht evenwel ook niet onstabiel en wordt er gesproken van een stabiel evenwicht. Epidemiologische modellen zijn over het algemeen (asymptotisch) stabiele systemen, waardoor er na een voldoende lange periode altijd een evenwicht bereikt wordt.

3.4 Conclusie

De paper van Karim *et al.* [98] beschrijft een, op het eerste zicht, logisch model voor de verspreiding van een hantavirus onder knaagdieren. Bij de implementatie ervan kwamen evenwel talrijke onvolledigheden en inconsistenties naar boven, ondanks de relatief hoge impactfactor van het tijdschrift waarin de paper gepubliceerd werd. De oplossingen voor deze problemen en de aanpassingen die gemaakt werden, zorgden er dan weer voor dat andere simulatieresultaten bekomen werden waardoor het uiterst moelijk wordt om de door de auteurs getrokken conclusies te verifiëren.

Wanneer de initiële conditie wordt aangepast, krijgt men een beter beeld van de dynamiek van het nieuwe model. Hieruit blijkt dat het aantal geïnfecteerden eerst sterk stijgt om dan een evenwicht te bereiken waarbij het aantal vatbaren constant blijft. Het invoeren van een regio waar geen infectie mogelijk is, leidt tot resultaten die binnen de verwachtingen vallen betreffende het aantal geïnfecteerde knaagdieren.

Volgens de berekende MLE is het systeem stabiel hetgeen te verwachten was aangezien epidemiologische modellen steeds een (asymptotisch) stabiel evenwicht bereiken. Het berekenen van de MLE voor een stochastisch, asynchroon CA-gebaseerd model is heel tijdsintensief. Om een beter beeld te krijgen van de stabiliteit van zulke modellen, dient er gezocht te worden naar minder rekenintensieve methoden.

4

Parametrisatie van een CASIR

In dit hoofdstuk wordt onderzocht of de optimale waarden van de modelparameters van het CAgebaseerde ziekteverspreidingsmodel CASIR dat wordt voorgesteld in Sectie 1.2.4, te vinden zijn aan de hand van geobserveerde gegevens. Eerst worden de eigenschappen van de gebruikte tesselatie gegeven (Sectie 4.1). Hierna worden de modelresultaten vergeleken met andere ziekteverspreidingsmodellen (Sectie 4.2). Dan wordt een gevoeligheidsanalyse uitgevoerd (Sectie 4.3) om het gedrag en de invloed van alle parameters op de simulatieresultaten te kennen, waarna het model geparametriseerd wordt aan de hand van gesimuleerde gegevens en simulaties met ruis met verschillende optimalisatietechnieken (Sectie 4.4).

4.1 Tesselatie

De gebruikte tesselatie bedekt België en is opgebouwd uit 250000 cellen met elk 50 inwoners. Om aan elke cel een landbedekkingsklasse toe te kennen werd gebruik gemaakt van de Corine databank [99]. Deze databank bevat informatie over de landbedekking binnen de EU en is gebaseerd op satellietfoto's en nationale landgebruik databanken. Het eerste niveau in deze databank bestaat uit vijf klassen, het tweede niveau heeft 15 klassen en het derde niveau heeft 44 klassen. Voor dit model werden de 44 originele klassen herleid tot zes verschillende klassen. Het aandeel van elke klasse in de tesselatie wordt gegeven in Tab. 4.1. De landbouwgronden hebben het grootste aandeel in de tesselatie, maar ook bossen e.d. bedekken een groot deel van de tesselatie. De cellen waarover geen gegevens beschikbaar zijn en de waterlichamen hebben een verwaarloosbaar aandeel in de tesselatie. Fig. 4.1 toont de gebruikte tesselatie waarin de verschillende landbedekkingsklassen zijn aangeduid. Fig. 4.2 toont het aantal buren $|B(c_i)|$ en het aandeel van c_i 's omtrek dat gevormd wordt door elk van zijn buren in $B(c_i)$.

Nr.	Klasse	Aandeel in de tesselatie (%)
1	Steden, fabrieken en industrie	15.1
2	Landbouwgrond	44.0
3	Weiden en natuurlijk grasland	17.1
4	Bos e.d.	20.1
5	Draslanden en waterlichamen	1.5
6	Geen gegevens	2.2

Tabel 4.1: De zes verschillende landbedekkingsklassen en hun aandeel in de gebruikte tesselatie [99]



Figuur 4.1: Landbedekking in het studiegebied

4.2 Vergelijking met andere ziekteverspreidingsmodellen

Enkele simulatieresultaten werden vergeleken met twee verschillende ziekteverspreidingsmodellen. Het eerste model is een model dat gebaseerd is op een stelsel DV (Vgl. (1.17)) dat voor het eerst werd beschreven door Kermack en McKendrick [88]. Hierbij ontbreekt de ruimtelijke component en wordt een perfecte menging van de populatie verondersteld. Figuur 4.3(a) toont de simulaties die met dit stelsel DV werd uitgevoerd waarbij $N = 10^7$, $I_0 = 15$, $R_0 = 0$, $a = 0.5 \ 10^{-7}$ en b = 0.1. Wanneer deze vergeleken worden met de simulatieresultaten van het CASIR model waarbij $N_i = 50$, $I_0 = 15$, $R_0 = 0$, $\mathbf{v_0} = [0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.5 \ 0.5 \ 0.5 \ 0.5 \ 0.5 \ 0.5 \ 0.6 \ 0.5$



Figuur 4.2: Een histogram van (a) $|B(c_i)|$ en (b) het aandeel van c_i 's omtrek dat gevormd wordt door elk van zijn buren $B(c_i)$

In tweede instantie werden enkele simulaties uitgevoerd met het CASIR model die gelijkaardig zijn aan deze die werden uitgevoerd in [1] en waarvan het model werd besproken in Sectie 1.2. Bij een homogene verdeling van de populatie en een constante virulentie in elke cel werden gelijkaardige simulatieresultaten bekomen (Fig. 4.4), hierbij verspreidt een infectiefront zich over de tesselatie. Ook als de tesselatie werd opgedeeld in vier kwartvlakken met elk een verschillende virulentie werden met het CASIR model simulatieresultaten bekomen die in overstemming zijn met de resultaten bekomen in [1] (Fig. 4.5).



Figuur 4.3: Vergelijking tussen het gesimuleerd aantal individuen in elke infectieklasse volgens (a) het model van Kermack en McKendrick [88] ($N = 10^7$, $I_0 = 15$, $R_0 = 0$, $a = 0.5 \ 10^{-7}$ en b = 0.1) en (b) het CASIR model [40]

4.3 Gevoeligheidsanalyse

Een gevoeligheidsanalyse wordt uitgevoerd om het gedrag van het model te kennen wanneer de modelparameters slechts weinig veranderen. Wanneer er bij een kleine perturbatie van een parameter het model geheel andere resultaten levert, zal een parameterschatting moeilijk worden.

4.3.1 Initiële conditie, evaluatiefunctie en modelparameters

Initiële conditie Simulaties met een ruimtelijk expliciet ziekteverspreidingsmodel betreffen meestal een ziekte-uitbraak vanuit één geïnfecteerde regio, wat in het beschouwde model kan bereikt worden door slechts in één cel een fractie geïnfecteerden in te stellen die verschillend is van nul. De epidemie zal zich dan als een golf verspreiden naar de aangrenzende cellen. Aangezien de gevoeligheidsanalyse zoals besproken in Sectie 1.3.1 geen rekening houdt met de ruimtelijke component, zullen de cellen waarin zich geen geïnfecteerden bevinden, de resultaten van de gevoeligheidsanalyse overheersen.



Figuur 4.4: Gesimuleerde spatio-temporele evolutie van het aantal geïnfecteerden gebruikmakend van (a) het model van White *et al.* [1] en (b) het CASIR model waarbij een homogene populatie en een constante virulentie wordt beschouwd

Om deze reden werd als initiële conditie een homogene verdeling van de infectie aangenomen, namelijk 10% geïnfecteerden en 90% vatbaren, dit is $V(c_i, 0) = 0.9 \land I(c_i, 0) = 0.1$, voor alle $c_i \in \mathcal{T}$.

Evaluatiefunctie Zoals reeds vermeld houdt de gevoeligheidsanalyse geen rekening met de ruimtelijke component van het model (cfr. Vgl. (1.30)), zodat een functie dient gekozen te worden die de gevoeligheden in elke cel moet bundelen. Vijf verschillende evaluatiefuncties werden bij elke gevoeligheidsanalyse gebruikt, namelijk de som, het gemiddelde, de mediaan, het minimum en het maximum van de gevoeligheden in alle cellen.

Modelparameters Het CASIR model, zoals beschreven in Sectie 1.2.4, heeft 13 modelparameters, namelijk de fractie geïnfecteerden die herstelt in één tijdstap (ρ), de zes intracelvirulenties (v_0) en de zes intercelvirulenties (v_1). Voor landbedekkingsklasse 6, de klasse waarvan geen gegevens bekend zijn, werden zowel de intra- als intercelvirulentie ($v_{0,6}$ en $v_{1,6}$) steeds gelijkgesteld aan nul. Het aandeel van draslanden en waterlichamen is heel klein (Tab. 4.1), zodat een perturbatie in $v_{0,5}$ en $v_{1,5}$ geen grote verschillen zal opleveren in de modelresultaten. Om deze reden werd er voor $v_{0,5}$ en $v_{1,5}$ geen gevoeligheidsanalyse uitgevoerd. Voor elke andere modelparameter werd een lokale gevoelig-



Figuur 4.5: Gesimuleerde spatio-temporele evolutie van het aantal geïnfecteerden gebruikmakend van (a) het model van White *et al.* [1] en (b) het CASIR model waarbij een homogene populatie en een niet-constante virulentie wordt beschouwd

heidsanalyse uitgevoerd, waarbij telkens één parameter werd geperturbeerd. De andere parameters werden constant verondersteld en hun waarde wordt weergegeven in Tab. 4.2.

4.3.2 Perturbatiefactor

De perturbatiefactor ε bepaalt de afwijking van de parameter θ waarvoor de gevoeligheid bepaald wordt. De optimale ε werd gevonden door het voorwaartse en achterwaartse verschil bij verschillende perturbatiefactoren te berekenen die zo dicht mogelijk bij elkaar moeten liggen (cfr. Sectie 1.3.1). Wanneer een te grote ε gekozen wordt, heeft de niet-lineariteit een invloed op de berekeningen, terwijl een te kleine ε aanleiding geeft tot afrondingsfouten zodat er geen correcte conclusies kunnen getrokken worden over de gevoeligheid.

De optimale perturbatiefactor werd gezocht door zowel de som van de gekwadrateerde fouten tussen het voorwaartse en het achterwaartse verschil (Vgl. (4.1)) als de som van de absolute fouten (Vgl. (4.2)) te berekenen voor $\varepsilon \in [10^{-1}, 10^{-9}]$ over 400 tijdstappen:

Modelparameter	Waarde	Modelparameter	Waarde
$v_{0,1}$	0.2	$v_{1,1}$	0.5
$v_{0,2}$	0.2	$v_{1,2}$	0.5
$v_{0,3}$	0.2	$v_{1,3}$	0.5
$v_{0,4}$	0.2	$v_{1,4}$	0.5
$v_{0,5}$	0.2	$v_{1,5}$	0.5
$v_{0,6}$	0	$v_{1,6}$	0
ρ	0.1		

 Tabel 4.2: De waarden voor de modelparameters wanneer er een lokale gevoeligheidsanalyse wordt uitgevoerd op een andere modelparameter

$$SSE = \sum_{t=1}^{400} \sum_{i=1}^{|\mathcal{T}|} \left(\frac{I^+(c_i, t) - I(c_i, t)}{\varepsilon \ \theta} - \frac{I(c_i, t) - I^-(c_i, t)}{\varepsilon \ \theta} \right)^2$$
(4.1)

$$SAE = \sum_{t=1}^{400} \sum_{i=1}^{|\mathcal{T}|} \left| \frac{I^+(c_i, t) - I(c_i, t)}{\varepsilon \theta} - \frac{I(c_i, t) - I^-(c_i, t)}{\varepsilon \theta} \right|$$
(4.2)

met $I^+(c_i,t)$ het aantal geïnfecteerden bij $\theta^+ = \theta + \varepsilon \ \theta$, $I^-(c_i,t)$ het aantal geïnfecteerden bij $\theta^- = \theta - \varepsilon \ \theta$ en $I(c_i,t)$ het aantal geïnfecteerden bij θ .

Fig. 4.6 toont de SSE en SAE in functie van ε voor het vinden van de optimale ε voor ρ en $v_{0,1}$, waarin duidelijk een minimum te zien is. De ε bij dit minimum werd gebruikt om de gevoeligheidsanalyses uit te voeren voor de belangrijkste modelparameters. Voor ρ treedt dit op bij $\varepsilon = 10^{-8}$ en voor $v_{0,1}$ bij $\varepsilon = 10^{-7}$. Voor de andere virulenties werd dezelfde ε als voor $v_{0,1}$ gebruikt.



Figuur 4.6: Som van de gekwadrateerde fouten (SSE) en absolute fouten (SAE) tussen het voorwaartse en het achterwaartse verschil bij verschillende perturbatiefactoren (ε) voor (a) de fractie geïnfecteerden die herstelt in één tijdstap (ρ) en (b) de intracelvirulentie binnen regio's met landbedekkingsklasse 1 ($v_{0,1}$)

4.3.3 Analyses

De gevoeligheidsanalyses werden uitgevoerd voor elke hierboven besproken modelparameter. Bij elke analyse werd één parameter equidistant gevarieerd tussen 0.1 en 0.9 en bleven de andere parameters onveranderd (Tab. 4.2). De absolute gevoeligheid (cfr. Vgl. (1.30)) werd berekend door het model uit te rekenen met een positief (θ^+) en negatief (θ^-) geperturbeerde parameter zoals hierboven beschreven en dan het centrale verschil te berekenen zoals gestipuleerd door Vgl. (1.33).

De relatieve gevoeligheid (cfr. Vgl. (1.34)) van het CASIR model voor ρ die werd bepaald met de verschillende evaluatiefuncties wordt gegeven in Fig. 4.7. Op deze figuren is duidelijk te zien dat bij de verschillende evaluatiefuncties de absolute waarden van $\rho \frac{\partial I}{\partial \rho}$ sterk verschillen, maar het verloop is wel gelijkaardig. De gevoeligheden evolueren naar 0 als $t \to +\infty$ wat wil zeggen dat er bij een kleine wijziging van ρ geen grote wijziging in de modelresultaten optreedt. Een kleine fout bij de parameterschatting zal dus geen grote invloed hebben op de voorspellingen, waardoor een parameterschatting op observaties mogelijk wordt.

Fig. 4.8 toont de resultaten van de relatieve gevoeligheid van het CASIR model voor de verschillende virulenties. Aangezien met Fig. 4.7 werd aangetoond dat het verloop bij alle evaluatiefuncties gelijk is, wordt enkel het gemiddelde van de gevoeligheden in elke cel weergegeven. Ook bij de verschillende virulenties evolueert de gevoeligheid naar 0, hierdoor zou een goede parameterschatting mogelijk moeten zijn bij gegevens met ruis. De relatieve gevoeligheid voor de intracelvirulentie ligt voor elke landbedekkingsklasse lager dan die van de intercelvirulentie, waaruit blijkt dat de verspreiding van de ziekte tussen de cellen in de tesselatie een grotere invloed heeft dan de verspreiding binnen de cel.

De gevoeligheid kan ook ruimtelijk weergegeven worden, namelijk door de gevoeligheid in elke cel in een bepaalde tijdstap af te beelden. Zo toont Fig. 4.9 de ruimtelijke verdeling van de gevoeligheid van de intra- en intercelvirulentie voor landbouwgronden ($v_{0,2}$ en $v_{1,2}$). Wat betreft de intracelvirulentie zijn de gevoeligheden het grootst in de cellen met landbouwgrond, doordat enkel in deze cellen de intracelvirulentie werd veranderd. Bij de intercelvirulentie zijn de gevoeligheden groter en meer verspreid, doordat de verschillen zich sneller doorzetten naar cellen in de omgeving van een cel met landbouwgrond.

De absolute gevoeligheden van de verschillende modelparameters worden vergeleken in Fig. 4.10. Uit Fig. 4.10(a) blijkt duidelijk dat ρ de grootste invloed heeft op de modelresultaten, want ρ heeft een grotere gevoeligheid, niettegenstaande er een kleinere ε gebruikt werd voor de gevoeligheidsanalyse. Wanneer een parameterschatting wordt uitgevoerd, zal een correcte schatting van ρ het belangrijkst zijn, aangezien er door een kleine afwijking van deze parameter, een groot verschil tussen de modelsimulaties ontstaat. Op Fig. 4.10(b) worden enkel de intra- en intercelvirulenties weergegeven, waaruit duidelijk blijkt dat de gevoeligheid van de virulenties positief gecorreleerd is met het aandeel in de landbedekking. Zo hebben $v_{0,2}$ en $v_{1,2}$ een grotere invloed op de modelresultaten dan $v_{0,1}$ en $v_{1,1}$.


Figuur 4.7: Relatieve gevoeligheid van het CASIR model voor ρ variërend van 0.1 tot 0.9 die wordt berekend als (a) de som, (b) het gemiddelde, (c) de mediaan, (d) het minimum en (e) het maximum van de gevoeligheden van de individuele cellen



Figuur 4.8: Het gemiddelde van de relatieve gevoeligheden van de individuele cellen van het CASIR model voor (a) $v_{0,1}$, (b) $v_{1,1}$, (c) $v_{0,2}$, (d) $v_{1,2}$, (e) $v_{0,3}$, (f) $v_{1,3}$, (g) $v_{0,4}$ en (h) $v_{1,4}$



(a)



(b)

Figuur 4.9: Ruimtelijke verspreiding van de gevoeligheid in elke cel van de onregelmatige tesselatie op tijdstap 8, voor (a) de intercelvirulentie bij landbouwgronden met $v_{0,2} = 0.1$ en (b) de intracelvirulentie bij landbouwgronden met $v_{1,2} = 0.1$



Figuur 4.10: De gemiddelde gevoeligheden van de individuele cellen van het CASIR model bij $\theta = 0.1$ voor (a) alle onderzochte modelparameters en (b) enkel de intra- en intercelvirulenties

4.4 Parametrisatie

4.4.1 Gesimuleerde gegevens

Er werd een ziekte-uitbraak gesimuleerd met het CASIR model gebruikmakende van de modelparameters uit Tab. 4.3. Om deze te simuleren werd de initiële conditie zo gekozen dat slechts één cel een fractie geïnfecteerde individuen bevatte die verschillend was van nul. Deze cel werd zo gekozen dat de infectiegolf in een minimaal aantal tijdstappen alle cellen in de tesselatie bereikt. De gesimuleerde fractie geïnfecteerden in elke cel en elke tijdstap werden dan gebruikt om na te gaan of deze modelparameters kunnen geschat worden met gekende optimalisatietechnieken.

Modelparameter	Waarde	Modelparameter	Waarde
$v_{0,1}$	0.3	$v_{1,1}$	0.6
$v_{0,2}$	0.25	$v_{1,2}$	0.55
$v_{0,3}$	0.2	$v_{1,3}$	0.5
$v_{0,4}$	0.15	$v_{1,4}$	0.45
$v_{0,5}$	0.1	$v_{1,5}$	0.4
$v_{0,6}$	0	$v_{1,6}$	0
ρ	0.1		

Tabel 4.3: De waarden voor alle modelparameters waarmee de gesimuleerde gegevens werden bekomen

Aangezien observaties nooit de waarden zoals gesimuleerd met het model aannemen, werd er ook ruis toegevoegd aan de gesimuleerde gegevens. Om deze ruis te bekomen werd een normaal verdeelde ruis met gemiddelde 0 en standaardafwijking 0.01 aan de gesimuleerde gegevens toegevoegd. Negatieve fracties zijn onmogelijk en werden dus vervangen door nul, fracties groter dan één werden om dezelfde reden vervangen door één.

4.4.2 Parameterschatting

Het gebruikte ziekteverspreidingsmodel heeft, zoals weergegeven in Tab. 4.2, 13 modelparameters. De waarden van deze parameters zijn afhankelijk van de epidemie. Aan de hand van observaties wordt een parameterschatting gemaakt door een doelfunctie te optimaliseren. In deze thesis werd als doelfunctie de som van de gekwadrateerde fouten tussen de geobserveerde en de berekende gegevens gebruikt:

$$SSE = \sum_{t=1}^{T} \sum_{i=1}^{|\mathcal{T}|} (I_{obs}(c_i, t) - I(c_i, t))^2,$$
(4.3)

met T het aantal tijdstappen. Het minimaliseren van deze fout leidt tot de optimale parameterwaarden. De gebruikte optimalisatietechnieken, namelijk Simulated Annealing [91], Nelder-Mead [92], Differential Evolution [94] en Random Search [95], werden reeds besproken in Sectie 1.3. De parametrisatie werd via een ingebouwde functie in Mathematica 7.0 (Wolfram Research Inc., Champaign, VSA) uitgevoerd. Het model, dat parallel geïmplementeerd was, diende hiervoor aangepast te worden naar een niet-parallele implementatie. Hierdoor nam de benodigde rekentijd voor het simuleren van het model drastisch toe.

Aangezien uit de gevoeligheidsanalyse blijkt dat ρ de meest gevoelige modelparameter is, werd onderzocht of een schatting van deze parameter mogelijk is. Eerst werden de simulaties zonder ruis gebruikt om ρ te schatten. Door het deterministisch karakter van het model, kan reeds na één tijdstap de juiste waarde voor de parameter gevonden worden. Indien echter observaties gebruikt worden, zullen meerdere tijdstappen nodig zijn om een goede parameterschatting te bekomen. In Fig. 4.11 wordt de rekentijd voor het vinden van ρ weergegeven voor de vier verschillende optimalisatietechnieken, in functie van het aantal tijdstappen. Deze rekentijd neemt exponentieel toe waardoor de rekentijd voor tien tijdstappen met Simulated Annealing meer dan 90 uur bedraagt. Uit deze resultaten blijkt ook dat Random Search het meest tijdsintensief is, omdat deze techniek heel goed uit lokale optima kan ontsnappen om het globale optimum te vinden.



Figuur 4.11: De rekentijden voor het vinden van de optimale waarde voor ρ met de verschillende optimalisatietechnieken in functie van het aantal tijdstappen

Met betrekking tot de simulaties met ruis is duidelijk te zien op Fig. 4.12 dat de kwaliteit van de

schatting bij een verhoging van het aantal tijdstappen steeds beter wordt. De vier verschillende technieken vinden in alle tijdstappen dezelfde waarde voor ρ , dus er kan aan de hand hiervan niet besloten worden welke de beste techniek is. De nodige rekentijd is echter wel een belangrijk criterium in het kiezen van een goede techniek. Er moet gezocht worden naar een methode die minder tijdsintensief is om een goede parameterschatting uit te voeren, zodat ook voor grote datasets de optimale waarden van alle modelparameters kunnen gevonden worden.



Figuur 4.12: De geschatte ρ met Simulated Annealing bij een oplopend aantal tijdstappen voor simulaties met ruis bij $\rho = 0.1$

4.5 Conclusie

In dit hoofdstuk werd onderzocht of de parametrisatie van een CA-gebaseerd ziekteverspreidingsmodel aan de hand van observaties mogelijk is. Eerst werd een tesselatie en een initiële conditie gekozen, waarmee gevoeligheidsanalyses werden uitgevoerd voor de belangrijkste modelparameters. Uit deze gevoeligheidsanalyses blijkt dat parametrisatie voor alle modelparameters mogelijk is, aangezien de gevoeligheid voor alle parameters evolueert naar nul. Dit is te wijten aan het feit dat ziekteverspreidingsmodellen een (asymptotisch) stabiel evenwicht hebben. De gevoeligheidsanalyses toonden verder aan dat ρ de meest gevoelige modelparameter is. Daarom werd deze geschat in de zoektocht naar een goede optimalisatietechniek.

Om de parameterschatting uit te voeren werd met het CASIR model een dataset aangemaakt. Vermits het een deterministisch model betreft leverden de beschouwde optimalisatietechnieken al na één tijdstap heel goede parameterschattingen. Echter, als het aantal tijdstappen groter wordt, neemt de rekentijd exponentieel toe wat zorgt voor grote problemen indien observaties gebruikt worden. Immers, deze gegevens worden gekenmerkt door veel ruis zodat er meer tijdstappen nodig zijn om de optimale waarde te vinden.

5

Algemene conclusies en verder werk

"I expect the children of 50 years from now will learn cellular automata before they learn algebra."

Stephen Wolfram

5.1 Algemene conclusies

Om te onderzoeken of een parametrisatie van een model mogelijk is, moet vooreerst de stabiliteit van het dynamisch systeem onderzocht worden. In Hoofdstuk 2 werd de invloed van de topologie op de stabiliteit van CA onderzocht. Voor alle 256 regels binnen de familie van (2,7) totalistische onregelmatige CA werd de maximale Lyapunov exponent (MLE) berekend bij een verschillende topologie. Hieruit bleek dat deze kunnen opgedeeld worden in vijf verschillende categorieën in functie van hun gedrag bij een veranderende topologie. Op basis van het bifurcatiepunt van de regels die bij hoge connectiviteit onstabiel zijn en bij lage connectiviteit superstabiel worden, kunnen deze eveneens in vijf categorieën opgedeeld worden. Er treden soms grote veranderingen op in het gedrag van CA bij een kleine wijziging in de topologie. De meeste studies naar het gedrag van CA gebruiken evenwel slechts één topologie waardoor de conclusies van deze studies niet éénduidig zijn.

In het volgende hoofdstuk werd een CA-gebaseerd SI model geïmplementeerd naar een artikel van Karim *et al.* [98] om de MLE te berekenen. Bij deze implementatie werden evenwel veel onvolledigheden en inconsistenties in het SI model gevonden. Na het aanbrengen van enkele aanpassingen, werden de simulaties uit [98] opnieuw uitgevoerd. De simulatieresultaten bleken sterk af te wijken van deze in [98], waaruit volgt dat de conclusies uit het artikel niet verifieerbaar zijn. Tot slot werd van dit systeem de MLE berekend, maar dit was heel tijdsintensief door het stochastische karakter van het SI model. Deze MLE bewees dat het systeem stabiel is, wat typisch is voor een epidemiologisch model.

Ten slotte werd in Hoofdstuk 4 een ziekteverspreidingsmodel met drie mogelijke toestanden, namelijk het CASIR model besproken [40]. Eerst werd onderzocht of een parametrisatie via optimalisatietechnieken mogelijk is door een gevoeligheidsanalyse uit te voeren van alle modelparameters. Uit deze

analyses bleek dat een kleine verandering in een modelparameter nooit grote veranderingen in de modelresultaten na vele tijdstappen teweeg brengt. De modelparameters kunnen dus geschat worden. Uit dezelfde analyses kon geconcludeerd worden dat een kleine wijziging van de fractie geïnfecteerden die herstelt in één tijdstap de grootste invloed heeft op de modelresultaten. In het licht van deze thesis werd enkel deze modelparameter geschat. Wanneer gesimuleerde gegevens gebruikt werden, kon de parameter al na één tijdstap geschat worden, maar bij geobserveerde gegevens met veel ruis zijn meerdere tijdstappen nodig om de modelparameters nauwkeurig te schatten. De rekentijd neemt echter exponentieel toe met het aantal tijdstappen waardoor de parameterschatting bijna onmogelijk wordt.

5.2 Verder werk

De invloed van de topologie op de stabiliteit van CA kan nog verder onderzocht worden met bijvoorbeeld topologische Lyapunov exponenten. Na verder onderzoek kunnen andere inzichten gevonden worden in het gebruik van deze Lyapunov exponenten om CA te classificeren. Het berekenen van de Lyapunov exponenten voor systemen met een groter aantal mogelijke toestanden of voor CML's moet ook nog verder uitgewerkt worden.

Voor de parametrisatie van het CASIR model moeten minder rekenintensieve oplossingen gevonden worden, zodat er optimale parameterwaarden kunnen geschat worden uitgaande van observaties. Want enkel met geobserveerde gegevens kan het model gevalideerd worden. Deze observaties hoeven niet per se afkomstig te zijn van een ziekte-uitbraak, maar er kunnen ook gegevens van bosbranden gebruikt worden omdat deze ongeveer dezelfde dynamiek als een ziekte-uitbraak vertonen. Verder zou het ook interessant zijn om andere factoren die een invloed hebben op de verspreiding van ziekten te onderzoeken en te implementeren in het CASIR model. Er dient bijvoobeeld rekening gehouden te worden met het internationaal transport van goederen en mensen.

Tot slot moet vermeld worden dat er nog veel onderzoek kan gebeuren naar het gebruik van CA voor de beschrijving van macroscopische processen. De eenvoudige transitieregels, waardoor een CA snel en precies berekend kan worden, hebben hierbij een grote meerwaarde.

Bibliografie

- S. H. White, A. M. del Rey, G. R. Sánchez, Modeling epidemics using cellular automata, Applied Mathematics and Computation 486 (2007) 193–202.
- [2] J. D. Murray, Mathematical Biology, Springer-Verlag, New York, VSA, 1993.
- [3] J. D. Murray, Mathematical Biology II: Spatial Models and Biomedical Applications, Springer-Verlag, New York, VSA, 2003.
- [4] L. Berec, Techniques of spatially explicit individual-based models: construction simulation and mean-field analysis, Ecological Modelling 150 (2002) 55–81.
- [5] A. Alexandridis, D. Vakalis, C. Siettos, G. Bafas, A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through Spetses Island in 1990, Applied Mathematics and Computation 204 (2008) 191–201.
- [6] C. Lett, C. Silber, N. Barret, Comparison of a cellular automata network and an individual-based model for the simulation of forest dynamics, Ecological Modelling 121 (1999) 277–293.
- [7] C. J. Rhodes, R. M. Anderson, Dynamics in a lattice epidemic model, Physics Letters A 210 (1996) 183–188.
- [8] W. Dzwinel, Spatially extended populations reproducing logistic map, Central European Journal of Physics 8 (2010) 33–41.
- [9] A. Brannstrom, D. Sumpter, Coupled map lattice approximations for spatially explicit individual-based models of ecology, Bulletin of Mathematical Biology 67 (2005) 663–682.
- [10] H. Sakaguchi, M. Ohtaki, A coupled map lattice model for dendritic patterns, Physica A 272 (1999) 300–313.
- [11] J. M. Baetens, B. De Baets, A generalized coupled-map lattice to model biological phenomena, Mathematical BiologyIn preparation.
- [12] S. Krone, Y. Guan, Spatial self-organization in a cyclic resource-species model, Journal of Theoretical Biology 241 (2006) 14–25.
- [13] M. Katori, Percolation transitions and wetting transitions in stochastic models, Brazilian Journal of Physics 30 (2000) 83–96.

- [14] M. Doebeli, G. Ruxton, Stabilization through spatial pattern formation in metapopulations with long-range dispersal, Proceedings of the Royal Society of London series B 265 (1998) 1325– 1332.
- [15] B. Szymanski, T. Caraco, Spatial-analysis of vector-borne disease A 4-species model, Evolutionary Ecology 8 (1994) 299–314.
- [16] T. Pei, A.-X. Zhu, C. Zhou, B. Li, C. Qin, Detecting feature from spatial point processes using Collective Nearest Neighbor, Computers Environment and Urban Systems 33 (2009) 435–447.
- [17] C. Ji, D. Merl, T. Kepler, M. West, Spatial mixture modelling for unobserved point processes: examples in Immunofluorescence histology, Bayesian Analysis 4 (2009) 297–315.
- [18] V. LeMay, A. Pommerening, P. Marshall, Spatio-temporal structure of multi-storied, multi-aged interior Douglas fir (Pseudotsuga menziesii var. glauca) stands, Journal of Ecology 97 (2009) 1062–1074.
- [19] L. Guo, S. Billings, D. Coca, Identification of partial differential equation models for a class of multiscale spatio-temporal dynamical systems, International Journal of Control 83 (2010) 40–48.
- [20] A. Matzavinos, M. Chaplain, V. Kuznetsov, Mathematical modelling of the spatio-temporal response of cytotoxic T-lymphocytes to a solid tumour, Mathematical Medicine and Biology 21 (2004) 1–34.
- [21] L. Guo, S. Billings, Identification of partial differential equation models for continuous spatiotemporal dynamical systems, IEEE Transactions on Circuits and Systems II-Express Briefs 53 (2006) 657–661.
- [22] J. von Neumann, Theory of Self-Reproducing Automata, University of Illinois Press, Illinois, VSA, 1966.
- [23] M. Gardner, Wheels, Life, and other Mathematical Amusements, W. H. Freeman and Company, New York, VSA, 1983.
- [24] S. Wolfram, A New Kind of Science, Wolfram Media, Illinois, VSA, 2002.
- [25] A. Ilachinski, Cellular Automata, A Discrete Universe, World Scientific Publishing, Singapore, 2001.
- [26] J. M. Baetens, B. De Baets, Tracking uncertainty in a spatially explicit susceptible-infected epidemic model, Lecture Notes in Computer Science (2010) 95–105.
- [27] A. Flache, R. Hegselmann, Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics, Journal of Artificial Societies and Social Simulation 4 (2001) 6.1–6.27.
- [28] A. Okabe, B. Boots, K. Sugihara, Spatial Tessellations Concepts and Applications of Voronoi Diagrams, Wiley & Sons, Chichester, VK, 1992.

- [29] J. M. Baetens, B. De Baets, Towards generalized measures grasping CA dynamics, Lecture Notes in Computer Science (2010) 177–187.
- [30] J. Schiff, Cellular Automata: A Discrete View of the World, John Wiley & Sons Ltd, Hoboken, VSA, 2008.
- [31] S. Wolfram, Cellular Automata and Complexity, Westview press, Boulder, Colorado, VSA, 1994.
- [32] F. Bagnoli, R. Rechtman, Damage spreading and Lyapunov exponents in cellular automata, Physics Letters A 172 (1992) 34–38.
- [33] S. Wolfram, Universality and complexity in cellular automata, Physica D 10 (1984) 1–35.
- [34] C. G. Langton, Computation at the edge of chaos, Physica D 42 (1990) 12–37.
- [35] M. Courbage, B. Kaminski, Space-time directional Lyapunov exponents for cellular automata, Journal of Statistical Physics 124 (2006) 1499–1509.
- [36] P. Tisseur, Cellular automata and Lyapunov exponents, Nonlinearity 13 (2000) 1547–1560.
- [37] J. Urías, R. Rechtman, A. Enciso, Sensitive dependence on initial conditions for cellular automata, Chaos 7 (1997) 688–693.
- [38] M. Cencini, F. Cecconi, A. Vulpiani, Chaos: From Simple Models to Complex Systems, World Scientific, Singapore, 2010.
- [39] F. Bagnoli, R. Rechtman, S. Ruffo, Damage spreading and Lyapunov exponents in cellular automata, Physics Letters A 172 (19920) 34–38.
- [40] J. M. Baetens, B. De Baets, Phenomenological study of irregular cellular automata based on Lyapunov exponents and Jacobians, Chaos 20 (2010) 033112.
- [41] P. Gerlee, A. R. A. Anderson, Stability analysis of a hybrid cellular automaton model of cell colony growth, Physical Review E 75 (2007) 1–8.
- [42] Y. Jiang, O. Sozinova, M. Alber, On modeling complex collective behavior in myxobacteria, Advances in Complex Systems 9 (2006) 353–367.
- [43] C. Picioreanu, M. van Loosdrecht, J. Heijnen, Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach, Biotechnology and Bioengineering 58 (1998) 101–116.
- [44] J. D. Chambless, S. M. Hunt, P. S. Stewart, A three-dimensional computer model of four hypothetical mechanisms protecting biofilms from antimicrobials, Applied and Environmental Microbiology 3 (2006) 2005–2013.
- [45] J.-U. Kreft, C. Picioreanu, J. W. T. Wimpenny, M. C. M. van Loosdrecht, Individual-based modelling of biofilms, Microbiology 147 (2001) 2897–2912.

- [46] T. Roose, S. J. Chapman, P. K. Maini, Mathematical models of avascular tumor growth, Siam Review 49 (2007) 179–208.
- [47] J. Moreira, A. Deutsch., Cellular automaton models of tumor development: A critical review, Advances in Complex Systems 5 (2002) 247–267.
- [48] A. R. Kansal, S. Torquato, G. R. Harsh, E. A. Chiocca, T. S. Deisboeck, Simulated brain tumor growth dynamics using a three-dimensional cellular automaton, Journal of Theoretical Biology 203 (2000) 367–382.
- [49] G. Schaller, M. Meyer-Hermann, Multicellular tumor spheroid in an off-lattice Voronoi-Delaunay cell model, Physical Review E 71 (2005) 051910.
- [50] H. Ueda, Y. Iwaya, T. Abe, T. Kinoshita, A cellular automata model considering diversity associated with HIV infection, Artif Life Robotics 10 (2006) 73–76.
- [51] F. Celada, P. E. Seiden, A computer model of cellular interactions in the immune system, Immunology Today 13 (1992) 56–62.
- [52] G. Folino, G. Mendicino, A. Senatore, G. Spezzano, S. Straface, A model based on cellular automata for the parallel simulation of 3D unsaturated flow, Parallel Computing 32 (2006) 357– 376.
- [53] D. Raabe, Overview of the lattice Boltzmann method for nano- and microscale fluid dynamics in materials science and engineering, Modelling and Simulation in Materials Science and Engineering 12 (2004) R13–R46.
- [54] S. Takesue, Fourier's law and the green-kubo formula in a cellular-automaton model, Physical Review Letters 64 (1990) 252–255.
- [55] K. Saito, S. Takesue, S. Miyashita, Transport phenomena at a critical point: Thermal conduction in the creutz cellular automaton, Physical Review E 59 (1999) 2783–2794.
- [56] J. Seo, J. Moore, P. Gordon, A sequential land use/transportation model with externalities: Linking the dynamics of regional economic growth and urban spatial structure, Transportation Research, Economics and Policy 4 (1996) 383–411.
- [57] S. Serneels, M. Linderman, E. F. Lambin, A multilevel analysis of the impact of land use on interannual land-cover change in east africa, Ecosystems 10 (2007) 402–418.
- [58] S. Rickebusch, M. Gellrich, H. Lischke, A. Guisan, N. E. Zimmermann, Combining probabilistic land-use change and tree population dynamics modelling to simulate responses in mountain forests, Ecological Modelling 209 (2007) 157–168.
- [59] C. Kamusoko, M. Aniya, B. Adi, M. Manjoro, Rural sustainability under threat in Zimbabwe simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model, Applied Geography (2008) 1–13.
- [60] N. N. Pinto, A. P. Antunes, A cellular automata model based on irregular cells: application to small urban areas, Environment and Planning B: Planning & Design 37 (2010) 1095–1114.

- [61] E. Marshall, T. O. Randhir, Spatial modeling of land cover change and watershed response using markovian cellular automata and simulation, Water Resources Research 44 (2008) 005514.
- [62] G. Engelen, R. White, I. Uljee, P. Drazan, Using cellular automata for integrated modelling of socio-environmental systems, Environmental Monitoring and Assessment 34 (1995) 203–214.
- [63] A. Hernández Encinas, L. Hernández Encinas, S. Hoya White, A. Martín del Rey, G. Rodríguez Sánchez, Simulation of forest fire fronts using cellular automata, Advances in Engineering Software 38 (2007) 372–378.
- [64] L. Bodrozic, D. Stipaniev, M. Seric, Forest fires spread modeling using cellular automata approach, in: Modern Trends in Control, Kosice, Slovakia, 2006, pp. 23–33.
- [65] S. Yassemi, S. Dragicevic, M. Schmidt, Design and implementation of an integrated GIS-based cellular automata model to characterize forest fire behaviour, Ecological Modelling 210 (2008) 71–84.
- [66] A. Collin, D. Bernardin, O. Sero-Guillaume, A physical-based cellular automaton model for forest-fire propagation, Combustion science and technology 183 (2011) 347–369.
- [67] C. Bone, S. Dragicevic, R. A., Integrating high resolution remote sensing, GIS and fuzzy set theory for identifying susceptibility areas of forest insect infestations, International Journal Of Remote Sensing 26 (2005) 4809–4828.
- [68] C. Bone, S. Dragicevic, A. Roberts, A fuzzy-constrained cellular automata model of forest insect infestations, Ecological Modelling 192 (2006) 107–125.
- [69] S. Franklin, M. Wulder, R. Skakun, A. Carroll, Mountain pine beetle red-attack forest damage classification using stratified Landsat TM data in British Columbia, Canada, Photogrammetric Engineering and Remote Sensing 69 (2003) 283–288.
- [70] Y. Chen, W. Jau-Wen, Agent-based simulation of agricultural prices volatility using cellular automata, in: 2007 Second International Conference on Innovative Computing, Information and Control, IEEE Computer Society, 2007, pp. 857–60, 2007 Second International Conference on Innovative Computing, Information and Control, 5-7 September 2007, Kumamoto, Japan.
- [71] Y. Fan, S.-J. Ying, B.-H. Wang, Y.-M. Wei, The effect of investor psychology on the complexity of stock market: an analysis based on cellular automaton model, Computers & Industrial Engineering 56 (2009) 63–9.
- [72] S. Darbha, K. Rajagopal, V. Tyagi, A review of mathematical models for the flow of traffic and some recent results, Nonlinear Analysis: Theory, Methods & Applications 69 (2008) 950 – 970.
- [73] B. Jia, X.-G. Li, Z.-Y. Gao, R. Jiang, A realistic two-lane cellular automata traffic model considering aggressive lane-changing behavior of fast vehicle, Physica A 367 (2006) 479–486.
- [74] X. Hai-lin, D. Min-yi, K. Ling-jiang, L. Mu-ren, Influence of people's moving on the opinion communication in the cellular automation public opinion model, Journal of Systems Engineering 20 (2005) 225–231.

- [75] S. G. Alves, N. M. O. Neto, M. L. Martins, Electoral surveys' influence on the voting processes: a cellular automata model, Physica A - Statistical Mechanics and Its Applications 316 (2002) 601–614.
- [76] K. C. Clarke, Getting Started with Geographic Information Systems, Prentice Hall, 1995.
- [77] D. Pullar, Integrating dynamic spatial models with discrete event simulation., in: SimTecT2000 proceedings, Sydney, Australia, 2000.
- [78] S. Watts, Epidemics and history: disease, power, and imperialism, Yale University Press, 1999.
- [79] M. R. Smallman-Raynor, A. D. Cliff, B. Trevelyan, Poliomyelitis: A World Geography: Emergence to Eradication, Oxford University Press, New York, VSA, 2006.
- [80] I. W. Sherman, The power of plagues, ASM Press, Washington D.C., VSA, 2006.
- [81] L. Yount, Epidemics, Lucent Books, San Diego, VSA, 2000.
- [82] L. O. Kallings, The first postmodern pandemic: 25 years of HIV/AIDS, Journal of Internal Medicine 263 (2008) 218–243.
- [83] S. C. Leão, F. Portaels, Tuberculosis 2007: From basic science to patient care, B. S. Kamps and P. Bourcillier., 2007.
- [84] P. I. Trigg, A. V. Kondrachine, Malaria: parasite biology, pathogenesis, and protection, American Society for Microbiology, Washington D.C., VSA, 1998.
- [85] R. S. Gottfried, The Black Death : natural and human disaster in Medieval Europe, Free press, New York, VSA, 1983.
- [86] G. Kolata, Flu: the story of the great influenza pandemic of 1918 and the search for the virus that caused it, Simon & Schuster, 2001.
- [87] J. Gani, Modelling epidemic diseases, Australian & New Zealand Journal of Statistics 52 (2010) 321–337.
- [88] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A 115 (1927) 700–721.
- [89] T. Turányi, H. Rabitz, Sensitivity Analysis, John Wiley & Sons Ltd., West Sussex, VK, 2000.
- [90] V. Kocabas, S. Dragicevic, Assessing cellular automata model behaviour using a sensitivity analysis approach, Computers, Environment and Urban Systems 30 (2006) 921–953.
- [91] Z. Michalewicz, D. Fogel, How to Solve It: Modern Heuristics, Springer, 2000.
- [92] J. A. Nelder, R. Mead, A simplex method for function minimization, Computer Journal 7 (1965) 308–313.
- [93] W. Spendley, G. R. Hext, F. R. Himsworth, Sequential application of simplex designs in optimisation and evolutionary operation, Technometrics 4 (1962) 441–461.

- [94] R. Storn, T. Price, Differential evolution a simple and efficient heuristic for global optimization over continuous spaces, Journal of Global Optimization 11 (1997) 341–359.
- [95] F. J. Solis, R. J.-B. Wets, Minimization by random search techniques, Mathematics of operations research 6 (1981) 19–30.
- [96] Y. Kuznetsov, Elements of Applied Bifurcation Theory, Second Edition, Springer-Verlag, New York, VSA, 1998.
- [97] J. M. Baetens, B. De Baets, On topological sensitivity of cellular automata, Chaos 21 (2011) 023108.
- [98] M. F. A. Karim, A. I. M. Ismail, Cellular automata modelling of hantarvirus infection, Chaos, Solitons & Fractals 41 (2009) 2847–2853.
- [99] M. Bosard, J. Feranec, J. Otahel, Corine land cover technical guide addendum 2000, Tech. Rep. 40, European Environmental Agency, Copenhagen, Denmark (2000).