

Techniques for stochastic spatial sensing in biology: From immunology to anthropology



Citlali Solís Salas
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2020

Acknowledgements

I want to thank the Mexican National Council of Science and Technology (CONACyT) for the granted scholarship, which carried me through the first 5 years of the doctorate. This thesis has been possible due to the support of the scholarship granted by the Secretaryship of Public Education and the Mexican Government. I also want to thank Fundación UNAM for the economic support. St. Catherine's College helped me, amongst other ways, through several hardship grants and the Light Senior Scholarship. Finally, I want to thank the University of Oxford for the support lent through the Hardship Scholarship.

Esta tesis se ha realizado con el apoyo de beca de la Secretaría de Educación Pública y del Gobierno Mexicano.

What follows now is an unusually long list (for a doctorate, I guess) of people I would like to thank. These are all people I admire and cherish, and there are so many of them because each gave me a bit of strength to keep going. I also now understand why winners of The Oscars must be shooed off-stage.

First and foremost, I would like to thank my supervisors Prof. Philip Maini, Dr. Thomas Woolley, Prof. Omer Dushek, Prof. Robin Dunbar, and Dr. Eiluned Pearce for their ideas, persistence, support, and guidance throughout these years. They did more than guide me academically, they cared for my well-being, and they put in a lot of effort to work together. Thank you for your patience and advice.

I would not have even gotten to Oxford without the support of many, amongst which are Prof. Eugenio Ley-Koo, Prof. and cheerleader Jocene Wild, Prof. Mauricio Fortes, Prof. Enrique Cabrera, Katie Leeke, and, particularly, Prof. Denis Boyer, who took me in, taught me the basics of mathematical biology, and was always kind to me. I want to emphatically thank Prof. Rubén Santamaría Ortiz, Martha Santamaría Ortiz, and Eng. Jaime Lomelín Guillen, whose support was relentless and opened doors. José Luis was instrumental to the first few years of my life in Oxford, from helping me to get to Oxford to teaching me how to live independently, and I am forever grateful.

During my life in Oxford I have met wonderful people in both expected and unexpected places. They lent invaluable helping hands or words of encouragement that helped me push on. From the institutions that supported me I have to thank Sandhya Patel, who went over and beyond her job to help graduate students, Prof. Ashok Handa and Richard Molyneux, who made sure I felt welcome and supported by St. Catherine's College, Prof. Elspeth Garman, who put effort and care into her role as director of the DTC's Systems Biology Programme, and Dr. Naomi Drury, who was attentive and kind when listening to my ailments and explaining the possible courses of action. I find they all share a will to do better than what is expected of them and an instinct of protection towards those in their care.

On the other side of the coin are the students with whom I shared laughs, complaints, meals, weather chats, coffee, tea, and hot chocolate. Amazing Judith shared both knowledge and possessions (like my first couple of glasses in Oxford) from the very first moment I met her. Barbara walked with me through the whole process and seeing her push through thick and thin gave me hope. Samara showed me confidence and power. Chico and Yayoi shared closeness and bubbiness. John, Marion, Dan, Annet, and Pavel made the DTC a place to look forward to. Rachel Hong Sheng, Abby, Ellie, Phil, and Marcus were joyful, warm, encouraging, and made me feel welcome in and outside the lab. I was lucky enough to share my

office with Chris and Ana, where lots of cheering and advice took place. Boy, were those times turning points for my path, but oh so necessary. Michael helped me keep young at heart, sharing quirkiness and goals in life. A special thanks to Jakub and Markéta for providing their care and friendship from the very first day I stepped into Oxford. They might not know how much they have helped, but having their support and feeling understood helped me look at myself more kindly. Liam saw me struggle and made it his goal to not let me feel alone. He has succeeded and he continues to be a rock when in need and fun when celebrating.

Life is fortunately not all about work, so I found support in many more places. Dinners and chats with Tom and Lucy made me happy to be in this new world, whilst later Francisco, Rhea, Ninad, Poncho, and Andrea gave me a sense of family away from home. Whole days of cooking, eating, and watching *One Punch Man* were spent in that Mexican house. Living on my own became much more enjoyable thanks to Jan and Nikola. I met a random guy on the bus once and he has become one of my best friends since; thanks, Luke. I loved chatting with Simon and Francis, they were either funny or had good gossip (mainly about themselves). Tristan, Joe, Avi, Archana, and Hamza let me be part of their cosy community, hurrah for board games and food. Leo and Sonia have made me laugh, dance, and be generally merry. Fabhiola has also made me dance like crazy, but her main role has been of cheerleader and confidant. Her happiness from my achievements strengthens the sense that we have gotten here together. Of course, I shared most of these events and this journey with Alexei, and I am thankful I did.

Finally, I must fully thank those who accompanied me from Mexico, those who cheered me on, shared themselves through time and space, and made me feel I had not fully left. I felt utmost support from Male, Diana, and Magdalena, who have known me all my life. Carmen, Laura, la tía Tere, Jorge, and Mari knew it was difficult to be away but kept me, my parents, and themselves cheerful. I had an abundance of encouraging chats with Irene, Laurita, Ana Paola, Zarcito, Itzcoatl, Rodrigo, Lals, Adriana, Vero, Ollintzin, and Pavel, some of which reminded me that I could do it, and some of which reminded me that I could just drop everything and go back. Do not underestimate the power of knowing that you can drop everything, go back, and still be cherished. For 14 years now I have enjoyed the friendship of Luigi, Iker, Corro, and David. We have made trips, gotten lost, explored new cities, given advice to one another, partied, studied, and done so much more together, and I hope this keeps on going for years to come. My list of support would not be complete without César, who has taken my hand and helped me push through the last bits. Sharing laughter and warmth with him has brought a piece of home to my heart. These are all people whose importance I can not understate.

Finally, it is clear I would not be me without my family. The love and support they have given me makes all possible. My brother, Luis Eduardo, has understood me best and known what to say when I needed a clearer view. My mother, Patricia, and father, Miguel Ángel, have done an incredible job learning with me and showing their love throughout thick and thin. I dedicate all that has been done here to them three. And Frijolín.

Statement of Originality

I declare that the contents of this thesis are, to the best of my knowledge, original and my own work, except where otherwise indicated, cited, or commonly known, nor has any part of this thesis been submitted for a degree at any other institution.

Abstract

There is a large branch of mathematics applied to biological problems, and many of these problems deal with the movement and behaviour of many bodies. Nevertheless, biology has many problems where the techniques used for mass movement is not accurate enough. In this thesis we use mathematical and computational tools applied to such biological problems where, furthermore, the relative position of each element is important to the outcome of the system. The first of the problems is in immunology, where we focus on the interaction between a protein hooked to the inside of a T cell and an enzyme floating in the cell's cytoplasm. The second biological problem we tackle on is in anthropology, where we ask ourselves whether nomadic groups of human can meet enough times in a year to maintain their friendships. In immunology we propose models of integrate-partial differential equations with which we can compare experimental data from the laboratory or results from computational simulations, and gain information about the system's reaction rates and particle lengths. These models use auto- and pair-correlation functions to take spatial effects into account. In Chapter 2 we deal with a dephosphorylating enzyme, SHP-1, and phosphorylated tails in the T cell membrane. In Chapter 3 we deal with a phosphorylating enzyme, ZAP-70, and both phosphorylated and dephosphorylated tails in the T cell membrane. In both chapters we find our models have a better fit to the experimentally obtained data, or the computational simulations in Chapter 3, than the traditional ODE models with mean-field approximation. These models also help as a proof of concept of what is thought to describe the interaction between the proteins involved.

Chapter 4 takes on the anthropological problem, where we seek to find what set of physical properties allow each human to meet 150 people (known as the Dunbar number) within a year. Although not many hunter-gatherer societies remain in the world, this question has more of a historical value, aiding in theories about the present-day hierarchical structure of society, the purpose of rituals such as gift-giving, and the importance of human interaction for mental health. Through the theoretical development we gain insight into stochastic movement in a grid. Agent-based simulations are used to provide the number of humans met for a range of possible parameters, such that, given a society's basic details, the particular number of meetings per person could be looked up. We find that, in the simplest case, hunter-gatherers would be able

to meet with enough people during a year to maintain their community network in most scenarios. As is the nature of mathematics, the models here developed could in future be adapted for other biological or physical scenarios where elements engage similarly to the elements in each of our problems. After all, life is a mixture of disciplines to be described in the language of mathematics.

Contents

1	General Introduction	1
1.1	Contributions	8
2	Tethered Enzymatic Reactions: Dephosphorylation	11
2.1	Introduction	11
2.1.1	Biological background	14
2.1.2	From experiment to mathematical abstraction	19
2.1.3	Reaction scheme	22
2.1.4	Stochastic Simulation	25
2.2	Mean field approximation for protein dephosphorylation by SHP-1 . .	28
2.2.1	Interaction kernel $\sigma(r)$	28
2.2.2	Mean Field Approximation	31
2.3	Multi-centre Particle Density Model of Dephosphorylation	38
2.3.1	Pair Correlation Functions	40
2.3.2	Kirkwood Superposition Approximation	48
2.3.3	Solving the MPDPDE model	51
2.4	Comparison between the MPDPDE and stochastic simulations with periodic boundary conditions	57
2.5	Geometry effect: two- and three- dimensions	59
2.5.1	Solving the MPDPDE model in 2D	62
2.5.2	Comparing models	66
2.5.3	MAG3D vs MAG2D	72
2.6	Conclusions	75
3	Tethered Enzymatic Reactions: Phosphorylation	79
3.1	Introduction	79
3.1.1	Biological background	82
3.1.2	From experiment to mathematical abstraction	86
3.1.3	Reaction Scheme	88
3.1.4	Stochastic Simulation	90
3.2	Mean field approximation for protein phosphorylation by an enzyme .	91
3.2.1	Interaction kernel $\sigma(r)$	93
3.2.2	Solving the Mean Field Approximation	93
3.3	Multi-centre Particle Density Model of Phosphorylation (MPDPDE) .	95
3.3.1	Pair Correlation Functions	97
3.4	Solving the PDEs: 3D	104

3.4.1	Comparing models	110
3.5	Conclusions	111
4	Hunter-Gatherer Friendships: Mathematical model of meeting	115
4.1	Introduction	115
4.1.1	Biological background	116
4.1.2	Mathematical background	121
4.1.2.1	Gas model and applications	122
4.1.3	Computational background	122
4.1.4	Data on Hunter-Gatherers	126
4.1.5	Presented research	126
4.2	Mathematical models of meeting	127
4.2.1	Diffusive motion	128
4.2.1.1	Computational simulation of diffusion	129
4.2.1.2	Gambler's ruin	130
4.2.2	Immediate Meeting Model	139
4.2.2.1	Particle movement in two dimensions	142
4.2.3	Meeting and Forgetting as a Reaction Model	146
4.2.3.1	Continuous time and space: mean-field approximation	148
4.3	Comparing Models	150
4.4	Model Application	156
4.4.1	Immediate Meeting Model	157
4.4.1.1	Application of IMM to biomes	157
4.4.1.2	Application of IMM to society-specific data	166
4.4.2	Parameter Sweep of MFRM	174
4.5	Conclusions	187
4.6	Further work	190
4.6.1	Mathematical model of resource use and meeting	191
4.6.1.1	Preliminary investigation into a model including re- source use	193
4.6.1.2	Combining meeting and resource use	197
4.6.1.3	An example of high eating rate	199
5	Conclusions	203
A	Integral in Mean Field model in Chapter 2	213
B	Ode23 and ode15s MATLAB solvers for the MPDPDE of Chap- ter 3	215
C	Gillespie and Smoluchowski comparison	219
D	Table of data used for parameter values	221
E	Stochastic to deterministic equation of diffusion	227

F	One-dimensional case of the Immediate Meeting Model	231
F.1	Parameter estimation	231
F.2	Simulations in one dimension	233
F.3	One-dimensional results analysis	235
G	Parameter estimation for all the biome-specific cases of the Immediate Meeting Model	241
H	Results for BiomeMin, MiomeMid and BiomeMax cases for each biome, Immediate Meeting Model.	245
H.1	Arctic	245
H.2	Australia: subtropical desert through temperate settings	247
H.3	Subarctic and continental midlatitude forests of North America and Asia	249
H.4	Tropical and subtropical Africa	251
H.5	Tropical and subtropical South America	253
	Bibliography	255

List of Figures

2.1	Sections of two cells. The higher, red one, is an undesired cell that presents antigens on its surface and is to be detected. The lower one is a T cell with both inhibitory and antigen receptors. LAIR-1 is an inhibiting receptor. When LAIR-1 binds to an inhibitory external protein it gets phosphorylated. SHP-1 enzyme, which floats in the cell cytoplasm, binds to the phosphorylated tail of LAIR-1 (its ITIM) to become activated and inhibit other protein reactions, like the phosphorylation of the T cell antigen receptor. The T cell antigen receptor would otherwise go on to activate the immune reaction of the T cell.	17
2.2	Representation of what happens in the <i>in vitro</i> experiment, using the surface plasmon resonance (SPR) assay. The lower, blue, rectangular section represents the dextran matrix to which the unphosphorylated (blue squiggly line) and phosphorylated (blue squiggly line with a yellow spot on top) tails are tethered. A buffer solution with a known concentration of SHP-1 is injected over the matrix and flows over it in the direction shown. The SHP-1 phosphatase then binds and unbinds to phosphorylated tails from the solution. SHP-1's catalytic domain dephosphorylates phosphorylated tails, inhibiting them, either from the solution or when bound to another phosphorylated tail.	18
2.3	Data from the <i>in vitro</i> experiment using the surface plasmon resonance (SPR) assay. The fraction of phosphorylated tails bound by SHP-1 is shown with a 10 Hz time resolution. Starting time is as soon as the solution of SHP-1 is injected.	19
2.4	Representation of the biological structures as used for our mathematical models. The phosphorylated receptor tail is species <i>A</i> (shown in (a)), the compound of the SHP-1 enzyme bound to the phosphorylated tail is species <i>B</i> (shown in (b)), and the unphosphorylated receptor tail is species <i>C</i> (shown in (c)). Adding a couple of phosphate groups (coloured yellow) to <i>C</i> makes molecule <i>A</i> , enclosed in the green rectangle. Finally, the whole group of <i>A</i> plus the SHP-1 enzyme (coloured in orange), makes molecule <i>B</i> , enclosed in the orange rectangle.	20
2.5	Representation of the laboratory experiment in the surface plasmon resonance (SPR) assay, seen in Figure 2.2, but substituting with general species names: <i>A</i> , <i>B</i> , and <i>C</i>	21

2.6	Graphic explanation of the Reach length, L , associated with B due to it being anchored at one end, but free at the other end. The theory is that B can dephosphorylate any phosphorylated protein within the sphere of radius L (although only half a sphere is depicted here the B can move around the full sphere).	21
2.7	Example of experimental data obtained from the surface plasmon resonance (SPR) for the fraction of B compared to stochastic simulations. The blue dotted line is the experimental data and the purple dashed line is a fit using the stochastic model, explained in Section 2.1.4. Parameters used for the stochastic simulation are stated in each subfigure caption.	24
2.7	(Continued) Example of experimental data obtained from the surface plasmon resonance (SPR) for the fraction of B compared to stochastic simulations. The blue dotted line is the experimental data and the purple dashed line is a fit using the stochastic model, explained in Section 2.1.4. Parameters used for the stochastic simulation are stated in each subfigure caption.	25
2.8	Representation of a random-flight chain, or freely-jointed chain, with $n = 16$ segments. Each segment's length is a , but the root-mean-square distance from the first to the last point is equal to L_A	29
2.9	Visual explanation of how both chains' end points point to the same point in space although chain A starts at the origin and chain B starts at $\vec{r}$	31
2.10	Phase Plane of ODE (2.16) with variables $k_{\text{on}}^* = 0.0888 \text{ s}^{-1}$, $k_{\text{off}} = 1.1988 \text{ s}^{-1}$, $\lambda = 0.0001 \text{ s}^{-1}$, $k_{\text{cat}}^* = 0.0043 \text{ } \mu\text{M/s}$, $a_T = 58,470$ initial A proteins, and 0 initial B proteins. Red arrows over the line of flow are at time 0.1, 1, and 5 s.	37
2.11	Example of experimental data and stochastic simulations versus mean field approximation model. Shown is the fraction of SHP-1 bound to ITIM. Experimental data obtained from the surface plasmon resonance (SPR). Blue dotted line denotes the experimental data, purple dashed line is the average of 500 stochastic simulations, fitted as explained in Section 2.1.4, and the red dotted line is the solution of the mean field equations (Equations (2.16)) with a different type of fit in each subfigure. Parameter values used for the stochastic simulations: $k_{\text{on}} = 0.117 \text{ s}^{-1}$, $k_{\text{off}} = 1.27 \text{ s}^{-1}$, $k_{\text{cat}} = 0.00837 \text{ } \mu\text{M/s}$, $L = 21.9 \text{ nm}$, $\lambda = 0.0005 \text{ s}^{-1}$. These values were obtained from Goyette et al. [1]. Periodic boundary conditions on a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube with 58,470 ITIMs were used for the stochastic simulations.	38

2.12	Representation of how the particle density can be measured from a point $\vec{x}$, as explained in the text. Note that both the pink (lighter colour) and the purple (darker colour) lines in Figure 2.12b, start at 0, because no particles are found surrounding the particle at $\vec{x}$ for a small distance in (a). Then the pink graph starts oscillating because there is a spatial relation between pink dots and the particle at $\vec{x}$, whilst the purple dots have no spatial relation, so the graph is uniform in space.	41
2.13	An example of the numerical solution to Equations (2.54), (2.57), and (2.59) to (2.61) using Equations (2.6) and (2.63) with $\epsilon = 10^{-8}$. Shown are the solutions to n_A and n_B , and each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in blue, to the furthest $r/L = 4.5$ in magenta. Hence there are 90 solutions to each function. Numerically solved by MATLAB's ode23. Parameter values are: $k_{\text{on}}^* = 1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 1 \text{ }\mu\text{M/s}$, $\lambda = 0.001 \text{ s}^{-1}$, $L_A = L_B = 10 \text{ nm}$. The initial conditions are $n_A(0) = 1$, $n_B(0) = \textit{eps}$ (where <i>eps</i> is the smallest possible number different from 0 in the computer, to avoid singularities, which is 2.2204×10^{-16}), and $X_A(r, 0) = X_B(r, 0) = Y(r, 0) = 1$ (assuming there is no correlation to start with because of homogeneity).	56
2.14	Comparing our MPDPDE model to Dr. Isaacson's stochastic simulations with periodic boundary conditions. A total number of $N = 15000$ tethered proteins in a $300 \times 413.4864 \times 413.4864 \text{ nm}^3$ cube was used for the simulations, presented by the continuous orange line. The average of a 1000 simulations is used. Simulations and the MPDPDE solution in dashed magenta were produced with $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.01 \text{ }\mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.0005$. Inset: Zoom in to seconds 50 to 200, to show difference between models.	58
2.15	Difference between the two- and three- dimensional representations of the gel matrix for the laboratory experiments.	61
2.16	Visualisation of how the width (and length) of the domain changes for a constant number of particles N with a three-, two-dimensional treatment, and how our expression in Equation (2.80) goes from one to the other. Width w , length l , and height h are rescaled by proteins length, L	68

- 2.17 Comparing the MAG3D to Dr. Isaacson’s stochastic simulations with periodic boundary conditions. A total number of $N = 268574$ initial A molecules in a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube was used for the stochastic simulations, presented in a continuous blue line. The average of 10 simulations was used. The stochastic simulations and the MAG3D solution in dashed magenta were produced with $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$, whilst the MAG3D solution in dotted green was produced with the fitted parameters $k_{\text{on}} = 0.1023 \text{ s}^{-1}$, $k_{\text{off}} = 1.0388 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0203 \mu\text{Ms}^{-1}$, $L = 18.4154 \text{ nm}$, and $\lambda = 0.0054$, fitted by MATLAB’s `lsqcurvefit`. The initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$. Inset: Zoom in to seconds 50 to 200, to show difference between models. 70
- 2.18 Comparing MAG2D to Dr. Isaacson’s stochastic simulations with periodic boundary conditions (blue continuous line). A total number of $N = 114592$ initial A molecules in a $4000 \times 4000 \text{ nm}^2$ square was used for the stochastic simulations. The average of 100 simulations plotted. The stochastic simulations and the MAG2D solution in dashed magenta were produced with $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{M s}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$, whilst the MAG2D solution in dotted green was produced with the fitted parameters $k_{\text{on}} = 0.1020 \text{ s}^{-1}$, $k_{\text{off}} = 1.0254 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0189 \mu\text{M s}^{-1}$, $L = 17.4858 \text{ nm}$, and $\lambda = 0.0054$, fitted by MATLAB’s `lsqcurvefit`. The initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$. The dashed orange line corresponds to stochastic simulations with the same parameter values as the blue continuous line, but without periodic boundary conditions. Inset: Zoom in to seconds 50 to 200, to show difference between models. 71
- 2.19 Data from a test with $N = 243,507$ initial A molecules. Ratio of fitted parameter value over each parameter value actually used to produce the stochastic simulations (labelled as “real”) at heights $h = 0.1, 1, 10, 50, 100$, and 150 nm . Fitted two- and three-dimensional MPDPDE models using MATLAB’s `lsqcurvefit`. In all cases the initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$. MAG2D in blue continuous line and MAG3D in orange dashed line. Corresponding to data in Table 2.1. 73
- 3.1 ZAP-70 enzyme binds to the phosphorylated tail of the T cell antigen receptor (its ITAM) to become activated and phosphorylate other proteins, including other ITAMs (as seen in Figure 3.1a), to which other ZAP-70 enzymes may bind (as seen in Figure 3.1b). Further phosphorylation by ZAP-70 helps activate the T cell, and hence the immune response. 83

3.2	Representation of the <i>in vitro</i> experiments, using the surface plasmon resonance (SPR) assay to emulate what happens in the cell membrane, Figure 3.1. The experiment consists of a dextran matrix that is much wider, longer, and taller than the length of the proteins, as shown in panel (a). This supporting matrix has ITAMs tethered all over and kinases (ZAP-70) freely flowing through the dextran matrix. A transversal cut of panel (a) is shown in panel (b) for a clearer view of the reactions.	85
3.3	Abstract representation of the laboratory experiment in the surface plasmon resonance (SPR) assay, seen in Figure 3.2b, substituted with general species names: A , B , and C . The kinase is purple, the ITAMs are the blue lines, and phosphorylation is denoted by the yellow circle. 86	
3.4	Graphic explanation of the reactions hypothesized to be taking place in the SPR experiment.	88
3.5	Comparison between the stochastic model and the numerical solution of the Mean Field approximation, Equations (3.13). The parameter values used were: $k_{\text{on}}^* = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{M}^{-1}\text{s}^{-1}$, $\lambda = 0.005 \text{ s}^{-1}$. The stochastic simulation used 300,000 initial C molecules in a volume of $1000 \times 1000 \times 1000 \text{ nm}^3$, which translates to $c_T = 498.1734 \mu\text{M}$ in the Mean Field model. Length of tail used for the stochastic simulation was $L = 20 \text{ nm}$, the simulation had periodic boundary conditions on all four sides, and a 100 repetitions were averaged. Run time of stochastic simulations was 3.5 days. . . .	95
3.6	Example of numerical solution of three-dimensional MPDPDE model of Chapter 3, shown are Equations (3.55), (3.56), and (3.57). Solved using Matlab's ode23. Parameter values are: $k_{\text{on}}^* = 1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 1 \mu\text{M}/\text{s}$, $\lambda = 0.001 \text{ s}^{-1}$, $L_A = L_B = 10 \text{ nm}$. The initial conditions were $n_A(0) = \text{eps}$, $n_B(0) = \text{eps}$ (where eps is the smallest possible number different from 0 in the computer, to avoid singularities, which is 2.2204×10^{-16}), $n_C(0) = 1$, and $X_A(r, 0) = X_B(r, 0) = X_C(r, 0) = Y_{AB}(r, 0) = Y_{BC}(r, 0) = Y_{AC}(r, 0) = 1$ (assuming there is no correlation to start with because of homogeneity).	108
3.7	Pair correlation functions of three-dimensional MPDPDE model of Chapter 3, Equations (3.55) to (3.63). Shown are the solutions to each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in green, to the furthest $r/L = 4.5$ in black. Hence 90 solutions to each function. Numerically solved by MATLAB's ode23. Solved using Matlab's ode23. Same parameters and starting conditions as in Figure 3.6.	109
3.8	Three-dimensional representation of pair correlation functions of the three-dimensional MPDPDE model in Chapter 3, Equations (3.55) to (3.63). Same data points as in Figure 3.7 as well as parameters and starting conditions.	110

3.9	Comparing the MPDPDE model and the Mean Field model to the data produced by the stochastic simulation model. The same set of parameters were used for all models presented here, namely: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$. The stochastic simulations used 300,000 initial C particles in a $1000 \times 1000 \times 1000 \text{ nm}^3$ area with periodic boundaries. 100 simulations were averaged for this solution. The stiff (ode15s) solver in MATLAB was used to numerically solve the MPDPDE model.	111
4.1	Network layers for an average hunter-gatherer. See text for details. .	119
4.2	Biomes: regions of land with similar weather, flora and fauna conditions. Different biomes are represented by different colours. Image by Ville Koistinen, taken from: https://commons.wikimedia.org/wiki/File:Vegetation.png , reproduced as permitted by the CC-BY-SA 3.0 licence.	121
4.3	Schematic diagram of stochastic diffusion illustrating parameter and variable definitions.	130
4.4	Comparison of the solution to our particular case of discrete space diffusion, Equation (4.21), in dark blue dots and the average of simulating 10^6 particles in light blue bars. Solution for $L = 10$, and an arbitrary $d = 6 \text{ km/day}^2$. The independent variable, x_k , represents the k^{th} compartment, and the dependent variable, $P(x_k, t)$, is the probability for the particles to be in that compartment at the time t indicated at the top of each figure.	136
4.5	Example solution from the Gillespie SSA of the model described in Section 4.2.2, for 2D. Average of 100 runs. Parameter values used were: 247 unmet families initially, forgetting time of a day, length of $71 \times 71 \text{ km}^2$, detection distance of 1 km, so each compartment is of 1 km^2 , and diffusion coefficient of $4.7 \text{ km}^2/\text{day}$. As this was taken from the North America: California and Northern Mexico biome, as seen in Section 4.4.1.1, the parameter values used here were the ones used for the BiomeMid case.	145
4.6	Density of families that were not yet met by the tracked family at different time points for three different models. A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200 \text{ km}$, with diffusion constant, D , of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are the numerical solutions of the PDE Equation (4.41) for U and the stochastic simulations that used the IMM and MFRM models. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MFRM simulations a forgetting rate of $f = 1$ per day was used, although $f = 0$ was used for the numerical solution of Equation (4.41). Results shown are an average of 500 simulations each, where applicable.	152

4.7	Density of tracked family at different time points for three different models. A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200$ km, with diffusion constant D of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are the numerical solutions of the PDEs Equations (4.40) for Y and the stochastic simulations that used the IMM and MFRM models. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MRM a forgetting time of one day is used, for the MFRM a forgetting rate of $f = 1$ per day was used. Results shown are an average of 500 simulations each, where applicable.	153
4.8	A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200$ km, with diffusion constant, D , of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are stochastic simulations of the IMM and MFRM models. The average of 500 simulations for each are shown in blue, whilst their respective standard deviations are shown in red and magenta. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MFRM simulations a forgetting rate of $f = 1$ per day was used.	154
4.9	The number of newly met families for: the numerical solutions of the PDEs, Equations (4.40) and (4.41); the IMM stochastic; and the MFRM stochastic simulations. Parameter values as in Figure 4.6. The stochastic simulations (blue, cyan) were only run until day 365 because that is the range of interest. The numerical solution of the PDEs (red) is not as computationally intensive to run, so can be run for a longer end time. Therefore, it is shown until it reaches the maximum (30 families in the domain) to show this limit is reached. Results shown are an average of 500 simulations each, where applicable.	155
4.10	Results for the two-dimensional IMM for data from hunter-gatherers in California and Northern Mexico. (a) The number of families met by Y as they were being met throughout the year. (b) The number of times each of the other families was met by family Y on average. BiomeMid parameter values were used, with a domain of 6006 km^2 with 245 “families” of 5 individuals each, an encounter distance of 1 km, and a diffusion coefficient of $9 \text{ km}^2/\text{day}$. Simulation of a year, repeated 2000 times.	163

4.11	Number of new families met over the year and number of times each family was met for the BiomeMin, BiomeMid, and BiomeMax cases for the Arctic biome. This is the case for the two-dimensional model with an average of 4.4 people per family and an encounter distance of 1 km. For the BiomeMin case we had 25 initially unmet families in an area of 370,881 km ² , and a diffusion coefficient of 9 km ² /day, with 100 simulations. For the BiomeMid case we had 318 initially unmet families in an area of 81,225 km ² , and a diffusion coefficient of 25.8 km ² /day, with 100 simulations. For the BiomeMax case we had 1,595 initially unmet families in an area of 18,769 km ² , and a diffusion coefficient of 132.3 km ² /day, with 26 simulations. All parameter values obtained as explained in Appendix G.	164
4.12	Close up of Figure 4.11e. The orange vertical line (furthest to the left) indicates when the 15-individual layer is met, the brown vertical line (middle one) for the 50-individual layer, and the red vertical line (furthest to the right) for the 150-individual layer.	164
4.13	Average number of people met (left) and times most frequently met family was met (right) by each biome, for the BiomeMin, BiomeMid, and BiomeMax cases, as explained in Section 4.4.1.1. The BiomeMin case can not be seen because of how small it is compared to the BiomeMax case. BiomeMid is shown in pink (lower bars) and BiomeMax is shown in purple (taller bars). Individual results for each society in the Arctic and tropical and subtropical South America biomes are shown in their respective biomes as asterisk marks so they can be compared, although the names of the societies are not given. That is, all the data points marked as asterisks above the Arctic biome label are results for the simulation of each society in the Arctic biome individually. This makes the comparison between the approach in Section 4.4.1.1 (the coloured columns) and the approach in Section 4.4.1.2 (the asterisks) easier to see.	168
4.14	Average number of people met (left) and times each family was met (right) by the Labrador Inuit, in the Arctic, over a year, in 100 simulations. This society had 255 unmet families in a 52500 km ² domain with an average family of 5.7 individuals and a speed of 33.75 km/day.	169
4.15	Average number of people met (left) and times each family was met (right) by the Caribou Inuit, in the Arctic, over a year, in 100 simulations. This society had 160 unmet families in a 236500 km ² domain with an average family of 4.4 individuals and a speed of 27.5 km/day.	169
4.16	Average number of people met (left) and times each family was met (right) by the Tehuelche, in Argentina, tropical and subtropical South America, over a year, in 100 simulations. This society had 1799 unmet families in a 475300 km ² domain with an average family of 5 individuals and a speed of 21.18 km/day.	172

4.17	Average number of people met (left) and times each family was met (right) by the Ona, in Argentina, tropical and subtropical South America, over a year, in 100 simulations. This society had 698 unmet families in a 48100 km ² domain with an average family of 5 individuals and a speed of 13.33 km/day.	172
4.18	Time taken to meet 150 people, given different: areas in which families move, speeds at which they move, and the number of families in total in the area, N_0 . Because the domains were taken as squares, only the length of one side of the domain is given. It can be seen that larger length increases the time to meet 150 people, whereas the increase in people in the area (across subfigures), or the increase in speed (lower lines in each graph), both decrease the time taken to meet the hierarchical level.	178
4.19	Log-log plot of Figure 4.18. Time taken to meet 150 people, given different: areas in which families move, speeds at which they move, and the number of families in total in the area, N_0 . Because the domains were taken as squares, only the length of one side of the domain is given. It can be seen that larger length increases the time to meet 150 people, whereas the increase in people in the area (across subfigures), or the increase in speed (lower lines in each graph), both decrease the time taken to meet the hierarchical level. These graphs show a power relationship.	179
4.20	Gradients from linear regression results corresponding to Figure 4.19a (lower line in light blue with squares), Figure 4.19b (dark blue line with crosses), Figure 4.19c (purple line with plus signs), and Figure 4.19d (top line in magenta with diamond markers). The gradient is the power term of the power law found in Figure 4.19.	180
4.21	Intercept with y-axis when doing linear regression to Figure 4.19a (top line in light blue with squares), Figure 4.19b (dark blue line with crosses), Figure 4.19c (purple line with plus signs), and Figure 4.19d (magenta line with diamond markers). The intercept is the constant term multiplying the powered independent variable of the power law.	181
4.22	Gradients from linear regression results similar to Figure 4.20 but corresponding to the time taken to meet the 15, 50, 500, and the 1500 social layers.	182
4.23	Intersect with Y-axis when carrying out linear regression similar to Figure 4.20 but corresponding to the time taken to meet the 15, 50, 500, and the 1500 social layers.	183
4.24	Out of the parameter values tested, those which hit the 150 level of number of people met throughout a year are marked as indicated. N_0 refers to the number of ‘other’ families in the domain. Meeting one family would mean meeting 5 people.	184

4.25	Out of the parameter values tested, those which, within the year, hit the number of people indicated below the subfigure are marked with a symbol accordingly.	185
4.26	Ranked families from most met to least. We obtain the number of times each family was met by family Y within a year and then sort them from most met to least, then average over all 100 simulations. (a) Minimum in all parameter values: 2 other families, apart from the main one, and movement at a speed of 5 km/day in a 9×9 km ² domain. (b) Maximum in all parameter values: 299 other families, apart from the main one, each moving at the speed of 45 km/day in a 799×7999 km ² domain. In each simulation Y is initialised in the middle of the domain and all other families are placed uniformly randomly throughout the domain.	186
4.27	Examining the plateau observed in the ranking of families, such as in Figure 4.26b. Varying the number of simulations for the mean of ranked families. 150 other families were used, with 399 km per side of the domain, and a speed of 40 km/day.	187
4.28	Numerical solution to Equations (4.56) (left) and (4.57) (right). Hunter-gatherers diffusing and eating in a one-dimensional space of length 200 km, diffusing at 25 km ² /day, with an eating rate of 50 resources per day, a growth rate of 1 per day, an initial maximum food level of $R_{max} = 5000$ units in every compartment and y initially as a step function of area 1 in the centre of the domain.	196
4.29	Numerical solution to Equations (4.56) (left) and (4.57) (right) at time $t = 0.01$ days. Same parameter values as in Figure 4.28.	196
4.30	Numerical solution to Equations (4.62) to (4.65) for y (left), u (centre), and r (right) at time $t = 0.01$ days. One Y and 30 unmet families diffusing, eating, and meeting in one-dimensional space of length 200 km, diffusing at 25 km ² /day, with an eating rate of 50 per day, a growth rate of 1 per day, a meeting rate of $\mu = 1$ per day, and an initial maximum food level of $R_{max} = 5000$	199
4.31	Average from 10,000 stochastic simulations. A family Y and 30 other initially unmet families diffusing in a one-dimensional domain of length $L = 200$ km, meeting and eating with rates $\mu = E = 1$ per day, forgetting each other every 24 hours, a detection distance of 1 km, a maximum resource of $R_{max} = 1$ unit, and a regrowth rate of $G = 1/365$ units per day. Each graph has the day stated at the top of the graph. All events treated as reactions except the forgetting event, which happens a day after a meeting occurs.	200
4.32	Number of new families Y meets when all families have a very high eating rate. Average from 10,000 stochastic simulations. All parameter values as in Figure 4.31.	201

B.1	Pair correlation functions of three-dimensional MPDPDE model of Chapter 3, Equations (3.55) to (3.63). Shown are the solutions to each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in blue, to the furthest $r/L = 4.5$ in red. Hence 90 solutions to each function. Numerically solved by MATLAB's ode23. Parameters used are a peptide concentration of 3×10^{-4} peptides/nm ³ (equivalent to a height and width of 1,000 nm with 300,000 peptides), $k_{on}^* = 0.1 \text{ s}^{-1}$, $k_{off} = 1 \text{ s}^{-1}$, $k_{cat} = 0.02 \mu\text{M}^{-1}\text{s}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$	216
B.2	Time taken for numerical solver, MATLAB's "ode23", to solve our MPDPDE system of equations up to a given 'Experimental Time'. $p_1 = 0.100000$, $p_2 = 9.961813$, $p_3 = 4.150755$, $p_4 = 0.005000$, $C_T = 0.0003$ peptides per nm ³ , and $k_{off} = 1$ per s.	217
B.3	Comparing the stiff (ode15s) and non-stiff (ode23) ODE solvers in MATLAB with the data produced by the stochastic simulations from Chapter 3. The blue line with asterisks is the difference in A concentration between non-stiff solver and data from stochastic simulations, whilst the red line with squares is the difference between the stiff solver and the same data from stochastic simulations. We note these are very similar. This is confirmed by plotting the difference between the stiff and non-stiff solutions, shown by the yellow line with asterisks, which seems flat (at zero) on the left. The difference can only be appreciated when we zoom in (on the right). Clearly, the solutions generated by the ODE solvers are not too different from each other, but ode15s managed to solve the equations much faster.	218
C.1	An example of the visualization of the <i>Smoldyn</i> dynamics at two points in time. Family Y in blue, unmet families in green and met families in red. Done with 245 initial unmet families, for 365 days, with a forgetting time of 1 day, a detection distance of 1 km, a diffusion constant of 9 km ² /day, and a domain of $78 \times 77 \text{ km}^2$	219
C.2	Number of families that family Y has met over time, only counting the first time they were met. Average of 2000 runs with the Gillespie SSA compared to two particular runs done with the Smoluchowski method using <i>Smoldyn</i> . The mean of the Gillespie simulations is shown in green, whilst its first standard deviations have been added above and below in a dashed magenta line. Done in two dimensions with 245 initial unmet families, for 365 days, with a forgetting time of 1 day, a detection distance of 1 km, a diffusion constant of 9 km ² /day, and a domain of $78 \times 77 \text{ km}^2$	220

F.1	Example solution from the Gillespie SSA for 1D for the model described in 4.2.2, IMM, average of 2,000 runs. Parameter values used were: 54 unmet families initially, forgetting time of a day, length of 400 km, detection distance of 0.5 km, and diffusion coefficient of 9 km ² /day.	235
F.2	The dotted green line represents the mean number of different families the family <i>Y</i> has met throughout the year. The dashed magenta lines represent ± 1 standard deviation. The solid blue line is the power-law approximation obtained using MATLAB's curve-fitting package. Parameter values are 54 families in a 400 km domain with a speed of 5.3 km/day, and averaged from 2,000 realisations.	236
F.3	Number of new families that family <i>Y</i> meets each day as an average over 2,000 realisations. These are families family <i>Y</i> had not met before in the year and so represents the number of uniquely met families each day. The solid blue line is the approximation obtained using MATLAB's curve-fitting package. Parameter values are 54 families in a 400 km domain with a speed of 5.3 km/day.	237
F.4	Ordered by average number of times they meet, we see here how many times family <i>Y</i> meets with other families throughout the year. Parameter values are 54 families in a 400 km domain, and 2,000 realisations.	239
H.1	Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case had 1 meeting, BiomeMid case had 116 meetings, and BiomeMax case had 10,021 meetings in average by the end of the year.	246
H.2	Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case met 0.43 new families, BiomeMid case met 63.69 new families, and BiomeMax case met 1,581.3 new families in average by the end of the year.	246
H.3	Number times each family was met by family <i>Y</i> over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. The most met family was met 0.53 times in the BiomeMin case, 6.52 times in the BiomeMid case, and 20.42 times in the BiomeMax case in average by the end of the year.	247
H.4	Example of 9 paths for each of the three cases for the Arctic biome. See text for parameter values.	247
H.5	Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 91 meetings, and BiomeMax case had 190,158 meetings in average by the end of the year.	248

H.6	Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for Australia. See text for parameter values. BiomeMin case met 0.40 new families, BiomeMid case met 43.64 new families, and BiomeMax case met 717 new families in average by the end of the year.	248
H.7	Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for Australia. See text for parameter values. The most met family was met 0.10 times in the BiomeMin case, 6.71 times in the BiomeMid case, and 279.38 times in the BiomeMax case in average by the end of the year.	249
H.8	Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 70 meetings, and BiomeMax case had 10,668 meetings in average by the end of the year.	250
H.9	Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. BiomeMin case met 0.15 new families, BiomeMid case met 39.64 new families, and BiomeMax case met 1,065.4 new families in average by the end of the year.	250
H.10	Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. The most met family was met 0.28 times in the BiomeMin case, 5.58 times in the BiomeMid case, and 25.67 times in the BiomeMax case in average by the end of the year.	251
H.11	Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the tropical and subtropical Africa. See text for parameter values. BiomeMin case had 3 meetings, BiomeMid case had 778 meetings, and BiomeMax case had 7,500 meetings in average by the end of the year.	252
H.12	Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical Africa. See text for parameter values. BiomeMin case met 1.29 new families, BiomeMid case met 89.89 new families, and BiomeMax case met 180 new families in average by the end of the year.	252
H.13	Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical Africa. See text for parameter values. The most met family was met 1.81 times in the BiomeMin case, 19.07 times in the BiomeMid case, and 59.97 times in the BiomeMax case in average by the end of the year.	253

H.14	Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the tropical and subtropical South America. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 329 meetings, and BiomeMax case had 234,720 meetings in average by the end of the year.	254
H.15	Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical South America. See text for parameter values. BiomeMin case met 0 new families, BiomeMid case met 165.48 new families, and BiomeMax case met 1,799 new families in average by the end of the year.	254
H.16	Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical South America. See text for parameter values. The most met family was met 0 times in the BiomeMin case, 8.48 times in the BiomeMid case, and 156.11 times in the BiomeMax case in average by the end of the year.	255

List of Tables

2.1	Fitting MAG2D and MAG3D for simulations with varying heights. Maximum error highlighted.	74
4.1	The assumed equivalences used in this project for social organization in hunter-gatherers as given in Pearce [2].	120
4.2	Definition of symbols/abbreviations used throughout this chapter. . .	142
4.3	Results for the BiomeMin sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged. The observed family did not meet any of its network layers in a year in any of the tested biomes.	159

4.4	Results for the BiomeMid sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; the number of days taken for family <i>Y</i> to meet its 15-, 50-, 150-, and 500-individual layer; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged.	160
4.5	Results for the BiomeMax sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; the number of days taken for family <i>Y</i> to meet its 15-, 50-, 150-, and 500-individual layer; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged.	161
4.6	Number of distinct people met by family <i>Y</i> at the end of the year and maximum number of meetings of a single family in each society in the Arctic biome. Both averages over 100 simulations.	167
4.7	Number of total new people met by family <i>Y</i> at the end of the year and number of times the family most met was met in each society in the tropical and subtropical South America biome.	170
4.8	Ranges of parameter values used in the computer simulations for the parameter sweep.	174
D.1	An extract of the data used from Grove et al. [3, Table 5.01]	221

Chapter 1

General Introduction

Much of Mathematical Biology uses deterministic descriptions to model the world. However, this approach is not always appropriate. Biology is full of systems where interactions between a small number of elements give rise to stochastic phenomena [4, 5, 6]. This means that systems with the same initial conditions can tend to different outcomes, as well as a small change in the starting conditions or a random fluctuation possibly causing large variations in the evolution of the system. Interest in systems with stochastic phenomena has been on the rise in recent years¹. There are no absolute rules on what constitutes small numbers, but they become important when the discrete nature of the elements in the system becomes significant, causing a deterministic model to fail to accurately describe the system. In this thesis we use stochastic techniques, alongside modelling that takes into account an element's position, to model different problems in biology.

Mathematical Biology uses deterministic descriptions because of the standardized development of equations as seen, for example, in Baker [7, Table 1], analytical

¹A search in Google Scholar of the terms “stochasticity, low, small, numbers” produces 349 results in the 4-year period from 2017 to 2020, and the number of results steadily decays going back in time, down to 91 results in the period 2001 – 2004

tractability, and because it is most likely to be numerically solvable without using much computer power. For example, continuum differential equations [8] can be used to model the average of many entities with a small number of equations. This modelling framework is used right across the field, from the modelling of circadian rhythms [9, 10], to ecology with predator-prey systems [11, 12], and carbon and nitrogen transport processes in water due to biofilms [13], to name but a few examples. However, when the entities are only present in low quantities it becomes necessary to resort to stochastic techniques to assess the influence of noise, as the usual deterministic description will not accurately represent the behaviour of the system [14]. Some examples of stochasticity in models are regulatory protein competition for a switch point [15], circadian rhythms [16] and in the cell cycle [17], again just naming a few.

Through the theory of stochastic processes [18, 19] we associate a continuous-time modelling of stochastic events, that depend solely on the previous state of the system and that accurately describes the behaviour of different species in the system, with an underlying master equation, which is a set of ordinary differential equations (ODEs). This master equation describes the evolution of the probability density of all the possible states of the system at each time point. But the master equation can be very nonlinear and have very high dimension so it quickly becomes intractable in practice. There are very few simple systems for which an analytical solution exists [20]. So, to overcome this, we often turn to stochastic simulations, such as the stochastic simulation algorithm (SSA) developed by Gillespie [21], which we use throughout this thesis and firstly introduce in Section 2.1.4. This kind of

simulation algorithm requires us to simulate all of the elements and reactions in the system, so it can be computationally intensive when population numbers grow. This is why exact algorithms, equivalent to the Gillespie algorithm, have been created, although the modelled system needs to be suitable for the algorithm. For example, we use the Gibson–Bruck Next Reaction Method [22] in Chapters 2 and 3 because it is generally thought that this is the best method to speed up most simulations using the Gillespie SSA [23], and the techniques used in this method are helpful in making the stochastic simulations faster. Another example is the Optimised Direct Method [23], which requires less computing time and memory than the Gillespie SSA, but where determining the next reaction to occur is slightly more complicated [24]. These algorithms can be further sped up by using parallel computing [25, 26].

Besides the importance of low numbers, we are going to be considering phenomena that are spatially dependent. Spatial dependence is important in areas such as ecology [27], where modelling of specific ecological problems needs to account for the dispersal of animals, patchiness of land, degree of aggregation, and distances between components. Examples of this include the factors influencing habitat suitability for sparrows [28], critical patch size for population density and dynamics [29], and the spatial spread of populations [30, 31]. Whereas some models account for space implicitly through variation in parameters, see, for example, Roughgarden [32], we will focus on a more explicit dependence on space because of its importance to our problems.

Stochasticity and spatial location are two distinct problems, but there are times when taking both into consideration is important for the modelling of a system.

We have come across three such problems at different ends of the spectrum: two regarding protein interaction in immunology and one regarding human interaction in anthropology.

Immunology is the study of the complex interaction of cells, tissues, organs, and substances that defend the body against infections and other diseases. It has made possible the study of cancer, vaccinations, organ transplantation, autoimmune diseases, and the behaviour of antibodies, amongst others [33, 34]. But the number of molecules involved in the immune system, and the amount of possible reaction pathways, is so large that the study can be approached from different frameworks. For example, multiscale modelling can be found in the cardiovascular field [35], neural systems [36, 37], tumour growth [38], vasculature formation [39], epidermal wound healing [40, 41, 42] and tissue engineering [43, 44].

On the microscopic level, we have biophysical and biochemical events [44, Figure 2], where molecules interact with each other. Protection from pathogens requires detection of foreign substances and, amongst other things, activation or inhibition of cells in the immune system [45]. These cells, such as white blood cells, obtain their cues from external signals that begin a signal transduction cascade inside the cell. Phosphorylation and dephosphorylation play important roles in such intracellular signalling and in switching the enzymatic activity on and off [46], aiding in the positive, or negative, feedback that leads to the fine control of the signalling intensity [45]. For example, phosphorylation and dephosphorylation of proteins can change the structural conformation of the proteins to allow for easier connection with other signalling proteins [47, 48]. Chapters 2 and 3 will focus on the modelling

and understanding of the dephosphorylation and phosphorylation, respectively, of different trans-membrane proteins, in an attempt to further the understanding of signal transduction and the cross-talk between signal pathways. After all, as stated by Enjalbert and Le Pechon-Vallee [49]: “Because phosphorylation and dephosphorylation play an essential role in almost every regulatable cellular process, any abnormality in protein kinase activity is likely to be a feature in human pathology.”

In Chapters 2 and 3 we study tethered signalling reactions (interaction between proteins that take part in the signalling pathway of the T cell, where at least one of the components is bound to a substrate), including binding and catalytic reactions, and molecular reach, which is the possible interaction between two tethered molecules depending on their length, flexibility, and the distance between them. Although binding and catalytic reactions have been well studied [50, 51, 52, 53], we include molecular reach, as its influence on other molecules is less understood [54].

We currently do not have a way of thinking about space in these problems, so we will develop a framework in which we can address these problems. The hope is that this work will contribute towards the ongoing effort to understand T cell regulatory mechanisms, a study that has been growing rapidly in the last few years because of the recognition of its importance and advancements in experimental technology, such as cell imaging [45].

Understanding the T cell regulatory system and the way in which phosphorylation and dephosphorylation affect surrounding proteins could, in future, be used for immune system treatments. Because dephosphorylation and phosphorylation are present in so many protein-protein interactions and dictate T cell activation, these

treatments could range from vaccine development to therapies for autoimmune diseases and cancer. But, furthermore, because the work is of a basic scientific nature, many more important questions and research paths in immunology could be found by shining a light on our ignorance. This theme has application and further research in mind.

Our second theme stems from anthropology. Anthropology is the study of human behaviour throughout time, ranging from human biology all the way through to its societies and culture [55]. It is a large field that has branched out into specialized fields, like physical anthropology, cultural anthropology (ethnology), social anthropology, linguistic anthropology, and psychological anthropology. Currently, anthropology is being applied to problems such as understanding global health [56]. Specifically, understanding the role of loneliness on mental health is an issue of great importance, particularly during periods of social isolation, which are presently being forced upon us by COVID-19.

Friends help with external threats, as well as internal ones, support us in dealing with stressful situations, provide help practically and economically when needed, and make us feel happier, amongst other advantages. As stated in Dunbar [57], “Friendship is the single most important factor influencing our health, well-being, and happiness”. This is also true for anthropoid primates like baboons [58, 59, 60, 61] and even wild horses and zebras [62, 63], as we share being social animals. Yet, group sizes in humans are larger than, say, similar apes. This is, in part, because group size correlates with brain size in primates [64, 65], where group size is identified as the number of people with whom a person has a relationship by personal contact

[65]. Human group size has been almost the same since we were hunter-gatherers. Dunbar's work [65] suggests that humans maintain interaction with around 150 people, the so-called "Dunbar number". Seeing as social bonds may have a strong influence on the longevity and quality of life, it strikes us as important to understand the processes that gave rise to the Dunbar number and the evolutionary origins of our societal dynamics. With this idea in mind, in Chapter 4 we explore how this number may emerge from a study of interactions between hunter-gatherers given different conditions of population size, area of roaming, and walking speed. This research hopes to help answer questions related to human bonding over different conditions, like the appearance of gift-giving when conditions were too harsh for our ancestors to meet regularly, and the importance of face-to-face interaction for our mental health.

Further, more detailed, information on each topic can be found in each chapter, as each chapter begins with an introductory section on the literature of the relevant theme.

In both the immunological and anthropological themes, stochasticity and location are important. In immunology, the number of interacting proteins is huge, but because of their tethered state they are restricted to effectively interacting with just a few other proteins, making stochasticity important for the calculation of some reaction rates. The anchored condition and specific length of each protein also mean that the distance between them is important for their interaction. On the other hand, hunter-gatherer societies have hundreds of individuals roaming around, but we concern ourselves with following one individual and their interactions with ev-

ery other individual or group. This means that, effectively, we are dealing with “one-against-one”, and thus stochastic effects become important. Similarly, because hunter-gatherers would only interact when close to each other, their location needs to be considered. Even today, with the existence of technological advances, letters, email and video-calls, we only have 150 close contacts [65, 66]. If we subscribe to Dunbar’s idea, we can see that these tools can help us maintain the 150 friends, but they do not extend this number.

As we can see, the themes of stochasticity and location run across a large range of spatial scales, from interactions between proteins in the cell, to the maintenance of societies. We show, by looking at problems from diverse biological backgrounds, that these seemingly very different examples all share underlying common themes, and we illustrate the importance of addressing them pertinently for understanding and prediction. We draw from mathematical and computational techniques to better understand our bodies and friendships.

1.1 Contributions

For Chapter 2: Jesse Goyette and Omer Dushek designed the research. Jesse Goyette, Nicola Coker-Gordon and Marcus Bridge performed the laboratory experiments with which the models were compared. Samuel A. Isaacson and Jun Allard developed the stochastic simulations. Citlali Solís Salas and Omer Dushek developed the analytical and numerical models. Citlali Solís Salas developed the numerical solvers, fitted the data, produced the graphs, and wrote the chapter. Philip Maini, Thomas Woolley and Omer Dushek read and commented on the written

work.

For Chapter 3: Omer Dushek and Citlali Solís Salas designed the research. Samuel A. Isaacson and Jun Allard developed the stochastic simulations. Citlali Solís Salas and Omer Dushek developed the analytical and numerical models. Citlali Solís Salas developed the numerical solvers, fitted the data, produced the graphs, and wrote the chapter. Philip Maini, Thomas Woolley and Omer Dushek read and commented on the written work.

For Chapter 4: Thomas Woolley, Philip Maini, Robin Dunbar, Eiluned Pearce and Citlali Solís Salas designed the research. Robin Dunbar and Eiluned Pearce provided the hunter-gatherer data on which the work is based. Citlali Solís Salas developed the stochastic simulations. Thomas Woolley and Citlali Solís Salas developed the analytical model. Citlali Solís Salas produced the data and graphs, and wrote the chapter. Philip Maini, Thomas Woolley and Citlali Solís Salas analysed the results. Philip Maini, Thomas Woolley and Robin Dunbar read and commented on the written work.

Chapter 2

Tethered Enzymatic Reactions: Dephosphorylation

2.1 Introduction

The work in this chapter, up to Section 2.4, has been published in the Appendix section of the paper Goyette, Solís Salas, et al. [1].

Our body's immune system is composed of cells and proteins dedicated to identify, restrain, and ultimately destroy, pathogens. T cells form a key part of the adaptive immune system and are able to recognise threats through their cell surface receptors. Understanding how these surface receptors work is important because they control the activation of T cells, which, in turn, control infections. Our current understanding of these receptors is already being exploited for cancer therapy and improvements in this understanding may improve therapies. It is also important to understand the mechanisms that “turn” the T cells “on and off”, as an over-reaction of the immune system may lead to autoimmune diseases and allergies, whereas an under-reaction may open the body to a huge amount of pathogens, which is what occurs in HIV, for example.

An example of an important surface receptor on T cells is LAIR-1, whose cytoplasmic tail can recruit an enzyme called SHP-1. When SHP-1 is bound to LAIR-1 at the cell surface, it can act on substrates that it can reach by dephosphorylating them. This is how the receptor LAIR-1 is thought to transmit signals and help reduce the T cell reaction. However, although the chemistry of this process can be verbally described, there has yet to be any quantitative analysis of this reaction. We hypothesise that the tethered property of LAIR-1 helps stop T cell activation, and with this work we seek to verify that LAIR-1 and SHP-1 interact as thought, as well as quantify their reaction rates.

The work in this chapter is motivated by experiments that have been performed to quantitatively assess the binding and catalytic interactions between SHP-1 and LAIR-1. The technique used is explained in Section 2.1.1. The data from these experiments were sufficiently precise to allow fitting of mathematical models. This is important because obtaining a suitable mathematical model means that more information could be obtained from the experiment, fewer experiments would be necessary, and possibly more could be learnt from other such experiments.

Once we tested a range of deterministic models it became apparent that we were in need of a faster way to verify the models and their representation of the physical phenomena we were attempting to model, because we had only a few laboratory experiments and producing more was timely and costly. We hence constructed a computational stochastic model (Section 2.1.4) which encompassed recreating the physical phenomena we considered to be happening in the experimental assay. Once the similarity of the computational, or *in silico*, data to the experimental, or *in vitro*,

data was shown we could use the *in silico* data to test and refine our deterministic model, as seen from Section 2.2.2 onwards. The fitting and comparison of our deterministic model against *in silico* data, where the parameters are known, gave us a good idea of the closeness of the model's predictions to the real values of the parameters, as seen in Section 2.4.

Although the computational stochastic model tests similar concepts to our deterministic model, the deterministic model proves to be much faster, which is why it is preferred. The deterministic model can then be used to inform the laboratory experiments on parameters that could yield interesting behaviour, saving both time and money.

In Section 2.1.1 we will describe the relevant biological interactions taking place in the cell, from which we abstract the mathematically relevant components in Section 2.1.3. In Section 2.1.4 we present the computational stochastic model with which we corroborate our deterministic model. Section 2.2 shows a simpler deterministic model, a coupled system of mean-field ordinary differential equations (ODEs), that do not capture the experimental behaviour appropriately. We therefore then derive a model that consists of a coupled system of non-linear partial differential equations that better describe the reaction scheme in Section 2.3, making use of the stochasticity of the system and taking location into account. We name this the MPDPDE model. We solve it numerically in Section 2.3.3. In Section 2.4 we compare our MPDPDE model with the computational stochastic model. We then realised some of the differences between the MPDPDE model and the stochastic simulations could be due to boundary conditions so, finally, in Section 2.5 we

adapt the MPDPDE model (which was previously solved in three dimensions) to two dimensions. We compare the two- and three-dimensional MPDPDE model to the stochastic model in Section 2.5.2, and then we test the suitability of our two- and three-dimensional solutions for a range of geometries going from a two-dimensional plate to a three-dimensional box to check when each model is best for each geometry, and at which height of box does this switch occur.

Goyette, Solís Salas, et al. [1] were able to adequately fit the MPDPDE model developed here to the experimental data available (seen in Figure 2.3). Using a statistical fit, they were able to isolate a specific parameter set to use. This parameter set is also used throughout this chapter to compare with other models and with MATLAB’s nonlinear least-squares solver. When referring to the “parameters obtained from Goyette et al. [1]”, this is what we refer to.

In the end, our aim is to understand SHP-1’s influence on the immune response and gain a better knowledge of how it works. In future, we could use this knowledge to improve our use of drugs, enhancing or stopping SHP-1 function when tackling illnesses such as cancer or auto-immune diseases.

2.1.1 Biological background

Immune cells detect their environment using proteins on the surface of their membrane, called receptors. As seen in Figure 2.1, these receptors have sections towards the outside of the cell and towards the inside. The outside section binds to other proteins and recognises them. The section inside, in the cytoplasm, is called the ‘tail’. The T cell is a specific example of an immune cell, containing many differ-

ent receptors on its surface. T cells can recognise components of pathogens when specific antigen receptors bind these pathogen components, inducing a large signal transduction pathway, which ultimately can lead to T cell activation. Activated T cells can perform various effector functions, such as the secretion of cytokines, that can mobilise a global immune response. On the other hand, the T cell has other receptors that can inhibit its activation, so that the system does not overreact. For example, LAIR-1 is one such receptor. Overall, the combination of activating and inhibiting receptors determines whether or not the T cell will become activated.

If we look closely at the reaction occurring inside the T cell, we will find receptor tails that become phosphorylated (activated) when an outside protein matches the specific receptor. Because the phosphorylation and dephosphorylation of these signalling tails determines whether or not a T cell will become activated, there is interest in understanding how receptor tails interact with enzymes, such as kinases and phosphatases, that catalyse these reactions. Phosphatase enzymes accelerate the dephosphorylation of phosphorylated proteins that they can reach nearby. The interaction between the receptor tails and corresponding enzymes will depend on the binding kinetics of the enzyme to the receptor tail, the catalytic rate of the enzyme acting on the substrate, and on the ability of the enzyme to reach the substrate [51, 67].

We look in particular at the interaction between a specific cytoplasmic tail that contains an immunoreceptor tyrosine-based inhibitory motif (ITIM) that serves as a binding site for the enzyme SHP-1. Inhibitory receptors on immune cells contain ITIMs in their cytoplasmic tails and, therefore, SHP-1 can bind to many different

inhibitory receptors when they are engaged with their ligands outside the cell. SHP-1 has three domains, two of which can bind to the phosphorylated ITIM and will leave SHP-1 “activated” [8]. The final domain in SHP-1 is the catalytic domain that dephosphorylates other proteins. An example of this pathway can be seen in Figure 2.1. It is known that SHP-1 reduces immune cell activation [68] by dephosphorylating other proteins in the surrounding area. This dephosphorylation would inhibit the signalling cascades that promote activation of the cell [69]. In other words, SHP-1 is thought to reduce phosphorylation at the cell membrane that is required for immune cell activation.

Although the function of SHP-1 relies on it binding (tethering) to an ITIM and then acting on other substances, most experiments that study SHP-1 have used systems that omit this tethering process. As a result, the ‘reach’ of the reaction has not been studied. In this project an experimental assay, known as surface plasmon resonance (SPR)[70, 71], was used to study SHP-1 tethered reactions. The objective of the work is to develop a quantitative assay that can provide the binding kinetics, catalytic rate, and molecular reach for SHP-1 tethered reactions. We think of the SPR experiment as a simplification of what may happen in the cell’s membrane, and a translation of Figure 2.1 into what happens in the SPR can be seen in Figure 2.2. In this technique the proteins of interest, in this case the receptor tails, or ITIMs, are fixed in a three-dimensional dextran matrix¹. Therefore, unlike the membrane, where the receptor tails move within the 2D membrane plane, in our experiment

¹A dextran matrix is a support matrix used as a medium to tether proteins for experiments, made of a glucose derivative. The matrix has a web-like structure, due to the branched condition.

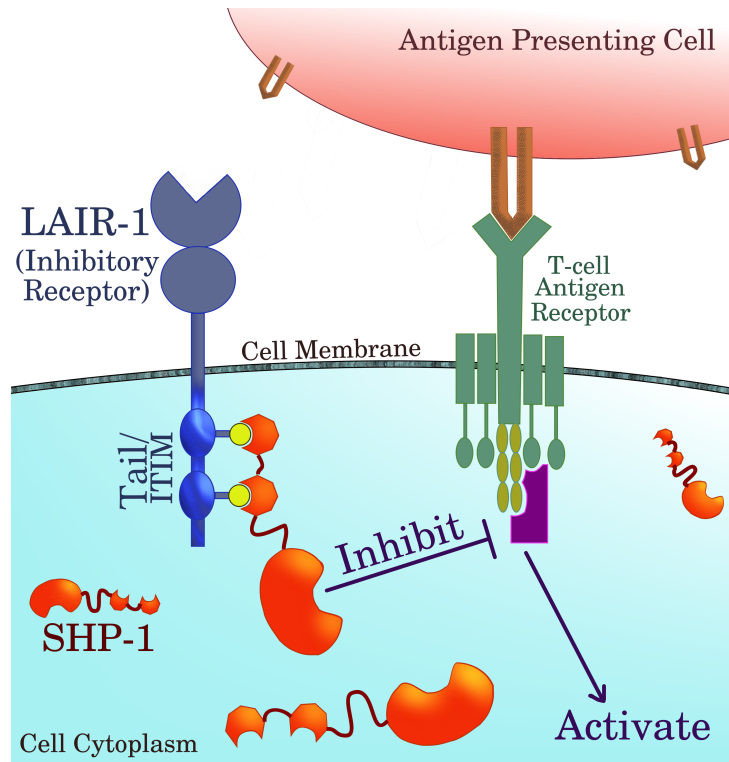


Figure 2.1: Sections of two cells. The higher, red one, is an undesired cell that presents antigens on its surface and is to be detected. The lower one is a T cell with both inhibitory and antigen receptors. LAIR-1 is an inhibiting receptor. When LAIR-1 binds to an inhibitory external protein it gets phosphorylated. SHP-1 enzyme, which floats in the cell cytoplasm, binds to the phosphorylated tail of LAIR-1 (its ITIM) to become activated and inhibit other protein reactions, like the phosphorylation of the T cell antigen receptor. The T cell antigen receptor would otherwise go on to activate the immune reaction of the T cell.

they are fixed and immobile. In the cell, SHP-1 would be floating in the cytoplasm, whilst in the SPR assay a solution with a known concentration of SHP-1 is constantly being injected over the dextran matrix. This means that there is never a shortage of SHP-1. A schematic representation of this can be seen in Figure 2.2. The SPR experiment provides a high sensitivity measure of the fraction of ITIMs bound by SHP-1 over time with 10 Hz time resolution and a very high signal-to-noise ratio. An example of the output can be seen in Figure 2.3.

The experiment starts with all ITIMs phosphorylated. As SHP-1 is washed over

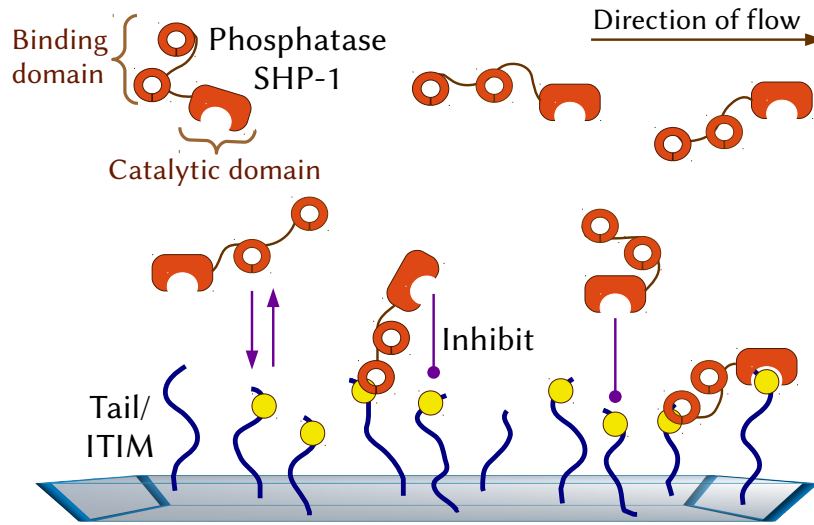


Figure 2.2: Representation of what happens in the *in vitro* experiment, using the surface plasmon resonance (SPR) assay. The lower, blue, rectangular section represents the dextran matrix to which the unphosphorylated (blue squiggly line) and phosphorylated (blue squiggly line with a yellow spot on top) tails are tethered. A buffer solution with a known concentration of SHP-1 is injected over the matrix and flows over it in the direction shown. The SHP-1 phosphatase then binds and unbinds to phosphorylated tails from the solution. SHP-1's catalytic domain dephosphorylates phosphorylated tails, inhibiting them, either from the solution or when bound to another phosphorylated tail.

the matrix, some SHP-1 binds to the phosphorylated ITIMs from the solution. This binding is reversible so that SHP-1 can bind and unbind from phosphorylated ITIMs. Once bound, a SHP-1 can reach out and dephosphorylate nearby phosphorylated ITIMs. Once dephosphorylated, an ITIM cannot go back to being phosphorylated. It is important to note that SHP-1 can only bind to phosphorylated ITIMs and, therefore, as it dephosphorylates ITIMs over time, it also reduces the total number of binding sites. This is why binding, after initially increasing, starts to decrease over time. It is also important to note that when an ITIM is bound by SHP-1, it cannot be dephosphorylated (either by the SHP-1 to which it is bound, or another SHP-1

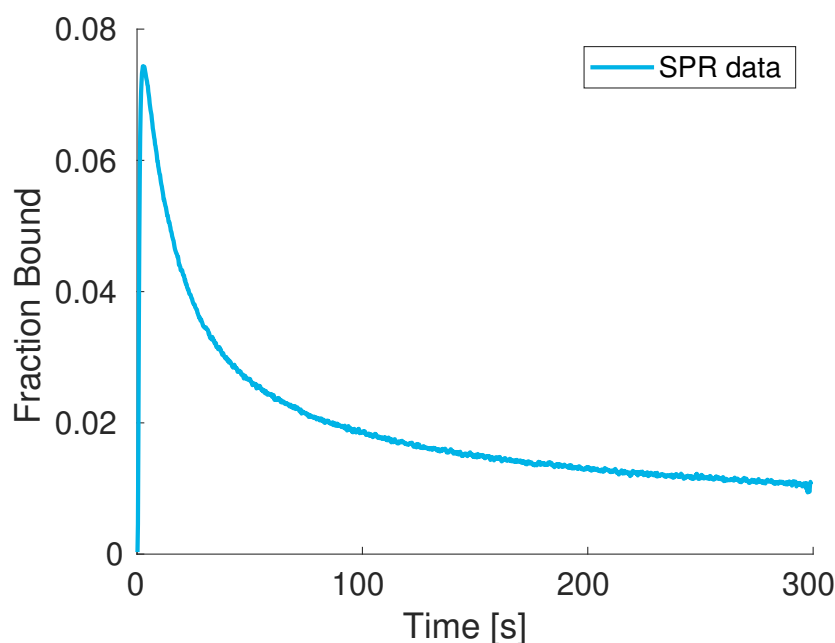


Figure 2.3: Data from the *in vitro* experiment using the surface plasmon resonance (SPR) assay. The fraction of phosphorylated tails bound by SHP-1 is shown with a 10 Hz time resolution. Starting time is as soon as the solution of SHP-1 is injected.

nearby). It is well known that binding protects these sites from dephosphorylation.

Although the majority of experiments are performed with a dextran matrix so that the ITIM is fixed in 3D space, experiments can also be performed without the matrix so that the ITIM is fixed directly on a 2D flat surface. We discuss these geometries in Section 2.5.

2.1.2 From experiment to mathematical abstraction

As mentioned in the Introduction, the models to be developed here can be applied to any couple of proteins or entities where one is tethered and the other one docks with it and interacts with its own docking site. Because we are interested in the reactions taking place amongst the proteins, we define and name each of the necessary species.

The phosphorylated ITIM, or tail, will be species A . The structure made by

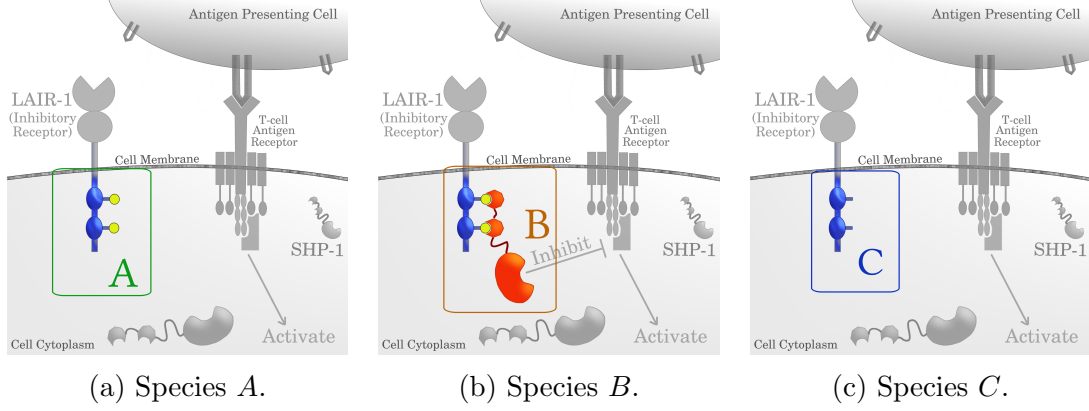


Figure 2.4: Representation of the biological structures as used for our mathematical models. The phosphorylated receptor tail is species *A* (shown in (a)), the compound of the SHP-1 enzyme bound to the phosphorylated tail is species *B* (shown in (b)), and the unphosphorylated receptor tail is species *C* (shown in (c)). Adding a couple of phosphate groups (coloured yellow) to *C* makes molecule *A*, enclosed in the green rectangle. Finally, the whole group of *A* plus the SHP-1 enzyme (coloured in orange), makes molecule *B*, enclosed in the orange rectangle.

an entity from species *A* bound by the SHP-1 enzyme will be species *B*. Finally, the unphosphorylated tail will be species *C*. The names can be seen in Figure 2.4, which is a representation of species *A*, *B*, and *C* in Figure 2.1. Figure 2.5 conveys how the experimental set-up from Figure 2.2 would translate into our abstraction.

Given that the ITIM is fixed in space, it follows that species *A*, *B*, and *C* are also all fixed in space. As a result, when a free ITIM (state *A*) is bound by SHP-1 (to form state *B*), it can convert nearby *A* molecules into *C* provided that they are within reach. This is illustrated in Figure 2.6.

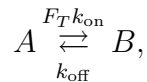
We note that *A* can turn into *B* and back, a *B* cannot be turned into *C*, and that once *A* has been turned into *C* it cannot go back. This means $A+B+C$ is a constant equal to the initial concentration of ITIM (i.e. $A(t) + B(t) + C(t) = A(t=0)$ since, initially, *B* and *C* are zero). Throughout the experiment we can observe the fraction of *B* with great accuracy. All reactions are further specified in Section 2.1.3.

We assume that it does not matter which binding site SHP-1 uses to bind with the tail in order to become B . Therefore, the objective is to use a deterministic model to infer the biochemical and biophysical constants of these reactions by only observing how much SHP-1 is bound over time. Nevertheless, with the help of the models developed here, the difference in reaction times presented by binding with different binding sites was further studied and the results are included in Goyette et al. [1, Fig. 6].

2.1.3 Reaction scheme

We start developing our models by equating the interaction of the proteins in the surface plasmon resonance (SPR) assay to chemical reactions.

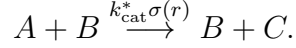
As SHP-1 binds and unbinds from A , we will have an *on* rate, k_{on} , and an *off* rate, k_{off} . SHP-1 that has un-docked effectively returns to the bulk pool of free SHP-1 in solution. Because of the constant flux of SHP-1 and because the amount of A is very small relative to the amount of bulk SHP-1, we assume that the pool of SHP-1 is not depleted by binding to A . This reaction may be written as:



where F_T is the concentration of injected SHP-1, which we consider constant.

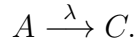
Once SHP-1 is bound to A (together labelled as B), it may dephosphorylate surrounding A with a *catalytic* rate, k_{cat}^* , resulting in C . For this to happen, B should be able to reach A , which means the reaction also depends on $\sigma(r)$, a function for the distance between the A and B reaction sites when anchored a distance r apart. $\sigma(r)$ is developed and further explained in Section 2.2.1. k_{cat}^* is the combined probability

of A and B coming together and B actually dephosphorylating A , rather than not reacting. Together, this reaction will look as follows:



The right hand side of the reaction has a B molecule because the A molecule is changed to C , but B is unaltered.

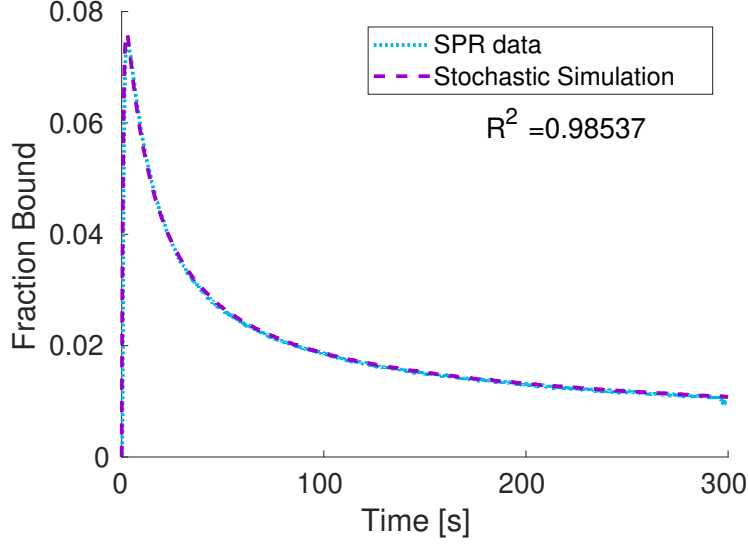
Finally, the possibility exists of SHP-1 dephosphorylating A directly from the solution without first binding to an ITIM, which we will denote as a solution catalytic rate, λ . We expect λ to be very low compared to k_{cat}^* at the beginning of the experiment because it seems more difficult for SHP-1 and A to find each other when SHP-1 is floating in the large space that the solution occupies, than for B reacting with A from a tethered point within reach, yet this is one of the issues we wish to explore. The reaction is therefore modelled as:



The idea that B will react with surrounding A whilst docked suggests that the combined length of the entities involved will play a role in the reaction speed. We will see in Section 2.2 that we require spatial correlations in our models to consider such effects.

Figure 2.7a shows a typical curve for B particles from the experiments. As SHP-1 starts to be injected it rapidly binds to A , but soon thereafter the binding decreases even though SHP-1 is still being injected and the association phase is still in process. We hypothesise that the reason for this decrease in binding is because

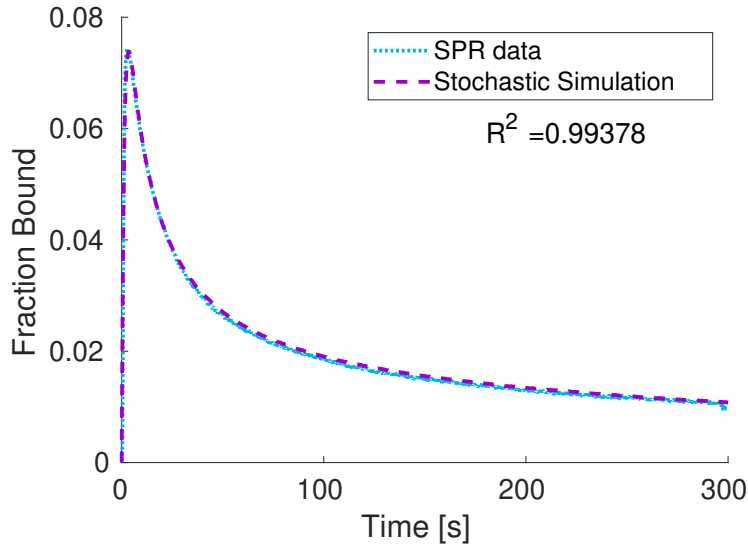
of the catalytic process, where docking sites are being dephosphorylated and SHP-1 has fewer binding sites.



(a) Parameters used for the stochastic simulation: $k_{\text{on}} = 0.117 \text{ s}^{-1}$, $k_{\text{off}} = 1.27 \text{ s}^{-1}$, $k_{\text{cat}} = 0.00837 \text{ } \mu\text{M/s}$, $L = 21.9 \text{ nm}$, $\lambda = 0.0005 \text{ s}^{-1}$. These parameter values were obtained from paper [1]. Periodic boundary conditions on a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube with $N_0 = 58470$ initial A were used. The average of 500 simulations, which took a total of 7.38 hours to run, is shown.

Figure 2.7: Example of experimental data obtained from the surface plasmon resonance (SPR) for the fraction of B compared to stochastic simulations. The blue dotted line is the experimental data and the purple dashed line is a fit using the stochastic model, explained in Section 2.1.4. Parameters used for the stochastic simulation are stated in each subfigure caption.

Following this description, the present deterministic model is developed in Section 2.3, which is hoped to better explain what can be observed in the experiments. A stochastic simulation model was made by Dr. Samuel Isaacson (University of Boston) and Dr. Jun Allard (University of California, Irvine), explained in Section 2.1.4, which is used to compare our model in Section 2.4.



(b) Parameters used for the stochastic simulations: $k_{\text{on}} = 0.0724 \text{ s}^{-1}$, $k_{\text{off}} = 0.7695 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0097 \text{ } \mu\text{M/s}$, $L = 20.1679 \text{ nm}$, $\lambda = 0.0097 \text{ s}^{-1}$. These parameter values were obtained from fitting our multi-particle density PDE model (developed in Section 2.3, Equations (2.59) to (2.57), called MPDPDE 3D model) to the experimental data. Periodic boundary conditions on a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube with 58470 ITIMs was used. 500 simulations, took a total of 6.55 hours to run.

Figure 2.7: (Continued) Example of experimental data obtained from the surface plasmon resonance (SPR) for the fraction of B compared to stochastic simulations. The blue dotted line is the experimental data and the purple dashed line is a fit using the stochastic model, explained in Section 2.1.4. Parameters used for the stochastic simulation are stated in each subfigure caption.

2.1.4 Stochastic Simulation

In order to ensure that our deterministic model accurately captures the chemistry in the SPR experiments, a stochastic simulation coded by Dr. Jun Allard and optimised by Dr. Samuel Isaacson was used. We note that deterministic models can often be solved analytically or numerically much more quickly than computational, stochastic models, making the former useful for more detailed data fitting. The stochastic simulation of the reaction scheme described in Section 2.1.3 was coded using MATLAB and C++. The Gibson-Bruck Next Reaction Method was used, a

variant of the Gillespie Stochastic Simulation Algorithm (SSA) [72, 73] where the next reaction time of each reaction is calculated and the reaction whose next reaction time is shortest is fired. A tree is used such that only reaction times that need to be updated after the chosen reaction are updated, saving time. This algorithm consists of 8 general steps:

1. Initialize, by setting A molecules randomly in a cuboid of side sizes w , l , and h , depending on the initial number of A , defined as N_0 , and the desired concentration. For the simulation, only the coordinates of each particle are recorded and the distance between these and the neighbours within a distance of r_{max} , making the simulation continuous in space. Set time, t , to 0.
2. Generate a dependency graph, G , detailing which reactions depend on which.
3. Determine the time step, τ_i , at which each chemical reaction, i , will occur by a process that draws random numbers from a uniform distribution in the unit interval. Store these time steps in an indexed priority queue, P .
4. Determine which reaction will occur by firing the chemical reaction i for which τ_i is smallest in P .
5. Update the time t to the chosen τ_i .
6. Update the corresponding population due to the chosen reaction.
7. Generate new time steps only for the reactions that need to be updated according to G . Replace in P the time values that had to change with the new reaction time, $t + \tau_i$.

8. Repeat from step 4 until the desired final time is reached.

When dealing with diffusion, the Gillespie SSA and its variants, like the Gibson-Bruck Next Reaction Method used here, are discrete in space, because diffusion needs to be calculated between blocks of space. But the ITIMs in our experiments are fixed in space and can be randomly allocated at the beginning of the simulation in a continuous space, therefore every species we model is in fixed in a continuous space. The distance between the ITIMs is then calculated and stored so this only needs to be done once and only reactions between molecules close enough to each other need to be considered. Therefore, the stochastic simulation in this chapter (and the next) is done in continuous space.

The agreement between the data and the stochastic simulations is very good, as can be seen in Figures 2.7a and 2.7b where the reported R^2 value is 0.99, very close to the perfect fit of 1. The experimental data in Figures 2.7a and 2.7b are the same, but the parameters used for the stochastic simulations in each are different. Figure 2.7a uses the parameters as obtained by Goyette et al. [1], where we do a parameter scan (testing the fit of different sets of parameters within the possible range of each parameter), whereas the parameters used for the stochastic simulations in Figure 2.7b were obtained by fitting the model developed in Section 2.3. This agreement between the SPR data and the stochastic simulations points towards the confirmation that the main behaviour of the proteins involved in the experiment is close to the behaviour detailed in Section 2.1.3, as previously thought.

These computational stochastic simulations were further used to generate arti-

ficial (*in silico*) data against which our deterministic models, developed in Section 2.3, could be tested. This means that the model parameters that generated the data are known and the model's accuracy can be tested.

2.2 Mean field approximation for protein dephosphorylation by SHP-1

We now have a computational model that we can fit to the experimental data but, as noted in Figure 2.7, it takes hours to run a sufficient number of simulations to have an average that smooths out the stochastic variations. This is not ideal time-wise, especially when several parameter sets have to be explored in search for the best fit. Hence we will develop an analytically tractable model that can hopefully be solved quicker. We will first start with the simplest model we can think of that takes into account the reactions proposed in Section 2.1.3, the classical mean-field model.

The chemical reactions considered, then, are as follows:



Here, $\sigma(r)$ gives the concentration of other A particles at a distance r from B , as described in the section that follows.

2.2.1 Interaction kernel $\sigma(r)$

We introduce the interaction kernel $\sigma(r)$, which is related to the probability that the end points of two proteins will be at the same coordinates despite their anchor

points being a distance r apart.

We can think of the proteins in these experiments as long chains of joint segments. These segments are composed of chains of amino acids, depending on the polymer. The first assumption will be that the segments can form any kind of angle at their junctions, even if that means the segment will overlap others. This is known as a freely-jointed chain, or random-flight chain, due to its similarity to random walk theory [74]. We can think of the chain as having n segments each of length a , as exemplified in Figure 2.8. We see in Yamakawa [75] that a measure for the average size of the chain can be obtained from:

$$L = \langle R^2 \rangle^{1/2} = n^{1/2}a, \quad (2.4)$$

where R is the distance from the beginning of the chain to the tip at any one time point and $\langle \rangle$ gives the average of all these time points. Hence, L can be thought of as the “reach”.

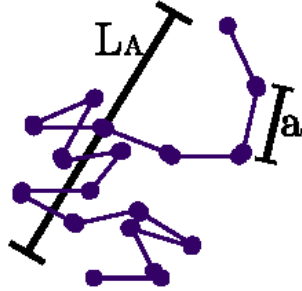


Figure 2.8: Representation of a random-flight chain, or freely-jointed chain, with $n = 16$ segments. Each segment’s length is a , but the root-mean-square distance from the first to the last point is equal to L_A .

Since our proteins have many sections, we can safely assume that $n \gg 1$. The random-flight chain theory also requires a final assumption that $R/na \ll 1$, which

implies that the chain bends easily. With these assumptions we can follow the work in Yamakawa [75] and obtain an expression for the probability distribution function of having the last point in the chain at a point $\vec{r}$ from the beginning of the chain:

$$P(\vec{r}) = \left(\frac{3}{2\pi L^2} \right)^{3/2} \exp \left(\frac{-3\vec{r} \cdot \vec{r}}{2L^2} \right). \quad (2.5)$$

Because of the properties of a random walk, this probability distribution has the form of a Gaussian. This is why this type of chain can also be called a Gaussian chain.

Now, we are interested in calculating the probability of particles A and B reaching each other depending on how far apart their starting points are. As can be seen in Figure 2.9, we have a probability distribution P_A for chain A and a different probability distribution P_B for chain B . $P_A(\vec{r}) d\vec{r}$ is the probability that the end point of chain A lies between $\vec{r}$ and $\vec{r} + d\vec{r}$. Similarly, $P_B(\vec{r}') d\vec{r}'$ is the probability that the end point of chain B lies between $\vec{r}'$ and $\vec{r}' + d\vec{r}'$. Multiplying $P_A(\vec{r}')P_B(\vec{r}' - \vec{r})$ determines a joint probability distribution function, which gives the number of conformations where A and B 's tips are in close proximity to each other divided over the number of all possible conformations for A and B 's tips. We assume that chain A has an average length L_A and chain B an average length L_B .

To calculate the probability density of both of the chains finding each other at $\vec{r}'$, regardless of where that $\vec{r}'$ is, we integrate this product over all space. This will

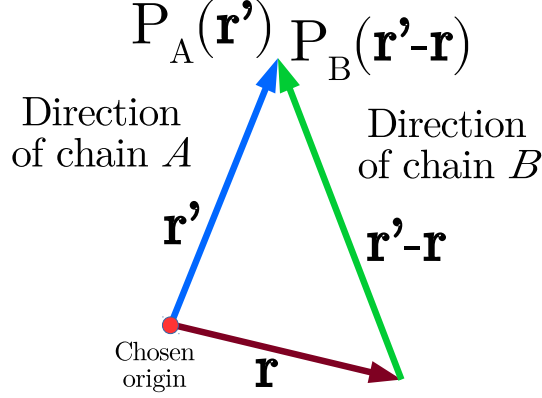


Figure 2.9: Visual explanation of how both chains' end points point to the same point in space although chain A starts at the origin and chain B starts at $\vec{r}$.

be our $\sigma(r)$:

$$\begin{aligned}
 \sigma(r) &= \int_{Volume} P_A(\vec{r}') P_B(\vec{r}' - \vec{r}) d^3 r' \\
 &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} r'^2 \sin \theta \left(\frac{3}{2\pi L_A^2} \right)^{3/2} \exp \left[\frac{-3\vec{r}' \cdot \vec{r}'}{2L_A^2} \right] \\
 &\quad \cdot \left(\frac{3}{2\pi L_B^2} \right)^{3/2} \exp \left[\frac{-3(\vec{r}' - \vec{r}) \cdot (\vec{r}' - \vec{r})}{2L_B^2} \right] dr' d\theta d\phi.
 \end{aligned}$$

Therefore:

$$\sigma(r) = \left[\frac{3}{2\pi(L_A^2 + L_B^2)} \right]^{3/2} \exp \left[\frac{-3r^2}{2(L_A^2 + L_B^2)} \right]. \quad (2.6)$$

From this expression we see that the probability density for the two chains finding each other is dependent on the sum of their reach ($L_A^2 + L_B^2$, which we define as $L_A^2 + L_B^2 = L^2$, due to their indistinguishability).

2.2.2 Mean Field Approximation

Since the equation for C decouples from those for A and B , we can simply use the conservation relationship to determine $C(t)$ once $A(t)$ and $B(t)$ are determined.

Therefore system (2.1) to (2.3) reduces to:



where $k_{\text{on}}^* = F_T k_{\text{on}}$, as we assume F_T , the amount of available SHP-1 in the substrate, will be constant.

When presented with a set of chemical reactions, like those in Equations (2.7) to (2.9), we would normally try to obtain the *reaction rate equation*. The reaction rate equation is the expression of the evolution of the mean A and B molecule numbers [7]. We derive a set of reaction rate equations from a phenomenological point of view. To introduce space, we will use $\sigma(r)$. Whilst for reactions between two molecules the division over the volume of the system would normally account for the fact that the larger the system is, the harder it is for two molecules to collide and react, $\sigma(r)$ will take care of this as it makes the reactions less likely the further away two molecules are from each other. The larger the system, the bigger r can become, and $\sigma(r)$ will decrease appropriately.

The set of equations will be:

$$\begin{aligned} \frac{\partial n_A(\vec{x}, t)}{\partial t} &= -(k_{\text{on}}^* + \lambda)n_A(\vec{x}, t) + k_{\text{off}}n_B(\vec{x}, t) \\ &\quad - k_{\text{cat}}^* \int \sigma(\vec{x} - \vec{x}') n_A(\vec{x}, t) n_B(\vec{x}', t) \, \text{d}^3 \vec{x}', \\ \frac{\partial n_B(\vec{x}, t)}{\partial t} &= k_{\text{on}}^* n_A(\vec{x}, t) - k_{\text{off}} n_B(\vec{x}, t), \end{aligned} \quad (2.10)$$

where $\vec{x}$ is the three-dimensional position vector, and n_A and n_B are the densities of

A and B molecules, respectively, at a point $\vec{x}$. Here, the first two terms on the right hand side of the n_A equation follow from the law of mass action [7], whilst the third term accounts for the interaction between each A and B molecule throughout the system, using $\sigma(r)$ to account for the space between molecules, as explained above. We assume the system volume grows to infinity for ease of calculation, because we know that the interaction between molecules will mostly depend on $\sigma(r)$, which decays rapidly, and because we know *a priori* that the sizes of the proteins in the SPR experiment are much smaller than the size of the dextran matrix, so we assume this to be a good enough approximation.

Because the initial conditions are spatially homogeneous (i.e. $n_A(t = 0) = A_0/V$, with A_0 being the initial number of A molecules in the system and V the volume of the system, and $n_B(t = 0) = 0$), there is no diffusion, and the reactions are symmetric, it follows that $n_B(\vec{x}, t)$ and $n_A(\vec{x}, t)$ should remain spatially homogeneous for all time. That is:

$$\begin{aligned} n_A(\vec{x}, t) &= a(t), \\ n_B(\vec{x}, t) &= b(t), \end{aligned} \tag{2.11}$$

where a and b are the density of A and B , respectively, throughout space.

As initial conditions we will have that:

$$\begin{aligned} a(t = 0) &= a_T, \\ b(t = 0) &= 0, \end{aligned} \tag{2.12}$$

with a_T the total density of initial A in the dextran matrix. Therefore, rescaling by

a_T , we can rewrite Equations (2.10) as:

$$\begin{aligned}\frac{da(t)}{dt} &= -(k_{\text{on}}^* + \lambda)a(t) + k_{\text{off}}b(t) \\ &\quad - k_{\text{cat}}^*a(t)b(t)a_T \int \sigma(\vec{x} - \vec{x}') d^3x', \\ \frac{db(t)}{dt} &= k_{\text{on}}^*a(t) - k_{\text{off}}b(t).\end{aligned}\tag{2.13}$$

This set of ordinary differential equations (ODEs) constitutes the traditional mean field approximation in the classical deterministic description.

As we assume an infinite space there are no boundaries in this mean-field model, as opposed to the periodic boundary conditions used in the Stochastic Simulations described in Section 2.1.4.

We now compute the integral term. Because $\sigma(r)$ only depends on the absolute distance between particles we can write it as $\sigma(|\vec{x} - \vec{x}'|)$. In spherical coordinates this is:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} r'^2 \sin \theta \sigma(|\vec{x} - \vec{x}'|) dr' d\theta d\phi.\tag{2.14}$$

The integral is done step by step in Appendix A. The result is:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} r'^2 \sin \theta \sigma(|\vec{x} - \vec{x}'|) dr' d\theta d\phi = 1.\tag{2.15}$$

Substituting Equation (2.15) into Equations (2.13) yields the complete mean-field equations:

$$\begin{aligned}\frac{da(t)}{dt} &= -(k_{\text{on}}^* + \lambda)a(t) + k_{\text{off}}b(t) - k_{\text{cat}}^*a(t)b(t)a_T, \\ \frac{db(t)}{dt} &= k_{\text{on}}^*a(t) - k_{\text{off}}b(t).\end{aligned}\tag{2.16}$$

The steady state of this system is easy to calculate, so we will obtain it for future

reference and to check against data. The steady state solutions to this system are:

$$\begin{aligned} a &= 0, \\ b &= 0, \end{aligned} \tag{2.17}$$

or:

$$\begin{aligned} a &= -\frac{\lambda L^4 k_{\text{off}}}{9k_{\text{cat}}^* k_{\text{on}}^*}, \\ b &= -\frac{\lambda L^4}{9k_{\text{cat}}^*}. \end{aligned} \tag{2.18}$$

The steady state of Equation (2.18) is negative and so not possible if we start with non-negative initial conditions (nor biologically realistic).

To analyse the linear stability of the steady state point (2.17), we linearise about the point. The eigenvalues that characterise the late-time behaviour around this steady state point, $(0, 0)$, are:

$$\mathcal{E} = \frac{k_{\text{on}}^* + \lambda + k_{\text{off}} \pm \sqrt{k_{\text{on}}^{*2} + \lambda^2 + k_{\text{off}}^2 + 2\lambda k_{\text{on}}^* + 2k_{\text{on}}^* k_{\text{off}} - 2\lambda k_{\text{off}}}}{-2}. \tag{2.19}$$

The corresponding eigenvectors are:

$$\vec{\mathcal{V}}_1 = \begin{pmatrix} \frac{k_{\text{on}}^* + \lambda - k_{\text{off}} + \sqrt{k_{\text{on}}^{*2} + \lambda^2 + k_{\text{off}}^2 + 2\lambda k_{\text{on}}^* + 2k_{\text{on}}^* k_{\text{off}} - 2\lambda k_{\text{off}}}}{-2k_{\text{on}}^*} \\ 1 \end{pmatrix}, \tag{2.20}$$

$$\vec{\mathcal{V}}_2 = \begin{pmatrix} \frac{k_{\text{on}}^* + \lambda - k_{\text{off}} - \sqrt{k_{\text{on}}^{*2} + \lambda^2 + k_{\text{off}}^2 + 2\lambda k_{\text{on}}^* + 2k_{\text{on}}^* k_{\text{off}} - 2\lambda k_{\text{off}}}}{-2k_{\text{on}}^*} \\ 1 \end{pmatrix}. \tag{2.21}$$

We now have solutions of the form:

$$\vec{\mathcal{V}}(t) = c_1 \vec{\mathcal{V}}_1 e^{\mathcal{E}_1 t} + c_2 \vec{\mathcal{V}}_2 e^{\mathcal{E}_2 t}, \tag{2.22}$$

where c_1 and c_2 are real constants that will depend on the initial conditions. $\mathcal{E}_1$ is the $\mathcal{E}$ with a plus sign between k_{off} and the square root in (2.19).

Experiments suggest that $k_{\text{on}}^* > \lambda$ so $\vec{\mathcal{V}}_1$ can not be equal to $\vec{\mathcal{V}}_2$. This also means that $\mathcal{E}_1 \neq \mathcal{E}_2$. In fact, simple algebra shows that $\mathcal{E}_1 < \mathcal{E}_2 < 0$. This means that as time goes to $-\infty$, from Equation (2.22), $\vec{\mathcal{V}}$ will be dominated by the $\vec{\mathcal{V}}_2$ term, unless $c_2 = 0$, in which case the solution will be along $\vec{\mathcal{V}}_1$. As time goes to ∞ , $\vec{\mathcal{V}}(t)$ tends to 0, making the steady state $(0, 0)$ a stable node.

We show the phase plane of the ODEs (2.16) in Figure 2.10. Note that in the beginning the system moves quickly: $b(t)$ leaves 0 and $a(t)$ starts dropping towards 0 at a fast pace. After 1 second, the trajectory has travelled a long distance. The evolution of the system slows down noticeably as it approaches the steady state, $(0, 0)$.

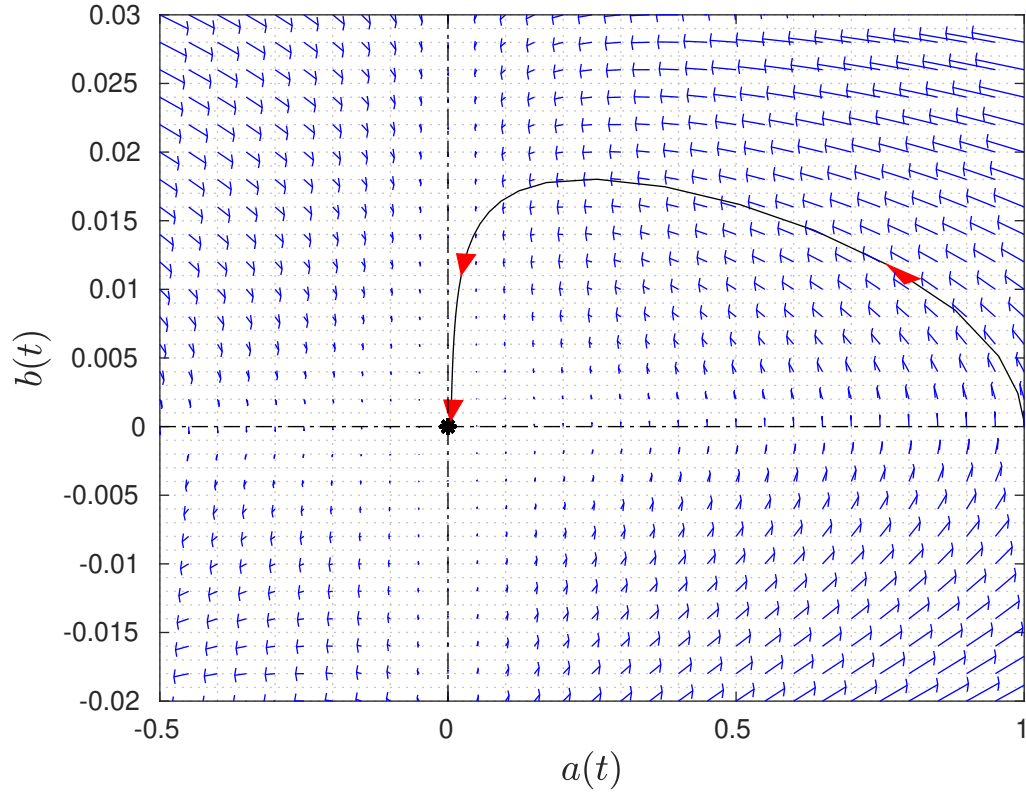
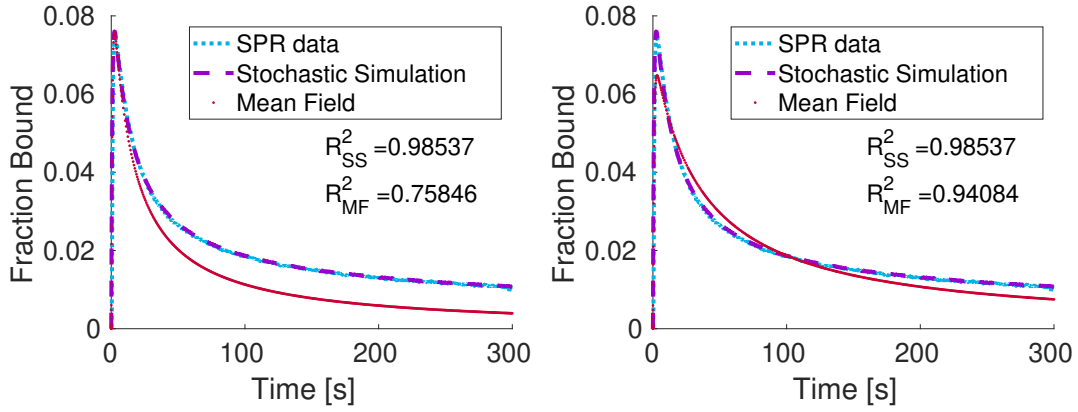


Figure 2.10: Phase Plane of ODE (2.16) with variables $k_{\text{on}}^* = 0.0888 \text{ s}^{-1}$, $k_{\text{off}} = 1.1988 \text{ s}^{-1}$, $\lambda = 0.0001 \text{ s}^{-1}$, $k_{\text{cat}}^* = 0.0043 \text{ } \mu\text{M/s}$, $a_T = 58,470$ initial A proteins, and 0 initial B proteins. Red arrows over the line of flow are at time 0.1, 1, and 5 s.

To show the difference between the experimental data and the mean-field model, we have chosen a particular set of experimental data, as done in Goyette et al. [1]. The comparison can be seen in Figure 2.11. Whilst Figure 2.11a uses the parameters estimated in Goyette et al. [1] for the mean-field model, Figure 2.11b uses a least-squares solver to fit the mean-field model. The model is not a good fit and exhibits systematic differences from both the data and the stochastic simulations, suggesting that this mean field approximation is insufficient to explain the experimental results.



(a) Mean field model with the same parameters as those used for the stochastic simulations.

(b) Parameter values used for the mean field model: $k_{\text{on}} = 0.0888 \text{ s}^{-1}$, $k_{\text{off}} = 1.1988 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0043 \text{ } \mu\text{M/s}$, $\lambda = 0.0001 \text{ s}^{-1}$. These were obtained by fitting the mean field model to the SPR data using MATLAB's nonlinear least-squares solver.

Figure 2.11: Example of experimental data and stochastic simulations versus mean field approximation model. Shown is the fraction of SHP-1 bound to ITIM. Experimental data obtained from the surface plasmon resonance (SPR). Blue dotted line denotes the experimental data, purple dashed line is the average of 500 stochastic simulations, fitted as explained in Section 2.1.4, and the red dotted line is the solution of the mean field equations (Equations (2.16)) with a different type of fit in each subfigure. Parameter values used for the stochastic simulations: $k_{\text{on}} = 0.117 \text{ s}^{-1}$, $k_{\text{off}} = 1.27 \text{ s}^{-1}$, $k_{\text{cat}} = 0.00837 \text{ } \mu\text{M/s}$, $L = 21.9 \text{ nm}$, $\lambda = 0.0005 \text{ s}^{-1}$. These values were obtained from Goyette et al. [1]. Periodic boundary conditions on a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube with 58,470 ITIMs were used for the stochastic simulations.

2.3 Multi-centre Particle Density Model of Dephosphorylation

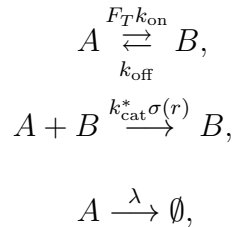
The mean-field model proposed in Section 2.2.2 did not provide a sufficiently good fit to experimental data, so in search of a better deterministic description of the reactions proposed in Section 2.1.3, we use the multi-centre particle density formalism [76].

The simplest kinetic approach takes into account a well mixed spatial distri-

bution, but this might not always be appropriate. In the case of our experiment, the A molecules are fixed, so they are unable to diffuse and mix uniformly across the domain, despite starting fixed uniformly in the domain. This also fixes the B molecules once SHP-1 has bound. Whilst B is tethered the number of A molecules that B can reach is effectively low (we estimate 8 A molecules within reach ²) and, as they become dephosphorylated, the number drops towards zero. In this case small numbers may play an important role, and stochastic variability could become important, whereas the mean-field approximation neglects this. Following the work by Schnörer et al. [77], we developed a model for reactants that are at low copy numbers and fixed in space, which we present here.

We consider *immobile* A and B particles initially distributed randomly in d -dimensional space. We will propose a reaction rate, $k_{\text{cat}}^* \sigma(r)$, that is, the rate at which A and B particles situated at a distance r apart react, where k_{cat}^* is the constant of reaction, whilst $\sigma(r)$ is the concentration of A molecules at a distance r from B . $\sigma(r)$ was discussed in Section 2.2.1.

Under our assumptions, the interaction between A and B satisfies the same chemical reactions in Equations (2.7) to (2.9), written down here again for convenience:



²This is a very rough estimate suggested by collaborators, but based on a concentration of around 117 μM in the SPR experiment, which translates to a concentration of $7.96 \times 10^{-5} \text{ nm}^{-3}$, giving around 9 ITIMs per spheres of 30 nm of radius.

where F_T is the constant concentration of SHP-1 in the substrate, which never gets depleted, A represents the concentration of phosphorylated ITIMs, and B represents the concentration of bound SHP-1 to A . As stated, we would expect λ to be small compared to the other reactions because of the difficulty for a large protein like SHP-1 to dock onto a relatively short chain like our ITIM whilst floating by, but we included it for completeness.

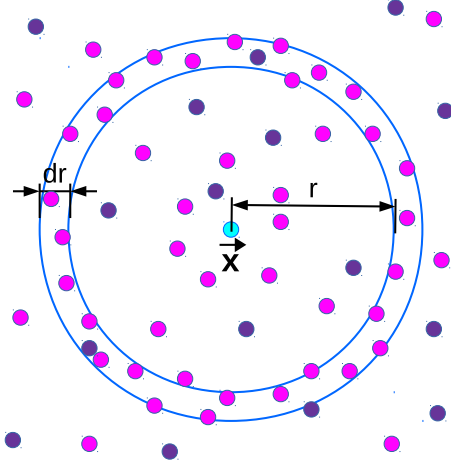
To obtain the rate equations for these we will need to first understand the concept of Pair Correlation Functions, as these will help us incorporate the spatial relation between our molecules into our model.

2.3.1 Pair Correlation Functions

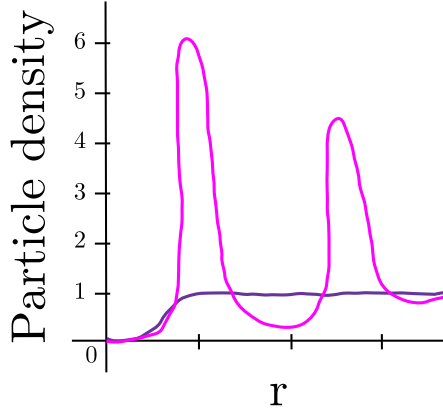
In order to keep track of the spatial relation amongst particles, we will introduce the notion of a k -point distribution function [76, 78, 79], $\rho^{(k)}$ ($k = 1, 2, \dots$). We will define the function later, but an intuitive way of seeing it is as follows:

We stand on a particle at $\vec{x}$, make a spherical shell centred at $\vec{x}$ of radius r to $r+dr$ (dr being very small), and count how many particles of interest are within that shell (see Figure 2.12a). If we do this for several spherical shells surrounding our particle at $\vec{x}$ we will have an idea of the number of particles of interest around our $\vec{x}$ point for each distance in r (of course, dividing the count of particles of interest by $4\pi r^2 dr$ to account for the fact that the further away we are, the larger the shell, and we will naturally find more particles). This can be seen in Figure 2.12b.

This knowledge of surrounding density is dependant on the position $\vec{x}$. But if we now take an average from the position of all, say, A particles, we will have



(a) Representation of a particle in point $\vec{x}$ surrounded by pink (lighter colour) and purple (darker colour) particles. A spherical shell composed of a boundary of radius r and one of radius $r + dr$ is shown.



(b) Density of pink particles and purple particles at distance r as seen from point $\vec{x}$. Density not measured exactly, but approximated for exemplifying purposes.

Figure 2.12: Representation of how the particle density can be measured from a point $\vec{x}$, as explained in the text. Note that both the pink (lighter colour) and the purple (darker colour) lines in Figure 2.12b, start at 0, because no particles are found surrounding the particle at $\vec{x}$ for a small distance in (a). Then the pink graph starts oscillating because there is a spatial relation between pink dots and the particle at $\vec{x}$, whilst the purple dots have no spatial relation, so the graph is uniform in space.

an idea of the density relation between particles over all of space. A good visual representation of an example of this can be seen at <http://www.physics.emory.edu/faculty/weeks//idl/gofr.html> [80].

In our case, since we only have particles A and B , we can write this probability distribution as in Kuzovkov and Kotomin [81]. We divide the volume of our system, V , into a large number, N , of small regions of the same volume, v , such that each region has at most one particle. Each region can be identified by a vector $\vec{r}$. We can describe the number of particles in each region $\vec{r}$ by $\nu_{\vec{r}}$, where $\nu = A$ or B . With this, we can describe the complete set of population numbers by $\{\nu_{\vec{r}}\}_N = \nu_{\vec{r}_1}, \dots, \nu_{\vec{r}_N}$. If $P(\{\nu_{\vec{r}}\}_N)$ is the probability of finding the system in an arbitrary physical state

$\{\nu_{\vec{r}}\}_N$, and we assume the continuous transition where $v \rightarrow 0$ and $N \rightarrow \infty$ such that $Nv = V$, then we can introduce the many-point distribution function:

$$\rho_m(\{\vec{r}\}_m) = v^{-m} P(\{\nu_{\vec{r}}\}_m),$$

where $\{\vec{r}\}_m \equiv \{\vec{r}_1, \dots, \vec{r}_m\}$. In particular, we can introduce the $(m + m')$ -point distribution functions, $\rho_{m,m'}$:

$$\rho_{m,m'}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}) = v^{-(m+m')} P(\{A_{\vec{r}}\}_m; \{B_{\vec{r}'}\}_{m'}), \quad (2.23)$$

where, as in Kotomin and Kuzovkov [82], $P(\{A_{\vec{r}}\}_m; \{B_{\vec{r}'}\}_{m'})$ gives the probability of finding particles A in m sub-volumes identified by the coordinates $\vec{r}_i (i = 1, \dots, m)$ and particles B in m' sub-volumes identified by the coordinates $\vec{r}'_j (j = 1, \dots, m')$. Physically speaking, $\rho_{m,m'} d\vec{r}_1, \dots, d\vec{r}'_{m'}$ is the mean number of configurations with m particles A in unit volumes from $\vec{r}_i$ to $\vec{r}_i + d\vec{r}_i$ ($i = 1, \dots, m$) and m' particles B in unit volumes from $\vec{r}'_j (j = 1, \dots, m')$ to $\vec{r}'_j + d\vec{r}'_j$ ($j = 1, \dots, m'$) [83]. Following this idea, $P(A_{\vec{r}})$ is the probability that we will find a particle A at a region around $\vec{r}$, which is also the mean number of A particles in the system as there can only be one particle per region. Therefore $\rho_{1,0}(\vec{r};) = P(A_{\vec{r}})/v$ is the density (concentration) of particles A at the region identified by the $\vec{r}$. For a homogeneous system ($V \rightarrow \infty$) $\rho_{1,0}$ and $\rho_{0,1}$ are independent of the coordinates, and so:

$$\rho_{1,0} \equiv n_A(t), \quad (2.24)$$

where we define $n_A(t)$ as the mean-field particle concentration of A . Similarly:

$$\rho_{0,1} \equiv n_B(t), \quad (2.25)$$

with $n_B(t)$ being the mean-field particle concentration of B .

On the other hand, $\rho_{1,1}(\vec{r}_1; \vec{r}_2)$ is the *joint* density of a particle A at region $\vec{r}_1$ and a particle B at region $\vec{r}_2$. As stated by Kotomin and Kuzovkov [82], “It can be shown that eq. (2.23) describes a set of macroscopic $(n + n')$ -point densities (not probability densities) which are nothing but averaged products of powers of the macroscopic densities $\rho_{1,0}$, $\rho_{0,1}$ taken at different spatial points [83].” In a homogeneous system the joint distribution functions, $\rho_{1,1}$, $\rho_{2,0}$ and $\rho_{0,2}$ depend on the relative coordinates $r = |\vec{r} - \vec{r}'|$.

From here we can define a correlation function, $\chi_{\nu,\nu'}$ such that [81, 84]:

$$\rho_{m,m'}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}) = n_\nu^m n_{\nu'}^{m'} \chi_{\nu,\nu'}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}), \quad (2.26)$$

where n_ν is the concentration of particle ν in the system. From Equation (2.23), the higher-order correlation functions are defined as [81, 85]:

$$\chi_{m,m'}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}) = \frac{P(\{A_{\vec{r}}\}_m; \{B_{\vec{r}'}\}_{m'})}{\prod_{i=1}^m P(A_{\vec{r}_i}) \prod_{j=1}^{m'} P(B_{\vec{r}'_j})}.$$

These correlation functions can serve as a measure of spatial correlation that depends on the relative distance $r = |\vec{r} - \vec{r}'|$. This is because for two different volumes, v and v' , that are non-overlapping, of the same size, centred at $\vec{r}$ and $\vec{r}'$, and whose reactants have no spatial correlation:

$$P(\{\nu_{\vec{r}}\}; \{\nu_{\vec{r}'}\}) = P(\nu_{\vec{r}})P(\nu_{\vec{r}'}),$$

and $\chi_{\nu,\nu'}^{(2)}(r, t) = 1$ in this case. Similarly, because the correlations become negligible when $r \rightarrow \infty$, $\chi_{\nu,\nu'}^{(2)}(\infty, t) = 1$.

As it is easier to start from the *many-point densities*, $\rho^{(k)} = \rho_{m,m'}$ (where $k = m + m'$) of a number of reactants when trying to derive the kinetic equations in question, we will start with:

$$\rho_{m,m'} = n_A^m(t)n_B^{m'}(t)\chi_{m,m'}. \quad (2.27)$$

For $k = 2$ the distribution function depends on two particles and their coordinates.

The possibilities are:

$$\rho_{2,0} \equiv n_A^2(t)X_A(|\vec{r}_1 - \vec{r}_2|, t), \quad (2.28)$$

$$\rho_{0,2} \equiv n_B^2(t)X_B(|\vec{r}_1 - \vec{r}_2|, t), \quad (2.29)$$

$$\rho_{1,1} \equiv n_A(t)n_B(t)Y(|\vec{r}_1 - \vec{r}_1|, t), \quad (2.30)$$

where $X_A(r, t)$ and $X_B(r, t)$ are the joint correlation functions of similar reactants, $\chi_{\nu,\nu'}^{(2)}(r, t) = X_\nu(r, t)$, and $Y(r, t)$ is for dissimilar reactants $\chi_{A,B}^{(2)}(r, t) = \chi_{B,A}^{(2)}(r, t) = Y(r, t)$. These are the pair correlation functions, which are “related to the probability of finding the centre of a particle a given distance from the centre of another particle” [80]. This indicates the spatial distribution of the different types of particles and the concentration in specific areas around each particle. We emphasize that these pair correlation functions have the property of equalling one when there is no correlation between the two particles. Note that

$$X_A = \frac{\rho_{2,0}}{\rho_{1,0}\rho_{1,0}}.$$

Independence of the events means the probability of both happening is equal to the multiplication of each one happening independently, $X_A = 1$ only when each of the A particles is independent of all the others. The same logic applies to X_B and Y .

Equation (2.28) states how likely you are to find another A particle a distance $|\vec{r}_1 - \vec{r}_2|$ away given that you have an A molecule. Equation (2.29) is similar but with B particles, whereas Equation (2.30) refers to having an A particle and trying to find a B particle a distance $|\vec{r}_1 - \vec{r}_1'|$ away. Equation (2.30) is symmetric for particles A and B .

From [81] we know that “the derived set of kinetic equations for $\rho_{m,m'}$ ($m, m' = 0, 1, \dots$) contains three kinds of contributions

$$\frac{\partial \rho_{m,m'}}{\partial t} = \left. \frac{\partial \rho_{m,m'}}{\partial t} \right|_{diff} + \left. \frac{\partial \rho_{m,m'}}{\partial t} \right|_{rec} + \left. \frac{\partial \rho_{m,m'}}{\partial t} \right|_{prod}, \quad (2.31)$$

due to reactant diffusion, recombination and production, respectively.” This approach assumes these three interactions to be combined linearly and to be the only interactions needed for this kind of reaction. We follow this framework as it has been used extensively [82, 86]. For our problem, the first term will be zero since we are considering no diffusion.

Following [81], the recombination contribution corresponds to the following system of coupled integro-differential equations:

$$\begin{aligned} \left. \frac{\partial \rho_{m,m'}}{\partial t} \right|_{rec} = & - \sum_{i=1}^m \sum_{j=1}^{m'} \sigma^*(r_i - r'_j) \rho_{m,m'} \\ & - \sum_{i=1}^m \int \sigma^*(r_i - r'_{m'+1}) \rho_{m,m'+1} d^d \vec{r}'_{m'+1} \\ & - \sum_{j=1}^{m'} \int \sigma^*(r'_j - r_{m+1}) \rho_{m+1,m'} d^d \vec{r}_{m+1}. \end{aligned} \quad (2.32)$$

It must be noted that here σ^* contains k_{cat}^* within it: $\sigma^* = k_{cat}^* \sigma$. This recombination term has the following parts: the first term refers to all the possible combinations amongst A and B pairs within the group consisting of m particles of A and m'

particles of B , the second term accounts for recombination of A particles in the group with B particles outside said group, and the third term accounts for recombination of B particles in the group with A particles outside the same group. The third term would mean the loss of a B particle from the group when it interacts with an A particle outside the considered group, but since B is conserved when reacting with A particles, we can dismiss the third term as it is zero.

For the production term we consider Equations (2.7) and (2.9), that correspond to the production part. An A particle at r_1 can be created by a B particle in r_1 losing its SHP-1 molecule with a rate k_{off} . In the same sense, an A particle at r_1 can be lost by an SHP-1 molecule docking on to an A molecule with a reaction rate of k_{on}^* , or a SHP-1 molecule dephosphorylating from the substrate at a rate λ . This gives:

$$\left. \frac{\partial \rho_{1,0}}{\partial t} \right|_{\text{prod}} = k_{\text{off}} \rho_{0,1} - k_{\text{on}}^* \rho_{1,0} - \lambda \rho_{1,0}. \quad (2.33)$$

Two A particles, one at r_1 and a second at r_2 , can be created by having an A molecule at r_1 and a B molecule at r_2 and the B molecule changing to A with a reaction rate of k_{off} or, symmetrically, having a B molecule at r_1 and an A molecule at r_2 and the B molecule changing to A with the same reaction rate. The same state with two A particles can be lost by any one of the two A molecules being docked by SHP-1 and turning to B with rate k_{on}^* , or any of them being dephosphorylated by an unattached SHP-1 molecule with rate λ . This leads to:

$$\begin{aligned} \left. \frac{\partial \rho_{2,0}}{\partial t} \right|_{\text{prod}} &= k_{\text{off}} \rho_{1,1} + k_{\text{off}} \rho_{1,1} - k_{\text{on}}^* \rho_{2,0} - k_{\text{on}}^* \rho_{2,0} - \lambda \rho_{2,0} - \lambda \rho_{2,0} \\ &= 2k_{\text{off}} \rho_{1,1} - 2(k_{\text{on}}^* + \lambda) \rho_{2,0}. \end{aligned} \quad (2.34)$$

From Equations (2.33) and (2.34) we can deduce that for m A particles we will have m different possibilities of having $m - 1$ A particles and a B particle that changes to A with reaction rate k_{off} . Each A particle can be turned into a B particle with rate k_{on}^* , or can be lost, each with rate λ . Applying the same logic to the existence of m' B particles:

$$\left. \frac{\partial \rho_{m,m'}}{\partial t} \right|_{\text{prod}} = mk_{\text{off}}\rho_{m-1,m'+1} - m(k_{\text{on}}^* + \lambda)\rho_{m,m'} + m'k_{\text{on}}^*\rho_{m+1,m'-1} - m'k_{\text{off}}\rho_{m,m'}. \quad (2.35)$$

Hence we can put together the general formula by substituting Equations (2.32) and (2.35) into (2.31):

$$\begin{aligned} \frac{\partial \rho_{m,m'}}{\partial t} = & - \sum_{i=1}^m \sum_{j=1}^{m'} k_{\text{cat}}^* \sigma(r_i - r'_j) \rho_{m,m'} - \sum_{i=1}^m \int k_{\text{cat}}^* \sigma(r_i - r'_{m'+1}) \rho_{m,m'+1} d^d \vec{r}'_{m'+1} \\ & + mk_{\text{off}}\rho_{m-1,m'+1} + m'k_{\text{on}}^*\rho_{m+1,m'-1} - m(k_{\text{on}}^* + \lambda)\rho_{m,m'} - m'k_{\text{off}}\rho_{m,m'}. \end{aligned} \quad (2.36)$$

This general formula will give us the integro-PDEs of our model dynamics. For this we will need to get equations in terms of n_A , n_B , and their joint correlation functions. To begin this process, we will consider only one A particle:

$$\frac{\partial \rho_{1,0}}{\partial t} = - \int k_{\text{cat}}^* \sigma(r_1 - r'_1) \rho_{1,1} d^d r'_1 + k_{\text{off}}\rho_{0,1} - (k_{\text{on}}^* + \lambda)\rho_{1,0}. \quad (2.37)$$

We can substitute the definitions in Equations (2.24), (2.25), and (2.30) into Equation (2.37):

$$\begin{aligned} \frac{dn_A(t)}{dt} = & - \int k_{\text{cat}}^* \sigma(r_1 - r'_1) n_A(t) n_B(t) Y(\vec{r}_1 - \vec{r}'_1, t) d^d \vec{r}'_1 \\ & + k_{\text{off}} n_B(t) - (k_{\text{on}}^* + \lambda) n_A(t). \end{aligned}$$

When making the substitution $\vec{r}_1 - \vec{r}'_1 = \vec{r}'$:

$$\frac{dn_A(t)}{dt} = - n_A(t) n_B(t) \int k_{\text{cat}}^* \sigma(r') Y(\vec{r}', t) d^d \vec{r}' + k_{\text{off}} n_B(t) - (k_{\text{on}}^* + \lambda) n_A(t). \quad (2.38)$$

In a similar way for one B particle, but simpler because of the lack of recombination term, we obtain:

$$\begin{aligned}\frac{\partial \rho_{0,1}}{\partial t} &= k_{\text{on}}^* \rho_{1,0} - k_{\text{off}} \rho_{0,1}, \\ \frac{dn_B(t)}{dt} &= k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t).\end{aligned}\tag{2.39}$$

For the following equations we will need three-point distribution functions, which we have not defined. For this we will need a closure approximation.

2.3.2 Kirkwood Superposition Approximation

Extended to its fullest, the general Equation (2.36) would lead to an infinite system of non-linear partial differential equations. This is because, as seen from the second right hand term in Equation (2.36), the A particles being considered will combine with B particles outside the system being considered, so that we need $\rho_{1,1}$ to describe $\rho_{1,0}$ (Equation (2.37)), $\rho_{1,2}$ to describe $\rho_{1,1}$, $\rho_{1,3}$ to describe $\rho_{1,2}$, $\rho_{2,1}$ to describe $\rho_{2,0}$, and so on. That is, we need to consider the correlations of $n+1$ particles to describe the rate of n particles when at least one A is involved. So a closure approximation is needed. This approximation was first introduced by Kirkwood [87], and consists of decoupling the three-centre correlation functions into the product of three pair correlation functions [85]:

$$\langle n(r_1)n(r_2)n(r_3) \rangle \approx \frac{\langle n(r_1)n(r_2) \rangle \langle n(r_2)n(r_3) \rangle \langle n(r_3)n(r_1) \rangle}{\langle n(r_1) \rangle \langle n(r_2) \rangle \langle n(r_3) \rangle},\tag{2.40}$$

where each r_i is a position vector in d dimensions.

A derivation of this approximation can be seen in Singer [88]. Its main validation is done by comparing its predictions to experiments or simulations [83, p.247]. This

approximation, as stated by Dettmann et al. [86], will be quantitatively valid for small densities of particles, as the further apart particles are on average, the less likely that higher order interactions will occur and have a strong effect. For higher densities we would obtain the right order of magnitude [86]. We consider stochasticity is important in our system, due precisely to the low density of particles, meaning this approximation can be valid.

We use Kirkwood's approximation for $\rho_{2,1}$ and for $\rho_{1,2}$:

$$\begin{aligned}
\rho_{2,1} &= \langle n_A(r_1, t) n_A(r_2, t) n_B(r'_1, t) \rangle \\
&\approx \frac{\langle n_A(r_1, t) n_A(r_2, t) \rangle \langle n_A(r_2, t) n_B(r'_1, t) \rangle \langle n_B(r'_1, t) n_A(r_1, t) \rangle}{\langle n_A(r_1, t) \rangle \langle n_A(r_2, t) \rangle \langle n_B(r'_1, t) \rangle} \\
&= \frac{n_A^2(t) X_A(r_1 - r_2, t) n_A^2(t) n_B^2(t) Y(r_1 - r'_1, t) Y(r_2 - r'_1, t)}{n_A^2(t) n_B(t)} \\
&= n_A^2(t) n_B(t) X_A(r_1 - r_2, t) Y(r_1 - r'_1, t) Y(r_2 - r'_1, t), \tag{2.41}
\end{aligned}$$

$$\begin{aligned}
\rho_{1,2} &= \langle n_A(r_1, t) n_B(r'_1, t) n_B(r'_2, t) \rangle \\
&\approx \frac{\langle n_A(r_1, t) n_B(r'_1, t) \rangle \langle n_B(r'_1, t) n_B(r'_2, t) \rangle \langle n_B(r'_2, t) n_A(r_1, t) \rangle}{\langle n_A(r_1, t) \rangle \langle n_B(r'_1, t) \rangle \langle n_B(r'_2, t) \rangle} \\
&= \frac{n_A^2(t) n_B^2(t) Y(r_1 - r'_1, t) Y(r_1 - r'_2, t) n_B^2(t) X_B(r'_1 - r'_2, t)}{n_A(t) n_B^2(t)} \\
&= n_A(t) n_B^2(t) X_B(r'_1 - r'_2, t) Y(r_1 - r'_1, t) Y(r_1 - r'_2, t). \tag{2.42}
\end{aligned}$$

Now we can put an A particle and a B particle into the general Equation (2.36):

$$\begin{aligned}
\frac{\partial \rho_{1,1}}{\partial t} &= -k_{\text{cat}}^* \sigma(r_1 - r'_1) \rho_{1,1} - \int k_{\text{cat}}^* \sigma(r_1 - r'_2) \rho_{1,2} \, d^d r'_2 \\
&\quad + k_{\text{off}} \rho_{0,2} + k_{\text{on}}^* \rho_{2,0} - (k_{\text{on}}^* + \lambda) \rho_{1,1} - k_{\text{off}} \rho_{1,1}. \tag{2.43}
\end{aligned}$$

Using Kirkwood's approximation (2.42) and substituting Equations (2.28) to (2.30)

into (2.43), we obtain:

$$\begin{aligned}
\frac{\partial \rho_{1,1}}{\partial t} = & -k_{\text{cat}}^* \sigma(r_1 - r'_1) n_A(t) n_B(t) Y(\vec{r}_1 - \vec{r}'_1, t) \\
& - \int \left[k_{\text{cat}}^* \sigma(r_1 - r'_2) n_A(t) n_B^2(t) \right. \\
& \cdot X_B(\vec{r}'_1 - \vec{r}'_2, t) Y(\vec{r}_1 - \vec{r}'_1, t) Y(\vec{r}_1 - \vec{r}'_2, t) \left. \right] d^d \vec{r}'_2 \\
& + k_{\text{off}} n_B^2(t) X_B(\vec{r}_1 - \vec{r}'_1, t) + k_{\text{on}}^* n_A^2(t) X_A(\vec{r}_1 - \vec{r}'_1, t) \\
& - (k_{\text{on}}^* + \lambda) n_A(t) n_B(t) Y(\vec{r}_1 - \vec{r}'_1, t) - k_{\text{off}} n_A(t) n_B(t) Y(\vec{r}_1 - \vec{r}'_1, t). \quad (2.44)
\end{aligned}$$

Making the substitutions $\vec{r}_1 - \vec{r}'_1 = \vec{r}$, and $\vec{r}_1 - \vec{r}'_2 = \vec{r}'$, we obtain:

$$\begin{aligned}
\frac{\partial \rho_{1,1}}{\partial t} = & -k_{\text{cat}}^* n_A(t) n_B(t) \sigma(\vec{r}) Y(\vec{r}, t) \\
& - k_{\text{cat}}^* n_A(t) n_B^2(t) Y(\vec{r}, t) \int \sigma(\vec{r}') X_B(\vec{r}' - \vec{r}, t) Y(\vec{r}', t) d^d \vec{r}' \\
& + k_{\text{off}} n_B^2(t) X_B(\vec{r}, t) + k_{\text{on}}^* n_A^2(t) X_A(\vec{r}, t) \\
& - (k_{\text{on}}^* + \lambda + k_{\text{off}}) n_A(t) n_B(t) Y(\vec{r}, t). \quad (2.45)
\end{aligned}$$

On the other hand, differentiating Equation (2.30):

$$\frac{\partial \rho_{1,1}}{\partial t} = n_A(t) n_B(t) \frac{\partial Y(\vec{r}, t)}{\partial t} + n_A(t) Y(\vec{r}, t) \frac{dn_B(t)}{dt} + n_B(t) Y(\vec{r}, t) \frac{dn_A(t)}{dt}. \quad (2.46)$$

We can substitute Equations (2.38), (2.39), and (2.45) into Equation (2.46) and, after simplification and solving for $\partial Y(\vec{r}, t)/\partial t$, we are left with the following dynamic equation:

$$\begin{aligned}
\frac{\partial}{\partial t} Y(\vec{r}, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(\vec{r}, t) - Y(\vec{r}, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(\vec{r}, t) - Y(\vec{r}, t)] \\
& - k_{\text{cat}}^* \sigma(\vec{r}) Y(\vec{r}, t) \\
& - k_{\text{cat}}^* n_B(t) Y(\vec{r}, t) \int \sigma(\vec{r}') Y(\vec{r}', t) [X_B(\vec{r}' - \vec{r}, t) - 1] d^d \vec{r}'. \quad (2.47)
\end{aligned}$$

The same process is applied to the general equation for two A particles and then for two B particles, using Equations (2.24) to (2.30), their differentials, (2.38), (2.39), and Kirkwood's approximation where needed. In the end, the full set of dynamic equations for a d -dimensional space is:

$$\frac{dn_A(t)}{dt} = -k_{\text{cat}}^* n_A(t) n_B(t) \int \sigma(\vec{r}') Y(\vec{r}', t) d^d \vec{r}' + k_{\text{off}} n_B(t) - (\lambda + k_{\text{on}}^*) n_A(t), \quad (2.48)$$

$$\frac{dn_B(t)}{dt} = k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t), \quad (2.49)$$

$$\begin{aligned} \frac{\partial}{\partial t} Y(\vec{r}, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(\vec{r}, t) - Y(\vec{r}, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(\vec{r}, t) - Y(\vec{r}, t)] \\ & - k_{\text{cat}}^* \sigma(\vec{r}) Y(\vec{r}, t) \\ & - k_{\text{cat}}^* n_B(t) Y(\vec{r}, t) \int \sigma(\vec{r}') Y(\vec{r}', t) [X_B(\vec{r}' - \vec{r}, t) - 1] d^d \vec{r}', \end{aligned} \quad (2.50)$$

$$\begin{aligned} \frac{\partial}{\partial t} X_A(\vec{r}, t) = & 2k_{\text{off}} \frac{n_B(t)}{n_A(t)} [Y(\vec{r}, t) - X_A(\vec{r}, t)] \\ & - 2k_{\text{cat}}^* n_B(t) X_A(\vec{r}, t) \\ & \cdot \int \sigma(\vec{r}') Y(\vec{r}', t) [Y(\vec{r}' - \vec{r}, t) + Y(\vec{r}' + \vec{r}, t) - 2] d^d \vec{r}', \end{aligned} \quad (2.51)$$

$$\frac{\partial}{\partial t} X_B(\vec{r}, t) = 2k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [Y(\vec{r}, t) - X_B(\vec{r}, t)]. \quad (2.52)$$

Equations (2.48) to (2.52) constitute our general multicentre particle density PDE model, which we refer to as the MPDPDE model³.

2.3.3 Solving the MPDPDE model

For its numerical solution we will try to write down Equations (2.48) to (2.52) as explicitly as possible for the three-dimensional case. We use spherical coordinates due to the natural rotational symmetries (most of the terms only depend on the

³This MPDPDE model, of Chapter 2, is not to be confused with the one to be developed in Chapter 3 of the same name but never referred to in this chapter

absolute distance between particles):

$$\begin{aligned} \frac{dn_A(t)}{dt} = & -k_{\text{cat}}^* n_A(t) n_B(t) \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} \sigma(\vec{r}') Y(\vec{r}', t) r'^2 \sin \theta \, dr' \, d\theta \, d\phi \\ & + k_{\text{off}} n_B(t) - (\lambda + k_{\text{on}}^*) n_A(t), \end{aligned} \quad (2.53)$$

$$\frac{dn_B(t)}{dt} = k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t), \quad (2.54)$$

$$\begin{aligned} \frac{\partial}{\partial t} Y(\vec{r}, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(\vec{r}, t) - Y(\vec{r}, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(\vec{r}, t) - Y(\vec{r}, t)] \\ & - k_{\text{cat}}^* \sigma(\vec{r}) Y(\vec{r}, t) \\ & - k_{\text{cat}}^* n_B(t) Y(\vec{r}, t) \\ & \cdot \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} \sigma(\vec{r}') Y(\vec{r}', t) [X_B(\vec{r}' - \vec{r}, t) - 1] r'^2 \sin \theta \, dr' \, d\theta \, d\phi, \end{aligned} \quad (2.55)$$

$$\begin{aligned} \frac{\partial}{\partial t} X_A(\vec{r}, t) = & 2k_{\text{off}} \frac{n_B(t)}{n_A(t)} [Y(\vec{r}, t) - X_A(\vec{r}, t)] \\ & - 2k_{\text{cat}}^* n_B(t) X_A(\vec{r}, t) \\ & \cdot \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} \sigma(\vec{r}') Y(\vec{r}', t) [Y(\vec{r}' - \vec{r}, t) - 1] r'^2 \sin \theta \, dr' \, d\theta \, d\phi, \end{aligned} \quad (2.56)$$

$$\frac{\partial}{\partial t} X_B(\vec{r}, t) = 2k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [Y(\vec{r}, t) - X_B(\vec{r}, t)], \quad (2.57)$$

where r' and r are only the radial coordinate and $\vec{r}'$ and $\vec{r}$ are the three dimensional vectors.

The pair correlation functions X_A , X_B , and Y , as well as the interaction kernel σ , only depend on the absolute distance between particles, so that, for example, $X_A(\vec{r}, t) = X_A(r, t)$ and $X_A(\vec{r}' - \vec{r}, t) = X_A(|\vec{r}' - \vec{r}|, t)$. With this in mind, we can use a variable change for the integrals in $\partial Y(\vec{r}, t)/\partial t$ and $\partial X_A(\vec{r}, t)/\partial t$: $q = |\vec{r}' - \vec{r}|$,

which is:

$$q = \sqrt{r'^2 + r^2 - 2rr' \cos \theta}. \quad (2.58)$$

This finally leads to:

$$\begin{aligned} \frac{dn_A(t)}{dt} = & -4\pi k_{\text{cat}}^* n_A(t) n_B(t) \int_0^\infty r'^2 \sigma(r') Y(r', t) dr' \\ & + k_{\text{off}} n_B(t) - (\lambda + k_{\text{on}}^*) n_A(t), \end{aligned} \quad (2.59)$$

$$\begin{aligned} \frac{\partial}{\partial t} Y(r, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(r, t) - Y(r, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(r, t) - Y(r, t)] \\ & - k_{\text{cat}}^* \sigma(r) Y(r, t) \\ & - \frac{2\pi k_{\text{cat}}^* n_B(t)}{r} Y(r, t) \\ & \cdot \int_0^\infty r' \sigma(r') Y(r', t) \int_{|r'-r|}^{|r'+r|} q [X_B(q, t) - 1] dq dr', \end{aligned} \quad (2.60)$$

$$\begin{aligned} \frac{\partial}{\partial t} X_A(r, t) = & 2k_{\text{off}} \frac{n_B(t)}{n_A(t)} [Y(r, t) - X_A(r, t)] \\ & - \frac{4\pi k_{\text{cat}}^* n_B(t)}{r} X_A(r, t) \\ & \cdot \int_0^\infty r' \sigma(r') Y(r', t) \int_{|r'-r|}^{|r'+r|} q [Y(q, t) - 1] dq dr'. \end{aligned} \quad (2.61)$$

When non-dimensionalizing, we can group the biophysical parameters into convenient parameter groupings for fitting purposes with the following definitions:

$$\begin{aligned} p_1 &= k_{\text{on}}^* &= k_{\text{on}}[\text{SHP-1}], \\ p_2 &= k_{\text{off}} &= k_{\text{off}}, \\ p_3 &= k_{\text{cat}}^* A_T &= k_{\text{cat}}^* (\text{tethered}) A_T, \\ p_4 &= \frac{k_{\text{cat}}^*}{L^3} &= \frac{k_{\text{cat}}^* (\text{tethered})}{L^3}, \\ p_5 &= \lambda &= k_{\text{cat}} (\text{solution}) [\text{SHP-1}], \end{aligned}$$

with A_T being the initial concentration of A .

The numerical solution was obtained using MATLAB's ode45 for the coupled differential equations, and the trapezoidal rule for the integrals, which states:

$$\int_a^b f(x) dx \approx (b - a) \left[\frac{f(a) + f(b)}{2} \right]. \quad (2.62)$$

Because of this trapezoidal rule, the radial coordinate had to be divided into small intervals of size dr .

Since the integrals for r' go to infinity we need to truncate them. This means that interactions between particles at a distance longer than the truncation distance will not be taken into account. We find an r_{max} for which the particles are so far away from each other that their interaction can be thought of as null. For this we will use the probability density $\sigma(r)$. We will assume that for a very small ϵ , there exists an r_ϵ such that if $\sigma(r_\epsilon) \leq \epsilon$ then $r_{max} = r_\epsilon$. This means:

$$r_{max} = \sqrt{(L_A^2 + L_B^2) \log \left[\frac{3}{2\pi(L_A^2 + L_B^2)\epsilon^{2/3}} \right]}. \quad (2.63)$$

This equation is derived by setting $\sigma(r) = \epsilon$ in Equation (2.6) and solving for r .

An initial analysis would suggest that this truncation would overestimate $dn_A(t)/dt$ because the truncation would reduce the first term on the right hand side of Equation (2.59), which is being subtracted. As $X_B(q, t)$ and $Y(q, t)$ are smaller than 1 most of the time, $[X_B(q, t) - 1]$ and $[Y(q, t) - 1]$ are negative most of the time, effectively making the last term on the right hand side of Equations (2.60) and (2.61) an addition. This suggests that $\partial Y(r, t)/\partial t$ and $\partial X_A(r, t)/\partial t$ will be underestimated because of the truncation that would reduce these last terms in Equations (2.60) and (2.61). But because Equations (2.54), (2.57), and (2.59) to (2.61) are coupled and must be solved simultaneously, it is difficult to say whether the truncation over- or under-estimates any function, and it probably changes over r and t .

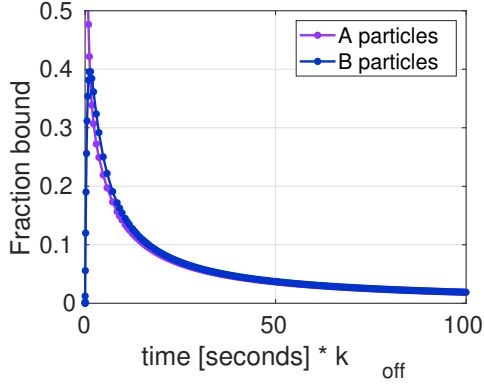
The numerical solution of the differential equations for the pair correlation func-

tions, Equations (2.57), (2.60) and (2.61), also requires the radial coordinate to be divided into small intervals. For simplicity, we used the same discretisation. The size of dr was chosen such that using a smaller interval than the chosen dr would produce a negligible difference on the results for n_A and n_B . Now, instead of one differential equation for $X_A(r, t)$, we will have an infinite number of differential equations: $X_A(0, t)$, $X_A(dr, t)$, $X_A(2dr, t)$, $X_A(3dr, t)$, ..., $X_A(\infty, t)$. The same applies for $X_B(r, t)$ and $Y(r, t)$. Because we cannot numerically calculate an infinite integral, we choose to truncate this expansion as well and, for simplicity, use the same r_{max} as the one used for the trapezoidal rule.

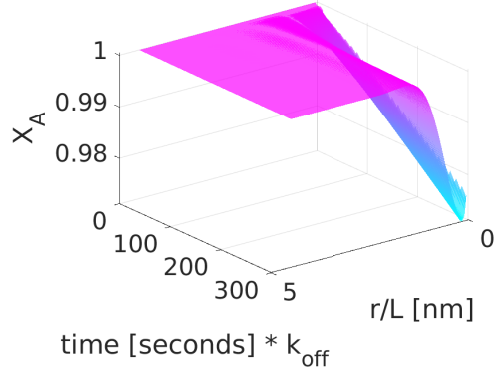
Numerical tests were done on the effective r_{max} such that using a larger r_{max} had no significant impact on the results. This r_{max} was found to be $3\sqrt{L_A^2 + L_B^2}$, so we used $4.5\sqrt{L_A^2 + L_B^2}$ to be safe. For example, for this r_{max} and a plausible case of $L_A \approx L_B \approx 10$ nm, $\sigma(r_{max}) = 7.5 \times 10^{-18}$ nm $^{-1/3}$, which is astronomically low. To put that into perspective, the distance in Matlab from 1.0 to the next larger double-precision number is 2.22×10^{-16} , two orders of magnitude larger than $\sigma(r_{max})$.

We now have all the ingredients we require for the numerical solution. By discretizing the integral PDEs, we are able to convert them into a system of coupled ODEs that are solved in Matlab. We can see an example of the solution in Figure 2.13a. We also see the pair correlation functions in the other panels of Figure 2.13. We show the time evolution of each function in each r interval, making each pair correlation function a three-dimensional solution. Because L_A and L_B are indistinguishable in all equations used, in future calculations we have chosen to assume

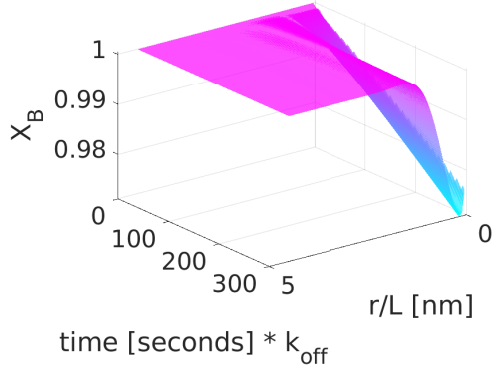
them of equal length: $L_A = L_B = L$.



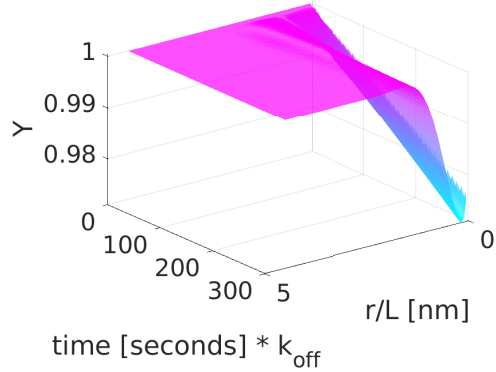
(a) A and B particles.



(b) X_A



(c) X_B



(d) Y

Figure 2.13: An example of the numerical solution to Equations (2.54), (2.57), and (2.59) to (2.61) using Equations (2.6) and (2.63) with $\epsilon = 10^{-8}$. Shown are the solutions to n_A and n_B , and each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in blue, to the furthest $r/L = 4.5$ in magenta. Hence there are 90 solutions to each function. Numerically solved by MATLAB's ode23. Parameter values are: $k_{\text{on}}^* = 1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 1 \text{ } \mu\text{M/s}$, $\lambda = 0.001 \text{ s}^{-1}$, $L_A = L_B = 10 \text{ nm}$. The initial conditions are $n_A(0) = 1$, $n_B(0) = \text{eps}$ (where eps is the smallest possible number different from 0 in the computer, to avoid singularities, which is 2.2204×10^{-16}), and $X_A(r, 0) = X_B(r, 0) = Y(r, 0) = 1$ (assuming there is no correlation to start with because of homogeneity).

2.4 Comparison between the MPDPDE and stochastic simulations with periodic boundary conditions

We have seen that our stochastic simulations fit the experimental data, so we will now use those simulations to test our numerical solutions.

The MPDPDE model does not include boundary conditions and assumes an infinite space that cannot be recreated in our computational stochastic model. Thus the stochastic simulations use a domain as large as possible and periodic boundary conditions. The assumption in both the stochastic and deterministic models, that the domain is effectively infinite, is reasonable because the reaction length scale, L , is much smaller than the dimensions of the dextran matrix (the physical space) where the reactions take place. The typical values for L are around 10 nm, whereas the dimensions of the matrix are approximately 20 nm in height with length and width much larger than 1000 nm. Nevertheless, this boundary effect is tested in Section 2.5.

After testing our model against the stochastic simulations with different parameters, we found similar results to those seen in Figure 2.14. Although different parameter values have been randomly tested and results are very similar to those shown, a more rigorous testing would be needed throughout the parameter range to ensure the numerical solution of the MPDPDE model is a good approximation of the stochastic simulations, but this is beyond the scope of this thesis.

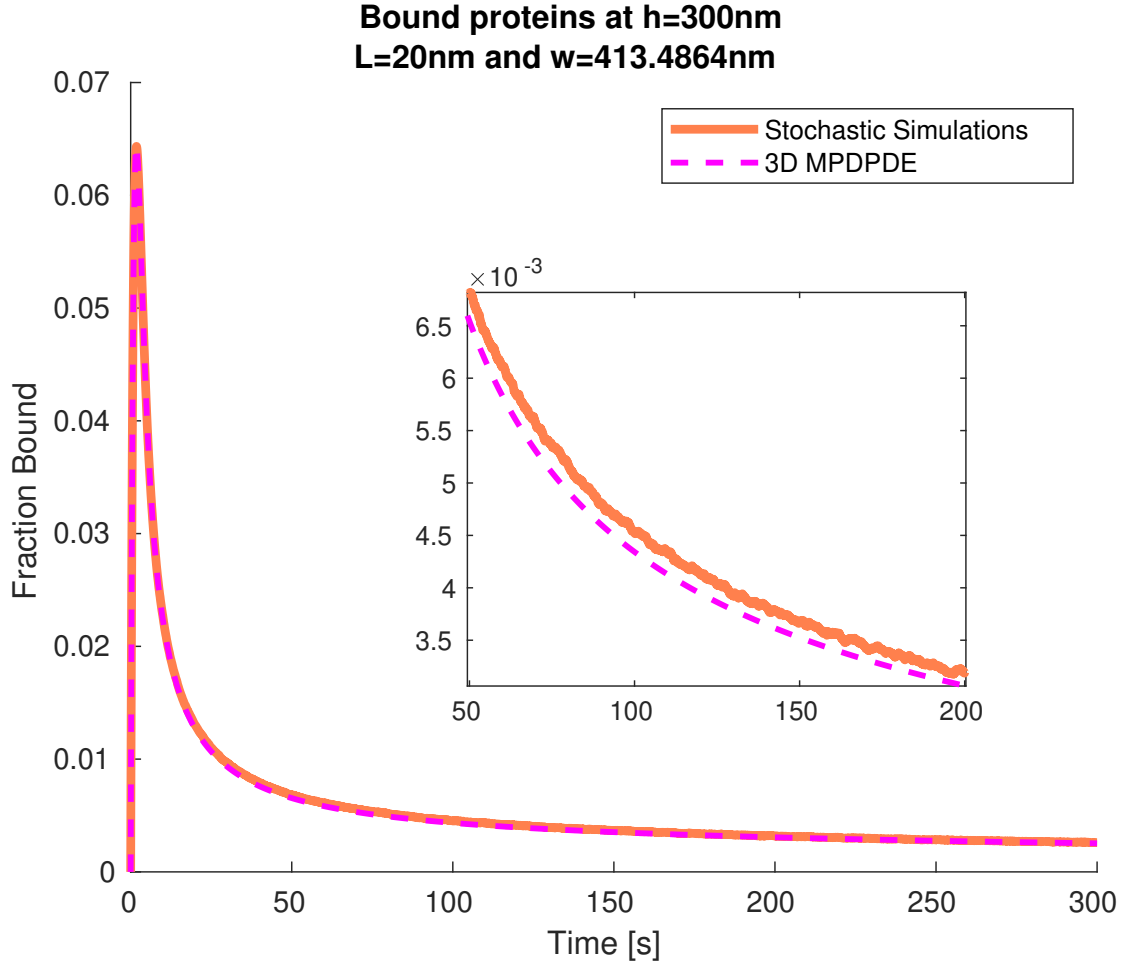


Figure 2.14: Comparing our MPDPDE model to Dr. Isaacson’s stochastic simulations with periodic boundary conditions. A total number of $N = 15000$ tethered proteins in a $300 \times 413.4864 \times 413.4864$ nm³ cube was used for the simulations, presented by the continuous orange line. The average of a 1000 simulations is used. Simulations and the MPDPDE solution in dashed magenta were produced with $k_{\text{on}} = 0.1$ s⁻¹, $k_{\text{off}} = 1$ s⁻¹, $k_{\text{cat}} = 0.01$ μMs^{-1} , $L = 20$ nm, and $\lambda = 0.0005$. Inset: Zoom in to seconds 50 to 200, to show difference between models.

We find that the numerical solution of the MPDPDE model is always slightly below the stochastic simulation solution (of the order of 10^{-4}). Possible sources for this systematic error could be the inevitable interaction with the boundaries that the stochastic simulations have, or the Kirkwood approximation in Section 2.3.2. Different geometries are used in Section 2.5 to study the effect the boundaries could

be having on the results.

To see the effect of the Kirkwood approximation we could exchange it for another closure approximation where three-centre correlation functions are accepted and four-centre correlation functions are decoupled into the product of four three-centre correlation functions. Whether this higher-order approximation still showed a systematic error and how much smaller this systematic error becomes would give us an idea of the influence of the Kirkwood approximation. This has not been done because it would produce a much larger set of equations that would be too cumbersome to write down and would be equally computationally intensive, whereas we consider that the trade-off of solving-time versus the accuracy we have with the MPDPDE model is good enough for experimental testing.

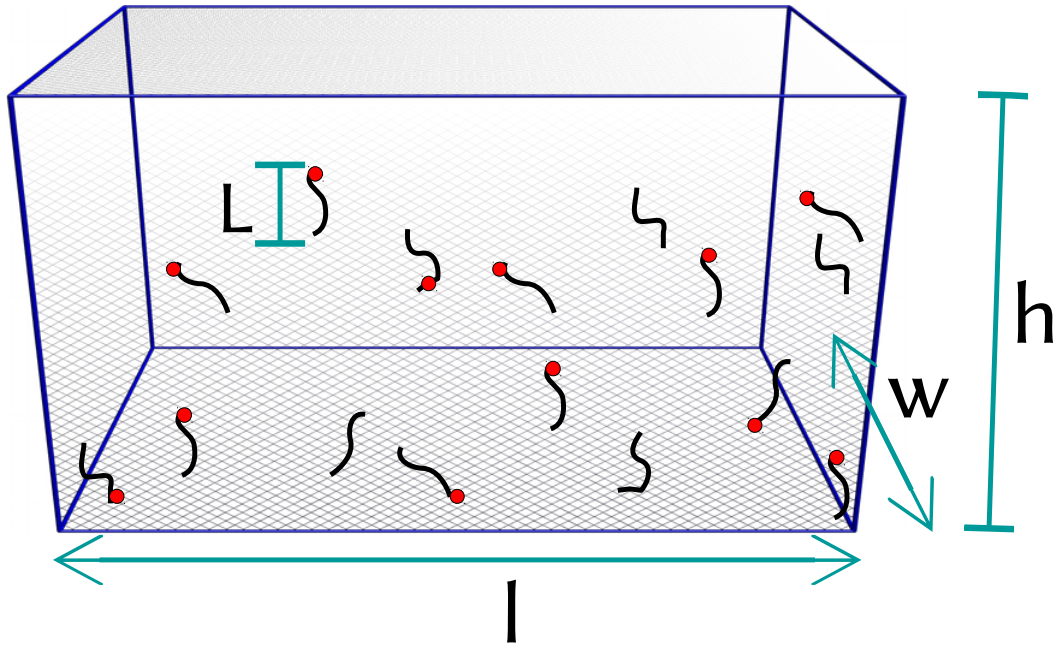
2.5 Geometry effect: two- and three- dimensions

As stated in Section 2.4, our MPDPDE deterministic model assumes an infinite domain, not taking into account boundaries. This is not the case for the experiments or the stochastic simulations, and the boundaries in these could be playing a role in the difference found in the results between these and our deterministic model. On the other hand, the interior of the dextran matrix might be causing an irregular flow of the injected SHP-1 to which the proteins are exposed, because of its size and web-like structure. Furthermore, the laboratory we are working with has the possibility of doing experiments on a different flow chamber where there is no dextran matrix and the peptides are directly immobilised to a 2D surface. We will therefore now adapt the three-dimensional MPDPDE model that was previ-

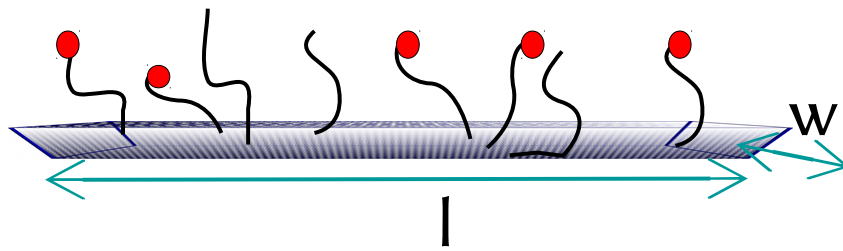
ously solved to a two-dimensional MPDPDE model. The idea is that we will deal with the boundaries and irregular flow, comparing the two- and three-dimensional MPDPDE models with similar two- and three-dimensional stochastic simulations. We hypothesise that the larger the volume in the three directions, the less the boundaries would make a difference, hence testing the “boundaries problem”. Also, there is a larger difference between a three-dimensional model without boundaries and its corresponding simulation, than a two-dimensional model without boundaries and its corresponding simulation, so comparing the differences will give us an idea of how much of a problem it is to compare our infinite model with the bounded system. On the other hand, the irregular flow problem would be addressed in the two-dimensional experiments. In the three-dimensional matrix of the experiment, most of the binding endings of A would lie within the gel matrix, exposing them to the turbulent flow that would be generated within, and only the few A molecules in the boundaries would be exposed to a more regular flow of SHP-1. On the other hand, with a flatter, two-dimensional “pill box” case, most of the binding endings of A would lie outside the gel, being exposed to a more regular flow. Finally, we will discern if the two-dimensional MPDPDE model would be better for the possible laboratory experiment, and determine the heights for which we need to switch from one model to the other.

For this comparison of geometries, experimental assays were carried out in two different volumes. One was the original three-dimensional gel matrix, where the height of the domain is much larger than the length of the molecules being studied. A representation of this can be seen in Figure 2.15a. The other one was a much

flatter domain, which we call the two-dimensional domain, or “pill box”, where the length and width are still much larger than the length of B , but the height is of the order of magnitude of the molecules’ length or smaller. This is represented in Figure 2.15b.



(a) Three-dimensional representation. Height of gel matrix, h , is much larger than the proteins, i.e. $L \ll l, w, h$.



(b) Two-dimensional representation, or “pill box”. Height of matrix is almost 0, so $L \ll l, w$ but $L \sim h$.

Figure 2.15: Difference between the two- and three- dimensional representations of the gel matrix for the laboratory experiments.

We needed to match the stochastic simulations and the MPDPDE model to the laboratory experiments. To match the experiments and test the effect of this transition from three dimensions to two dimensions, we carried out stochastic simulations with varying heights. Because the laboratory experiments were made in domains which were much larger in height, width, and length than could be recreated in our computational simulations because of time requirements and limits on computational memory ⁴, periodic boundary conditions in the three directions were implemented. To match the stochastic model, we solved our general MPDPDE model, Equations (2.48) to (2.52), in two dimensions, such that we had a three-dimensional and a two-dimensional model to test against the stochastic simulations and the experimental data.

2.5.1 Solving the MPDPDE model in 2D

For the case where the experiments are done in such thin plates, Equations (2.48) to (2.52) had to be re-evaluated for the appropriate case, choosing the variable $q = |\vec{r} - \vec{r}'| = \sqrt{r^2 + r'^2 - 2rr' \cos(\theta)}$ and assuming, without loss of generality, that $\vec{r} = r\hat{x}$, yielding:

$$\begin{aligned} \frac{dn_A(t)}{dt} = & -2\pi k_{\text{cat}}^* n_A(t) n_B(t) \int_{r'=0}^{\infty} \sigma(r') Y(r', t) r' dr' \\ & + k_{\text{off}} n_B(t) - (\lambda + k_{\text{on}}^*) n_A(t), \end{aligned} \quad (2.64)$$

$$\frac{dn_B(t)}{dt} = k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t), \quad (2.65)$$

⁴In typical experiments, a density of A molecules in the matrix of about 6.023×10^{-5} molecules per nm^3 is used. So, to get to length-scales of 10,000 nm, which is the length-scale in the laboratory experiments, we would need about 10^8 particles, which is not feasible, computationally.

$$\begin{aligned}
\frac{\partial}{\partial t} Y(r, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(r, t) - Y(r, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(r, t) - Y(r, t)] \\
& - k_{\text{cat}}^* \sigma(r) Y(r, t) \\
& - 4k_{\text{cat}}^* n_B(t) Y(r, t) \int_{r'=0}^{\infty} r' \sigma(r') Y(r', t) \\
& \cdot \int_{q=|r'-r|}^{|r'+r|} \frac{[X_B(q, t) - 1]q}{\sqrt{[q^2 - (r' + r)^2][q^2 - (r' - r)^2]}} dq dr', \tag{2.66}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial t} X_A(r, t) = & 2k_{\text{off}} \frac{n_B(t)}{n_A(t)} [Y(r, t) - X_A(r, t)] \\
& - 8k_{\text{cat}}^* n_B(t) X_A(r, t) \int_{r'=0}^{\infty} r' \sigma(r') Y(r', t) \\
& \cdot \int_{q=|r'-r|}^{|r'+r|} \frac{[Y(q, t) - 1]q}{\sqrt{[q^2 - (r' + r)^2][q^2 - (r' - r)^2]}} dq dr', \tag{2.67}
\end{aligned}$$

$$\frac{\partial}{\partial t} X_B(r, t) = 2k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [Y(r, t) - X_B(r, t)]. \tag{2.68}$$

Rescaling the variables as follows: n_A by A_T (where A_T is the initial number of A molecules), n_B by A_T as well, and r by L (where L is the length of A and B , as seen at the end of Section 2.3.3) leads to the following system:

$$\begin{aligned}
\frac{dn_A(t)}{dt} = & - \frac{2\pi k_{\text{cat}}^* A_T n_A(t) n_B(t)}{L} \left(\frac{3}{2\pi} \right)^{3/2} \int_{r'=0}^{\infty} e^{-\frac{3r'^2}{2}} Y(r', t) r' dr' \\
& + k_{\text{off}} n_B(t) - (\lambda + k_{\text{on}}^*) n_A(t), \tag{2.69}
\end{aligned}$$

$$\frac{dn_B(t)}{dt} = k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t), \tag{2.70}$$

$$\begin{aligned}
\frac{\partial}{\partial t} Y(r, t) = & k_{\text{off}} \frac{n_B(t)}{n_A(t)} [X_B(r, t) - Y(r, t)] + k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [X_A(r, t) - Y(r, t)] \\
& - \frac{k_{\text{cat}}^*}{L^3} \left(\frac{3}{2\pi} \right)^{3/2} e^{-\frac{3r^2}{2}} Y(r, t) \\
& - \frac{4k_{\text{cat}}^* A_T n_B(t)}{L} \left(\frac{3}{2\pi} \right)^{3/2} Y(r, t) \int_{r'=0}^{\infty} r' e^{-\frac{3r'^2}{2}} Y(r', t) \\
& \cdot \int_{q=|r'-r|}^{|r'+r|} \frac{[X_B(q, t) - 1]q}{\sqrt{[q^2 - (r' + r)^2][q^2 - (r' - r)^2]}} dq dr', \tag{2.71}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial t} X_A(r, t) = & 2k_{\text{off}} \frac{n_B(t)}{n_A(t)} [Y(r, t) - X_A(r, t)] \\
& - \frac{8k_{\text{cat}}^* A_T n_B(t)}{L} \left(\frac{3}{2\pi} \right)^{3/2} X_A(r, t) \int_{r'=0}^{\infty} r' e^{-\frac{3r'^2}{2}} Y(r', t) \\
& \cdot \int_{q=|r'-r|}^{|r'+r|} \frac{[Y(q, t) - 1]q}{\sqrt{-(q^2 - (r' + r)^2)[q^2 - (r' - r)^2]}} dq dr', \tag{2.72}
\end{aligned}$$

$$\frac{\partial}{\partial t} X_B(r, t) = 2k_{\text{on}}^* \frac{n_A(t)}{n_B(t)} [Y(r, t) - X_B(r, t)]. \tag{2.73}$$

We can group the biophysical parameters into convenient parameters for fitting purposes with the following definitions:

$$\begin{aligned}
p_1 &= k_{\text{on}}^* &= k_{\text{on}}[\text{SHP-1}], \\
p_2 &= k_{\text{off}} &= k_{\text{off}}, \\
p_3 &= \frac{k_{\text{cat}}^* A_T}{L} &= \frac{k_{\text{cat}}^*(\text{tethered}) A_T}{L}, \\
p_4 &= \frac{k_{\text{cat}}^*}{L^3} &= \frac{k_{\text{cat}}^*(\text{tethered})}{L^3}, \\
p_5 &= \lambda &= k_{\text{cat}}(\text{solution})[\text{SHP-1}].
\end{aligned}$$

We note that these parameters are very similar to the ones used for the three-dimensional solution, since p_1 , p_2 , p_4 , and p_5 are the same. It is only p_3 that differs by having an extra $1/L$ for this three-dimensional case. These definitions result in a fully non-dimensional MPDPDE system.

Numerical evaluation of the integrals is complicated because of singularities at the integration limits. To obtain a numerical discretization without blow up, we first separate the inner integral into two and then use a mid-point rule. This method consists of decoupling the integrals into a part that can be evaluated numerically by discretization, and a part that can be evaluated analytically, and will not blow up

in the limits. This works as follows: Splitting the integral produces:

$$\begin{aligned}
I &= \int_{|r'-r|}^{|r'+r|} \frac{[X_B(q, t) - 1]q}{\sqrt{-(q^2 - (r' + r)^2)[q^2 - (r' - r)^2]}} dq \\
&= \int_{|r'-r|}^{\eta} \left[\frac{[X_B(q, t) - 1]q}{\sqrt{-(q^2 - (r' + r)^2)[q^2 - (r' - r)^2]}} \right] dq \\
&+ \int_{\eta}^{|r+r'|} \left[\frac{[X_B(p, t) - 1]p}{\sqrt{-(p^2 - (r' + r)^2)[p^2 - (r' - r)^2]}} \right] dp.
\end{aligned}$$

η is chosen so that it is away from the integration limits (e.g. $\eta = (|r-r'| + |r+r'|)/2$).

Next we discretize the inner integral using the mid-point rule:

$$\begin{aligned}
I &\approx \sum_{i=1}^N \frac{[X_B(q_{i+1/2}, t) - 1]}{\sqrt{-(q_{i+1/2}^2 - (r' + r)^2)}} \int_{q_i}^{q_{i+1}} \frac{q}{\sqrt{q^2 - (r' - r)^2}} dq \\
&+ \sum_{i=1}^N \frac{[X_B(p_{i+1/2}, t) - 1]}{\sqrt{-(p_{i+1/2}^2 - (r' - r)^2)}} \int_{p_i}^{p_{i+1}} \frac{p}{\sqrt{p^2 - (r' + r)^2}} dp \\
&\approx \sum_{i=1}^N \frac{[X_B(q_{i+1/2}, t) - 1]}{\sqrt{-(q_{i+1/2}^2 - (r' + r)^2)}} \left[\sqrt{q_{i+1}^2 - (r' - r)^2} - \sqrt{q_i^2 - (r' - r)^2} \right] \\
&+ \sum_{i=1}^N \frac{[X_B(p_{i+1/2}, t) - 1]}{\sqrt{-(p_{i+1/2}^2 - (r' - r)^2)}} \left[-\sqrt{(r' + r)^2 - p_{i+1}^2} + \sqrt{(r' + r)^2 - p_i^2} \right],
\end{aligned}$$

with $q_i = i\Delta r + |r - r'|$, $q_N = \eta$, $p_i = i\Delta r + \eta$, $p_N = |r + r'| - \Delta r$, $q_{i+1/2}$ is the middle point between q_i and q_{i+1} , and $p_{i+1/2}$ is the middle point between p_i and p_{i+1} . We note that $q_{N+1} = (N + 1)\Delta r + |r - r'|$. We redefine indices so that X_B is evaluated at whole Δr intervals, making $q_{i+1/2} = q_i$ and $q_i = q_{i-1/2}$, and the same for p . We shift the summation, from $\hat{i} = 1/2$ to $\hat{i} = N - 1/2$, to $\hat{i} = 1$ to $\hat{i} = N$. We drop i 's

that for convenience, hence:

$$I = \sum_{i=1}^N \frac{[X_B(q_i, t) - 1]}{\sqrt{(r' + r)^2 - q_i^2}} \left[\sqrt{q_{i+1/2}^2 - (r' - r)^2} - \sqrt{q_{i-1/2}^2 - (r' - r)^2} \right] \\ + \sum_{i=1}^N \frac{[X_B(p_i, t) - 1]}{\sqrt{p_i^2 - (r' - r)^2}} \left[\sqrt{(r' + r)^2 - p_{i-1/2}^2} - \sqrt{(r' + r)^2 - p_{i+1/2}^2} \right]. \quad (2.74)$$

Everything is analogous for the integrals with Y instead of X_B in $\partial X_A / \partial t$ in Equation (2.67). We note that $\sigma(r)$ will be the same in the two-dimensional case as in the three-dimensional case, as the protein chain can still move within the upper half of a sphere, in the three-dimensional space above the experimental plane. So, the protein will move in hemisphere but will spend twice as much time in any point of this hemisphere.

Once again, for the numerical solution MATLAB's ode45 was used for the coupled differential equations, and the trapezoidal rule for the remaining integrals.

2.5.2 Comparing models

We compare here our MPDPDE model with the results from the stochastic simulations. To make a useful comparison, the program that produced the stochastic simulations was adapted by Dr. Samuel Isaacson such that it could not only produce simulations in three dimensions, but also in a two-dimensional plate with $h = 0$ nm. Periodic boundary conditions were used in both the three-dimensional stochastic simulations (3DSS) and on the two-dimensional stochastic simulations (2DSS) to resemble the infinite space of our MPDPDE as closely as possible.

Restrictions on the size of the domain of the stochastic simulations were imposed

by the number of proteins the computer's memory could handle and our required density. The density in experiments is typically 5×10^{-5} of A per nm^3 , or about 9 molecules within a sphere of radius $L = 30$ nm. This meant there were 9 proteins per sphere of radius L in the three-dimensional case, or per circle of radius L in the two-dimensional case. This would have meant that the number of proteins, N , that we would use would be given by:

$$N_{3D} = \rho_{3D}(hlw) = \frac{9}{4/3\pi L^3}(hlw), \quad (2.75)$$

$$N_{2D} = \rho_{2D}(lw) = \frac{9}{\pi L^2}(lw), \quad (2.76)$$

for three and two dimensions, respectively.

Because we want to maintain a fixed number of proteins throughout the simulations (the maximum number handled by the computer, around 300,000 particles) and we want to observe what happens when we gradually go from three to two dimensions, it is important that decreasing the height in the 3D simulations converges to the parameters of the 2D simulations so that they can be compared. For a fixed number of proteins, $l \times w$ has to go to infinity as h approaches 0. To address this problem we made a combination of both Equations (2.75) and (2.76):

$$N = \frac{9}{4/3\pi L^3}(hlw) + \frac{9}{\pi L^2}(lw). \quad (2.77)$$

This means that in the extremes where $h \gg L$, N will behave accurately for the three-dimensional case, and it will do so for the two-dimensional case in the extreme where $h = 0$, but will give us a smooth transition. Since we want N to be fixed and can vary the length and width of the domain depending on the desired height,

we can rewrite Equations (2.75), (2.76), and (2.77) in terms of the width. We will assume $l = w$ and rescale all lengths by L , denoting them with a hat:

$$\hat{w}_{3D} = \sqrt{\frac{4N\pi}{27\hat{h}}}, \quad (2.78)$$

$$\hat{w}_{2D} = \sqrt{\frac{N\pi}{9}}, \quad (2.79)$$

$$\hat{w} = \sqrt{\frac{4N\pi}{9(3\hat{h} + 4)}}. \quad (2.80)$$

The comparison between Equations (2.78), (2.79), and (2.80) can be seen in Figure 2.16. We can observe that, had we followed the 2D approach until $\hat{w}_{2D}$ intersects with $\hat{w}_{3D}$, and then followed $\hat{w}_{3D}$, the biggest difference between these and the approximation we used, $\hat{w}$, occurs when $h = 4L/3$. At this point, our approximation has a smaller width and, therefore, a slightly larger concentration than using $\hat{w}_{3D}$.

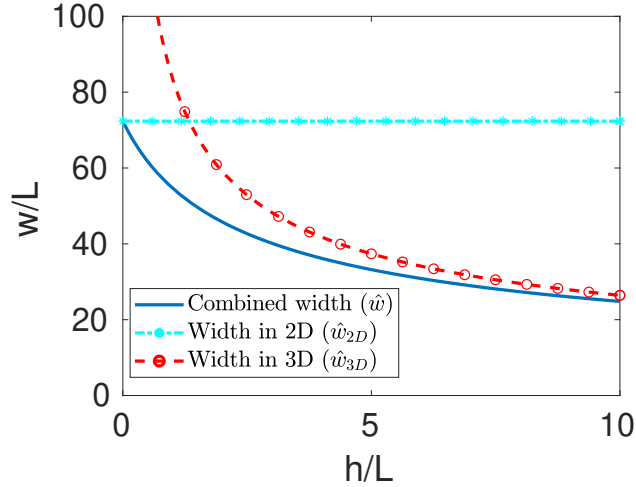


Figure 2.16: Visualisation of how the width (and length) of the domain changes for a constant number of particles N with a three-, two-dimensional treatment, and how our expression in Equation (2.80) goes from one to the other. Width w , length l , and height h are rescaled by proteins length, L .

To verify our MPDPDE model, we first solve the stochastic model with periodic boundary conditions on as large a domain as possible. Given Equation (2.80), this

would be approximately $1000 \times 1000 \times 1000 \text{ nm}^3$ for three dimensions and $5800 \times 5800 \text{ nm}^2$ for two dimensions. We label our two- and three-dimensional MPDPDE models as MAG2D and MAG3D, respectively, and show the results from the three-dimensional case in Figure 2.17, whilst the results from the two-dimensional case are shown in Figure 2.18. Because the biggest two-dimensional domain was larger than needed to see the boundary interactions, and to save computational time, the concentration of molecules (9 proteins per circle of radius L) was kept but the domain was reduced to $4000 \times 4000 \text{ nm}^2$ for Figure 2.18. As stated before, the change between two and three dimensions is given by the height of the matrix in the simulations compared to the length of A and B . For example, because we are using a protein length of 20 nm, we say that an entity similar or smaller in length is considered two-dimensional, while an entity 10 times longer or more is considered three-dimensional. Both figures show the solution to our MPDPDE model with the parameters used for the stochastic simulations, as well as the solution to our MPDPDE model using the parameters found to fit best by MATLAB's `lsqcurvefit`⁵. In both cases the fit between the stochastic simulations and the curve produced by our MPDPDE model with the fitted parameters is remarkable. We cannot guarantee that the fitted results are a global minimum, but assumed the correct minimum was

⁵Matlab's `lsqcurvefit` function is an algorithm for solving least-squares problems. This means that, given observed data $ydata$ and input data $xdata$, the solver searches for the set of parameters x such that

$$\frac{1}{2} \sum_i (F(x, xdata_i) - ydata_i)^2 \quad (2.81)$$

is minimized. F is a function defined by the user, in this case our MPDPDE model. The minimization is done following the method suggested by Coleman and Li [89]. This method is a subspace trust region method [89], which establishes global convergence under reasonable assumptions, and uses an interior-reflective Newton method described in Coleman and Li [90], which minimizes a smooth nonlinear function within a region of its multi-dimensional variable.

achieved because of such good agreement between the stochastic simulations and the results.

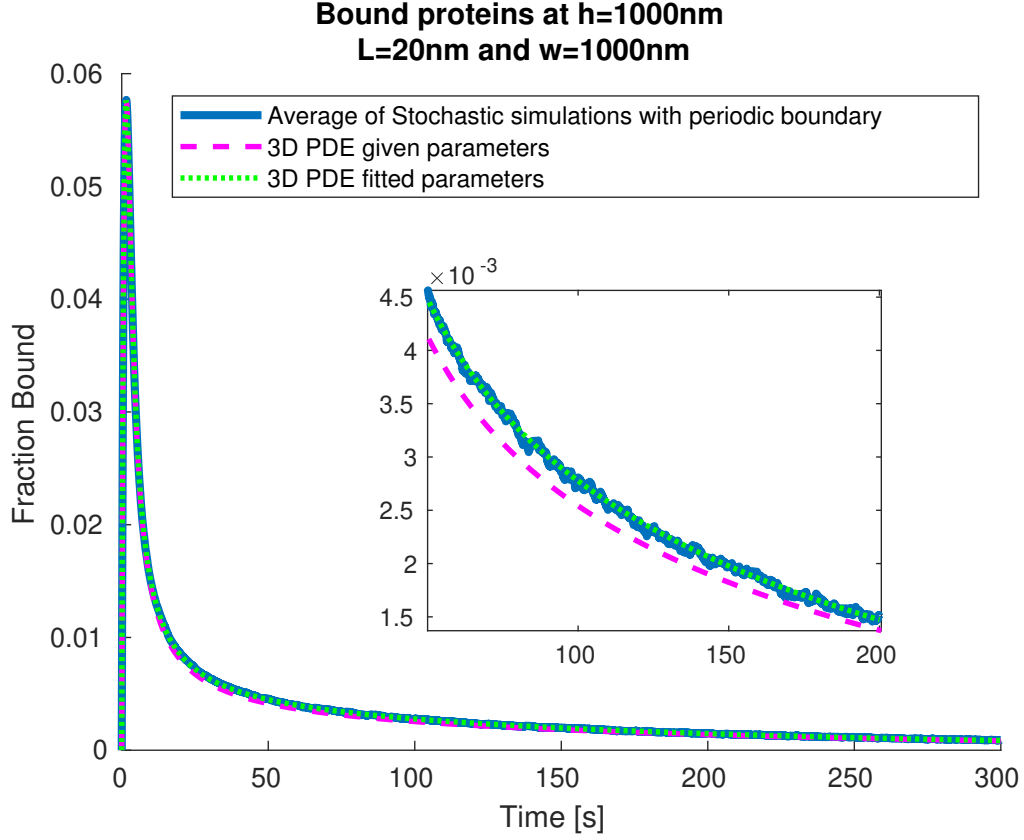


Figure 2.17: Comparing the MAG3D to Dr. Isaacson’s stochastic simulations with periodic boundary conditions. A total number of $N = 268574$ initial A molecules in a $1000 \times 1000 \times 1000 \text{ nm}^3$ cube was used for the stochastic simulations, presented in a continuous blue line. The average of 10 simulations was used. The stochastic simulations and the MAG3D solution in dashed magenta were produced with $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$, whilst the MAG3D solution in dotted green was produced with the fitted parameters $k_{\text{on}} = 0.1023 \text{ s}^{-1}$, $k_{\text{off}} = 1.0388 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0203 \mu\text{Ms}^{-1}$, $L = 18.4154 \text{ nm}$, and $\lambda = 0.0054$, fitted by MATLAB’s `lsqcurvefit`. The initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$. Inset: Zoom in to seconds 50 to 200, to show difference between models.

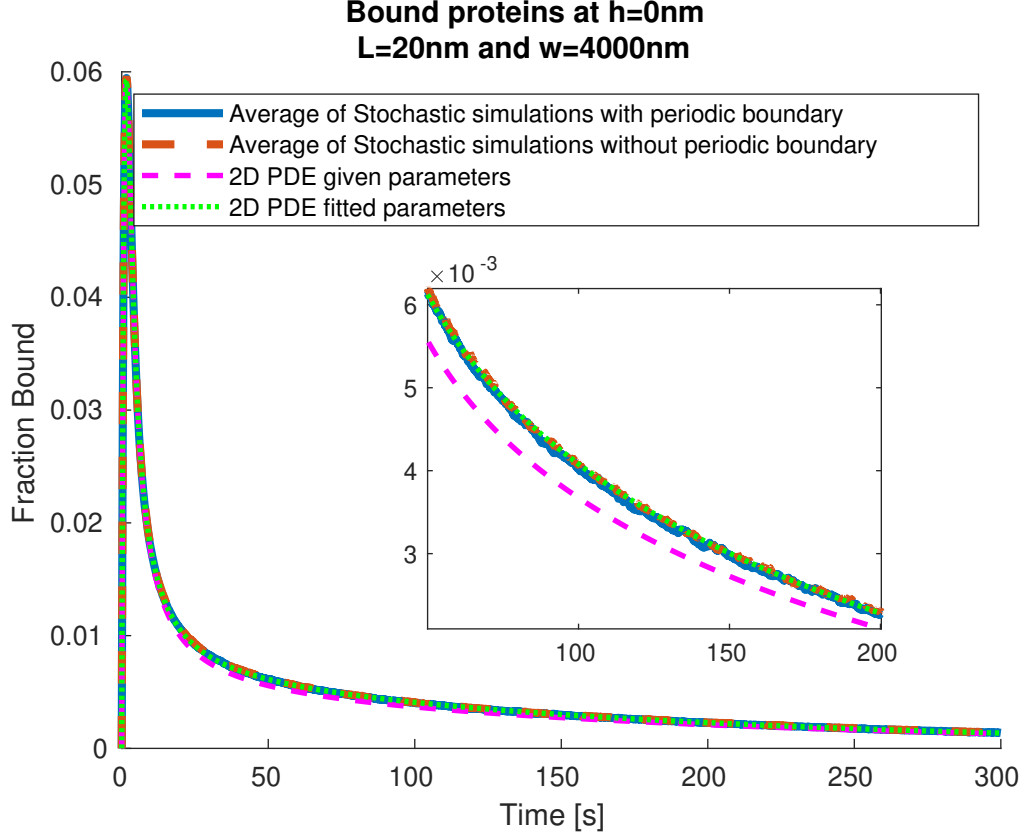


Figure 2.18: Comparing MAG2D to Dr. Isaacson’s stochastic simulations with periodic boundary conditions (blue continuous line). A total number of $N = 114592$ initial A molecules in a $4000 \times 4000 \text{ nm}^2$ square was used for the stochastic simulations. The average of 100 simulations plotted. The stochastic simulations and the MAG2D solution in dashed magenta were produced with $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{M s}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$, whilst the MAG2D solution in dotted green was produced with the fitted parameters $k_{\text{on}} = 0.1020 \text{ s}^{-1}$, $k_{\text{off}} = 1.0254 \text{ s}^{-1}$, $k_{\text{cat}} = 0.0189 \mu\text{M s}^{-1}$, $L = 17.4858 \text{ nm}$, and $\lambda = 0.0054$, fitted by MATLAB’s `lsqcurvefit`. The initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{M s}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005$. The dashed orange line corresponds to stochastic simulations with the same parameter values as the blue continuous line, but without periodic boundary conditions. Inset: Zoom in to seconds 50 to 200, to show difference between models.

Both the three- and the two-dimensional MPDPDE models fit their stochastic models very closely, and have a similar difference in their comparisons, which makes us think that the boundary does not have that large an effect and that assuming infinite space in the MPDPDE model does not impact the results significantly. If

the assumption of infinite space made a noticeable difference, we would expect the difference in the two-dimensional model to be smaller than in the three-dimensional model, which it is not.

2.5.3 MAG3D vs MAG2D

This study will consist of verifying that the 2D PDE model fits the two-dimensional stochastic simulations better than the 3D PDE model, as well as the 3D PDE model fitting better the three-dimensional stochastic simulations than the 2D PDE model. This means there should be a crossover between model use when the height goes from “small” to “large”. We will also see where this crossover happens in terms of height.

Setting the boundary conditions to periodic in the stochastic simulations, we have created simulations with different heights, whilst keeping to $N = 243,507$ tethered proteins, using our width approximation in Expression (2.80). We have then fitted our two- and three-dimensional MPDPDE models, labelled as MAG2D and MAG3D, respectively. The results can be seen in Figure 2.19, whilst numbers can be checked in Table 2.1.

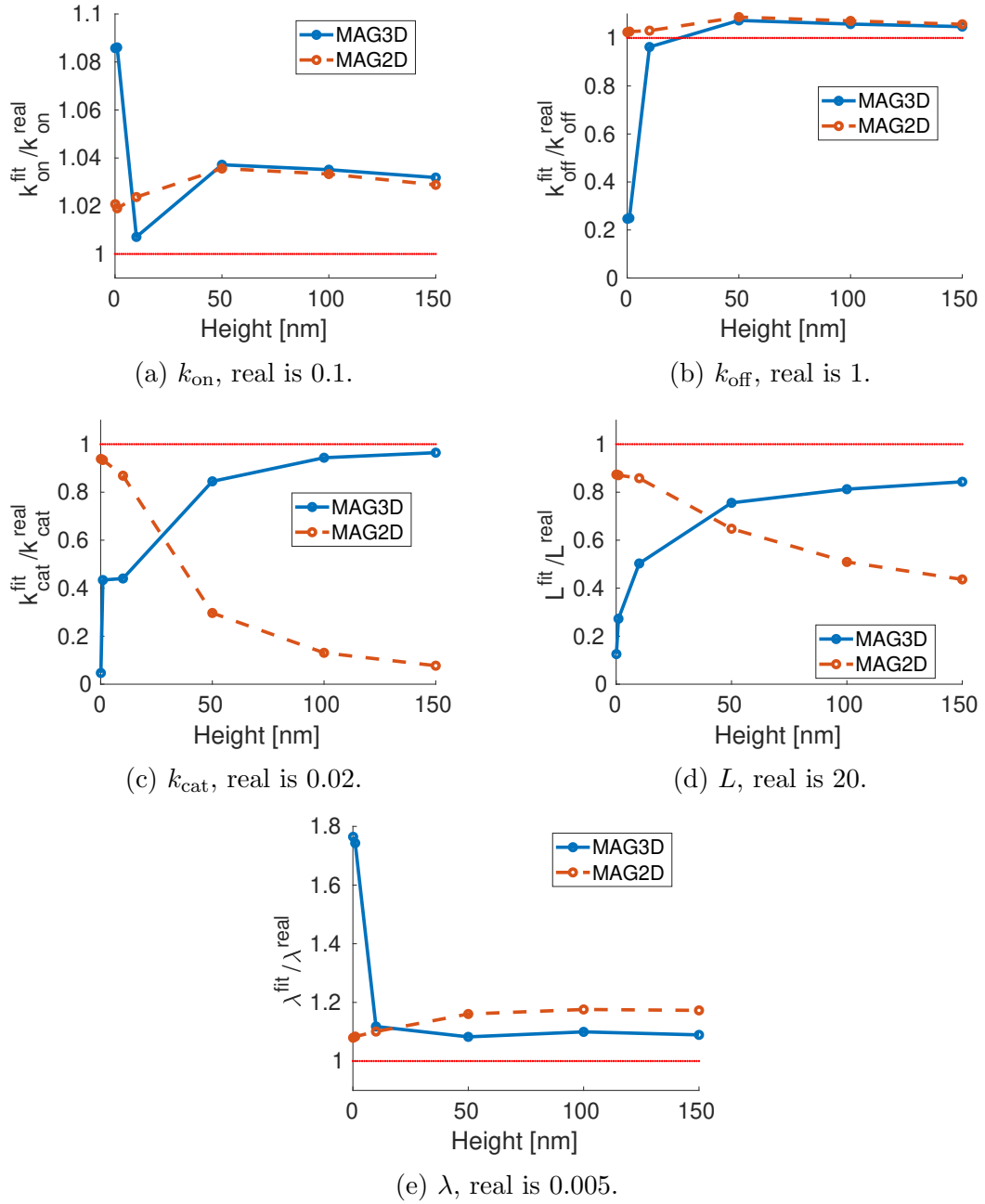


Figure 2.19: Data from a test with $N = 243,507$ initial A molecules. Ratio of fitted parameter value over each parameter value actually used to produce the stochastic simulations (labelled as “real”) at heights $h = 0.1, 1, 10, 50, 100$, and 150 nm. Fitted two- and three-dimensional MPDPDE models using MATLAB’s `lsqcurvefit`. In all cases the initial guess for the fitting function was: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$. MAG2D in blue continuous line and MAG3D in orange dashed line. Corresponding to data in Table 2.1.

Table 2.1: Fitting MAG2D and MAG3D for simulations with varying heights. Maximum error highlighted.

	Height [nm]	Used value	Best fit in MAG2D	Best fit in MAG3D
k_{on}	0.1	0.1	0.1021	0.1086
	1		0.1019	0.1086
	10		0.1024	0.1007
	50		0.1036	0.1037
	100		0.1033	0.1035
	150		0.1029	0.1032
k_{off}	0.1	1	1.0242	0.2467
	1		1.0265	0.2494
	10		1.0305	0.9625
	50		1.0871	1.0731
	100		1.0711	1.0580
	150		1.0570	1.0473
k_{cat} or μ	0.1	0.02	0.0188	0.0009
	1		0.0187	0.0087
	10		0.0174	0.0088
	50		0.0059	0.0169
	100		0.0026	0.0189
	150		0.0015	0.0193
L	0.1	20	17.4678	2.5068
	1		17.4160	5.4565
	10		17.1650	10.0605
	50		12.9564	15.1123
	100		10.1858	16.2505
	150		8.7298	16.8678
λ	0.1	0.005	0.0054	0.0088
	1		0.0054	0.0087
	10		0.0055	0.0056
	50		0.0058	0.0054
	100		0.0059	0.0055
	150		0.0059	0.0054

Ideally, the quotient of the fitted parameter over the real parameter should equal to 1, especially for MAG2D when the height is 0 nm for the stochastic simulations, and for MAG3D when the height is much larger than 20 nm for the stochastic simulations. We can see that, whilst the quotient is not quite 1, MAG2D does better than MAG3D for heights below 25 nm, but by heights of 50 nm the results

have crossed over and MAG3D fares better. Also, the fitted results for the extreme cases (0 nm or larger than 100 nm) are never more than 10% off, often being much closer to the actual, real, parameters used to produce the stochastic simulations.

Experimental assays have been done in a three-dimensional gel matrix as well as in a flat matrix. The robustness of our three-dimensional and two-dimensional MPDPDE models for the dephosphorylation reaction have yet to be tested.

2.6 Conclusions

To better understand the interaction between phosphorylated proteins in the T cell membrane and SHP-1 enzyme, we developed a group of deterministic and stochastic models. These models have a better fit to the experimentally obtained data than the traditional ODE models with mean-field approximation. These models can be used as a proof of concept of what is thought to describe the interaction between the enzyme SHP-1 and phosphorylated tails in the T cell membrane, but the models could also be adapted for other interactions where elements engage with their own docking sites and those of surrounding stationary counterparts.

The whole set of parameters of the *in vitro* experiments is unknown, but we tested the accuracy of our deterministic model against the known parameters of *in silico* generated data. It is noticeable that fitted parameters had a better curve fit than using the known parameters, when testing against *in silico* experiments with known parameters, but the difference was usually smaller than 5% for most parameters. The biggest discrepancy was in the length of the proteins, L , which could be up to 15%. This discrepancy is likely a result of having to approximate

the deterministic model by the use of the Kirkwood approximation. As a result of the complexity of including higher order terms, it was not possible to explicitly test whether this is the case.

Furthermore, we tested the use of the MPDPDE models in two and three dimensions against stochastic simulations for different heights. With this we found, as expected, that it is better to use the two-dimensional numerical model for experiments with heights shorter than the proteins' length, and the three-dimensional numerical model for experiments with heights larger than twice the proteins' length. This suggests that if the fitted reach parameter (L) is on the scale of the SPR chip height (h), then the 2D model may be more appropriate.

These models can hopefully now be used to further study unknown compound reaction rates and lengths of proteins in other experiments. Indeed, two studies have now been published that have utilised the 3D model to fit tethered reactions in SPR (Goyette et al. [1], Clemens et al. [91]). The deterministic models are preferred to the stochastic simulations because of their shorter running time and ease of use. These models could also, in future, be plugged into larger models of protein interaction inside the T cell, to further understand the role of tethering and reach in signalling pathways.

Some of the reasons for the difference between the MPDPDE models and the stochastic simulations could be due to the approximations used. We used the Kirkwood approximation, where we approximate three-particle interaction with the product of all possible two-particle interactions. Whilst this is a commonly used approximation, it could be introducing errors. In future work this could be explored by

not approximating the three-body interaction but the four-body one. This would suppose a larger number of equations and running time, but still finite.

In conclusion, our MPDPDE model can be used to analyse tethered catalytic reactions in SPR and has recently been applied in the study of PD-1 and LAIR-1 tethered SHP-1 reactions [1, 91]. This analysis is non-trivial because it comprises both stochastic and spatial elements, which is why simpler ODE models fail to fit the data. We provide both three- and two-dimensional versions of the model that should be used for their corresponding SPR experiment (i.e. 3D when there is a dextran matrix and 2D when there is no dextran matrix). Our approach may be widely applicable to reactions whereby proteins tether to one molecule but act on another.

In this chapter we have spoken about the importance of protein interaction in the cell membrane and, specifically, of how dephosphorylation of receptor tails could lead to a diminished reaction of the T cell immune system. We focused on enzymes like SHP-1 that bind and dephosphorylate, creating a negative feedback loop, and could also dephosphorylate other tethered proteins nearby. But if we look back to Figure 2.1, we can see that there are similar enzymes that have a completely different role: they bind and phosphorylate. In Figure 2.1 this is the case for ZAP-70. If enzymes like SHP-1 could be involved in downregulating the cell's activity, enzymes like ZAP-70 could be involved in upregulating it. In the next chapter we will focus on, and attempt to model, that kind of protein interaction.

Chapter 3

Tethered Enzymatic Reactions: Phosphorylation

3.1 Introduction

We mentioned in the introduction of the previous chapter how the immune system uses activating and inhibiting receptors to detect pathogens and attack them, or stop an over-reaction of the immune system. We modelled the behaviour of a particular inhibiting receptor, but its counterpart, the activating receptor, is just as important. In Chapter 2 we developed the forward reaction, where the enzyme SHP-1 would dock unto proteins fixed to the cell membrane and then dephosphorylate surrounding ones. We now proceed to study the reverse event in which fixed, dephosphorylated, proteins (or receptor tails) are phosphorylated by enzymes in the cytoplasm that can attach to a phosphorylated activating receptor tail and phosphorylate surrounding tails.

The dephosphorylating reaction seen in Chapter 2 could be indicative of the action of an enzyme (such as SHP-1) that down-regulates the immune reaction in a T-cell. In this chapter we are modelling an enzyme that could help up-regulate the

immune reaction in said T-cell. If, as seen for the SHP-1 case in Goyette et al. [1], the tethering of the enzyme in this chapter increases its phosphorylation rate, then this would mean that a small phosphorylation event could produce a much bigger reaction in the T-cell. In other words, only one external phosphorylation event would be necessary to activate the phosphorylating enzyme, probably caused by the interaction of the antigen receptor with the corresponding antigen. Once this event happens, the tethering enzyme could help phosphorylate surrounding receptors more quickly and, therefore, produce a bigger reaction.

The phosphorylating event we will study here is present in several cells of the immune system. In particular, the zeta-chain-associated protein kinase ZAP-70 is important for T cell receptor signalling [92]. Whereas only 5 proteins in the *Homo sapiens* species have the structure of SHP-1, there are 49 proteins in the *Homo sapiens* species that have the structure of ZAP-70 [93, 94]. This particular enzyme is therefore important to study because of its presence and involvement in the immune system. Malfunction of ZAP-70 results in problems such as persistent infections and severe combined immunodeficiency [95]. Moreover a general understanding of tethered phosphorylation would be useful in research against diseases like diabetes, colon carcinoma [96], and Thrombocytopenia 6 (a clotting disease) [97], amongst many others. These examples underline the fact that the activation, or lack, of the immune system is crucial for health.

To better understand the intricacies of the immune system, these initial interactions with pathogens and inhibitors are important. We must therefore study both kinds of receptors and their interactions with the surrounding proteins. Given the

mathematical techniques that we have already developed, their expansion for the modelling of this new interaction is of interest and the logical next step. This chapter will develop a mathematical model that can be used to understand interactions amongst enzymes that bind and phosphorylate, such as the afore mentioned ZAP-70.

In the same way that there are experiments being done *in vitro* to study SHP-1, the surface plasmon resonance (SPR) technique could be used to experiment with phosphorylating enzymes like ZAP-70 and provide us with data to validate the model developed in this chapter, but these experiments have not yet been done.

In this chapter we explain the biological reaction to be modelled and show how we mathematically abstract it in Sections 3.1.1 to 3.1.3. We then present a computational stochastic model, to which our deterministic models will be compared, in Section 3.1.4. Section 3.2 derives a Mean Field model that approximates the phosphorylating reactions in which we are interested, and we see it does not accurately capture the behaviour of the stochastic simulations. Section 3.3 uses similar mathematical techniques to those used in Chapter 2 and develops a Multi-centre Particle Density PDE (MPDPDE) model that better describes the interaction between tethered proteins, a particular enzyme, and their joint phosphorylation. Section 3.4 solves the MPDPDE model in three dimensions, deals with the problems of solving a stiff set of differential equations, and finally compares the results of the MPDPDE model to the mean-field and the computational stochastic model. The MPDEPDE model in this chapter has the same acronym as the model in the previous chapter because they use the same technique, but the MPDPDE model of Chapter 2 is not used in this chapter.

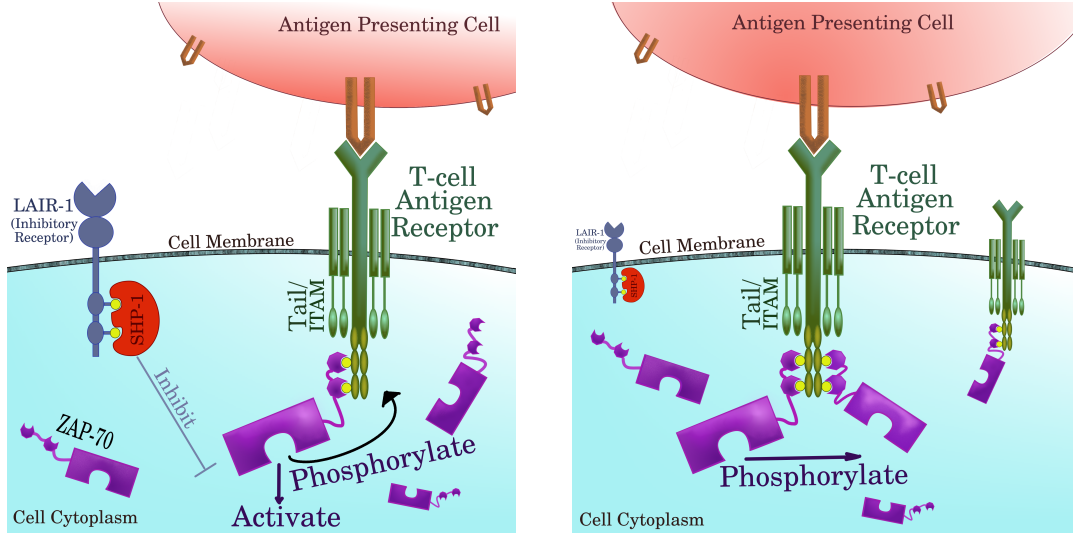
3.1.1 Biological background

Surface receptors on immune cells control their activation decision. Their extracellular parts bind to protein fragments to detect whether the agent may be dangerous or not. The intracellular part, or ‘tail’, interacts with cytoplasmic proteins inside the cell. In the case of antigen receptors recognising a pathogen, the tail becomes activated to initiate a cascade of reactions that may activate the immune cell.

The molecular steps of this process are as follows: Upon binding to the antigen, the tail of the immune receptor will become phosphorylated on multiple residues. These residues are known as immunoreceptor tyrosine-based activation motifs (ITAMs). When these ITAMs are phosphorylated, they generate docking sites for a cytoplasmic enzyme known as ZAP-70. When ZAP-70 docks or binds to an ITAM, it can start phosphorylating substrates that it can reach nearby, including nearby ITAMs [98] (Figure 3.1). This is a form of positive feedback and, ultimately, the signalling cascade that is initiated can lead to the activation of the immune cell.

It was shown in Goyette et al. [1], and followed up in Clemens et al. [91], that the binding or tethering of phosphatase SHP-1 to phosphorylated ITIMs can greatly increase the perceived local concentration of SHP-1. Following these observations, it stands to reason that the tethering of phosphorylating enzymes, or kinases such as ZAP-70, would also increase their perceived concentration at the membrane and hence facilitate their reactions.

The study of how these surface receptors initiate and propagate signals in cells is difficult because each molecular component cannot be easily controlled. In order



(a) ZAP-70 about to phosphorylate another ITAM from the same T cell antigen receptor.

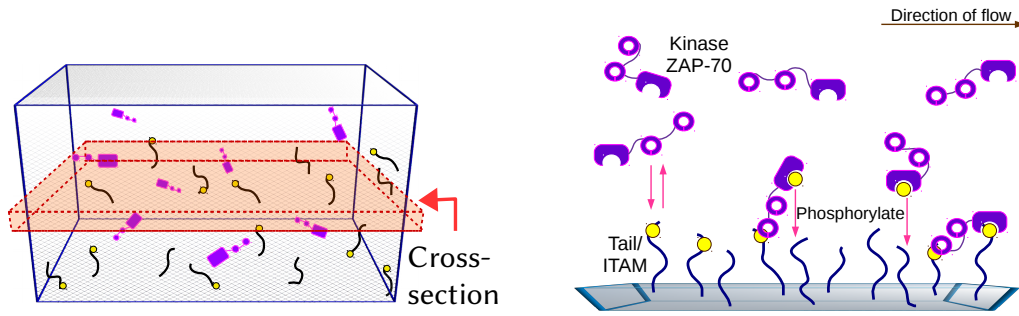
(b) Another ZAP-70 binds to the phosphorylated ITAMs and both ZAP-70s phosphorylate surrounding ITAMs.

Figure 3.1: ZAP-70 enzyme binds to the phosphorylated tail of the T cell antigen receptor (its ITAM) to become activated and phosphorylate other proteins, including other ITAMs (as seen in Figure 3.1a), to which other ZAP-70 enzymes may bind (as seen in Figure 3.1b). Further phosphorylation by ZAP-70 helps activate the T cell, and hence the immune response.

to isolate and control each interaction, *in vitro* experiments are often used where specific proteins can be studied in isolation. The previous chapter focused on an *in vitro* assay to study tethered phosphatases in surface plasmon resonance (SPR). Building on the success of that work, in this chapter we develop a mathematical PDE model that can be used to study kinases, like ZAP-70, in SPR. Although the experiments have yet to be performed, the PDE model that is developed here can be used to guide the experiments.

A T cell receptor contains multiple tails on the inside of the cell (the cytoplasm), as can be seen in Figure 3.1b. Each tail contains ITAMs that, when phosphorylated, generate a docking site for ZAP-70 that itself can phosphorylate other ITAMs. For our SPR experiments the tail containing an unphosphorylated ITAM will be im-

mobilised to a three-dimensional dextran matrix. The experimental set up can be thought of as an upside down representation of the membrane. The SPR experiment will consist of an initial number of unphosphorylated ITAMs, C_0 , immobilised on one end, attached to a three-dimensional dextran matrix, as can be seen in Figure 3.2a. Being immobilised implies that the ITAMs will not be able to diffuse. All ITAMs will be unphosphorylated at the beginning. The matrix will be constantly washed with a mixture of a liquid buffer and a kinase, such as ZAP-70. Therefore, the concentration of this kinase will be known and remain constant, such that there is never a shortage of this enzyme in the experiment. It is noted that the utility of this SPR experiment, just as in the previous chapter, is to recover relevant biophysical and biochemical parameters. Therefore, although the 3D geometry of the SPR experiment may not recapitulate the 2D geometry within cells, by using a mathematical model that accounts for the geometry, we can extract parameter values that are independent of the geometry. For example, the binding affinity is not expected to change whether ZAP-70 binds from solution to a 2D surface (cell membrane) or a 3D matrix (SPR).



(a) Large dextran matrix, as found in the SPR assay.

(b) Cross-section of panel (a) for better view of interactions.

Figure 3.2: Representation of the *in vitro* experiments, using the surface plasmon resonance (SPR) assay to emulate what happens in the cell membrane, Figure 3.1. The experiment consists of a dextran matrix that is much wider, longer, and taller than the length of the proteins, as shown in panel (a). This supporting matrix has ITAMs tethered all over and kinases (ZAP-70) freely flowing through the dextran matrix. A transversal cut of panel (a) is shown in panel (b) for a clearer view of the reactions.

As the kinases undergo diffusion around the tails, they will be able to phosphorylate ITAMs directly from solution without tethering. This phosphorylation from solution is expected to be slow because it is determined, in part, by the concentration of the injected kinase. Once phosphorylated, these ITAMs would then become available to the kinase to bind to phosphorylated ITAMs. These bound, or tethered, kinases would then be able to further phosphorylate surrounding, unphosphorylated, ITAMs. This phosphorylation, when tethered, is expected to be fast because it is determined, in part, by the high local concentration of the tethered kinases that nearby ITAMs will experience. The kinases will, ideally, continue phosphorylating ITAMs, either from the solution or from the bound state, until all ITAMs are phosphorylated. A representation of this experiment can be seen in Figure 3.2b. Given the constant flow, which brings new kinases into the matrix, it is

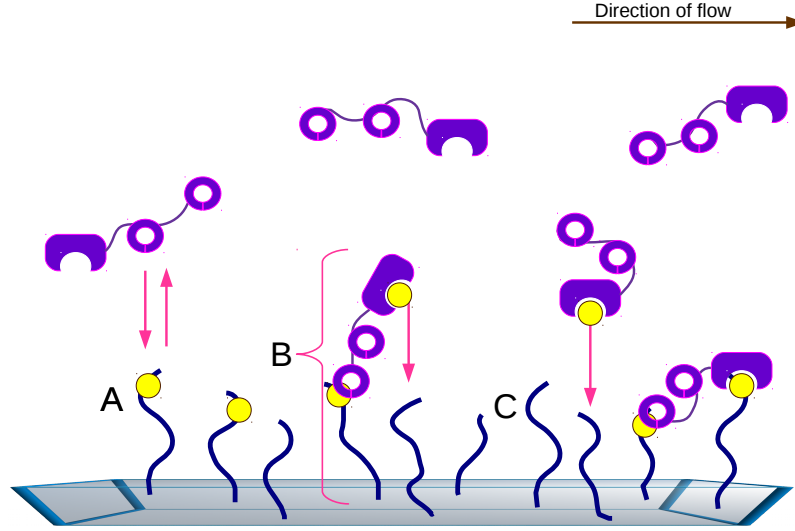


Figure 3.3: Abstract representation of the laboratory experiment in the surface plasmon resonance (SPR) assay, seen in Figure 3.2b, substituted with general species names: A , B , and C . The kinase is purple, the ITAMs are the blue lines, and phosphorylation is denoted by the yellow circle.

reasonable to assume that the concentration of active kinase in solution is fixed so that any chemical reaction has a fixed rate that does not depend on, for example, the depletion of active kinases.

3.1.2 From experiment to mathematical abstraction

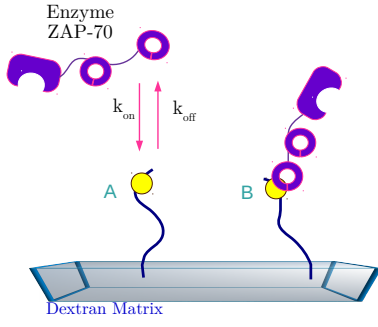
In the SPR assay, the ITAMs are immobilised within the dextran matrix and unable to diffuse. We start by calling these unphosphorylated ITAMs species C . Once an unphosphorylated ITAM has been phosphorylated it becomes species A . The binding of an entity A by a kinase is denoted as species B . An illustration of these labels in the *in vitro* experiment can be seen in Figure 3.3.

As a result of the ITAM being immobilised within the dextran matrix, any modification to it also remains immobile within the dextran matrix (i.e. the diffusion

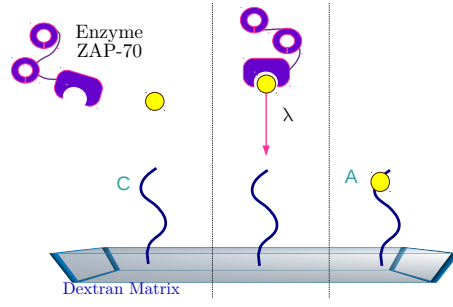
coefficients of A , B , and C are zero). The only molecule that can diffuse is the free kinase in solution. When the kinase binds to a phosphorylated ITAM, it can explore a volume that is determined by the molecular reach of the ITAM tether and the enzyme itself (Figure 3.4d). Given that the ITAM tail is fixed in space onto the dextran matrix, the molecular reach determines, in part, how many other ITAM tails will be within reach and it is these other ITAMs that will experience a high concentration of the tethered kinase. The number of these ITAMs within reach is expected to be very small. This will bring spatial and stochastic elements into our modelling.

Initially, the ITAMs would be unphosphorylated, and we term this the initial amount of C , C_0 . For our abstraction, we will have C become an A through interaction with a kinase in solution, as can be seen in Figure 3.4b. A can then be bound by a kinase to become B . As this kinase can bind and unbind freely, B can turn back into A . B can help phosphorylate C within reach, because of its bound kinase. In all, $A + B + C$ is a constant.

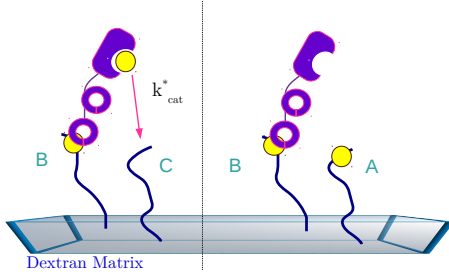
In the SPR assay the enzyme is constantly injected over the dextran matrix. Because we have a constant flow of the enzyme in the experiment, this resource would not get depleted, meaning it is always available and we do not need to consider it as a separate species. The details of the reactions are shown in Section 3.1.3.



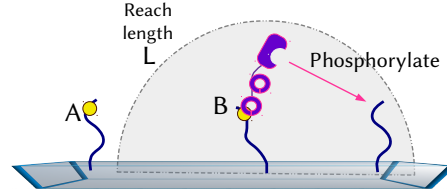
(a) A phosphorylated ITAM, or species A , being bound by the phosphorylating enzyme, ZAP-70, to turn it into species B . Binding rate k_{on} , unbinding rate k_{off} .



(b) An unphosphorylated ITAM, or species C , being phosphorylated by kinase ZAP-70 and turning into species A with a reaction rate λ .



(c) Representation of species B reaching over to a C within its reach length, L , to turn it into an A species, with a reaction rate $k_{\text{cat}}\sigma(r)$.



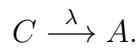
(d) Reach length, L , which B has due to it being anchored at one end but free at the other end. The hypothesis is that B can phosphorylate any dephosphorylated protein within a distance L .

Figure 3.4: Graphic explanation of the reactions hypothesized to be taking place in the SPR experiment.

3.1.3 Reaction Scheme

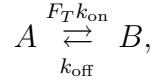
We can now formalise the interactions into a chemical reaction scheme.

As we start with all entities being C , the first interaction that can happen is that of C being turned into A by an enzyme from the solution, with a reaction rate of λ . The first-order reaction for this interaction is:



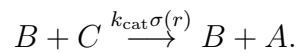
This is also what is depicted in Figure 3.4b, where λ has units of s^{-1} .

The production of state A means that the kinase can now undergo reversible binding and unbinding from A to form B and B to form A with k_{on} and k_{off} , respectively. This reaction is depicted as:



where F_T is the concentration of kinase in the solution, taken as constant due to the constant flow of the enzyme, as mentioned above. This means that the reaction is effectively first-order, and is depicted in Figure 3.4a.

The final chemical equation corresponds to the phosphorylation of surrounding C by compound B reaching over. In this case we consider a catalysing rate, k_{cat} , but we propose that the rate of interaction will also depend on the distance between the reactants. We denote this function as $\sigma(r)$, which has been introduced previously in Section 2.2.1 (restated in Section 3.2.1 here). We will use that same function as the way this compound moves is very similar to the one studied in Chapter 2. The chemical reaction is:



This bimolecular second-order reaction is depicted in Figure 3.4c.

These three chemical reactions form the basis of our deterministic models for a kinase injected over a substrate in SPR. Comparing our model to experimental results in the future should confirm whether the proposed reactions can accurately capture the data. For now, we will compare our deterministic model to a computa-

tional stochastic model, to make sure our assumptions work. This stochastic model is explained in the next section.

3.1.4 Stochastic Simulation

In order to ensure that simplifications in our deterministic model do not introduce large errors into our solution and that we accurately represent the reactions happening in the SPR experiments, a stochastic simulation was developed. We note that, once experimental data are available, the deterministic model will probably be preferred to the stochastic model. A stochastic simulation cannot easily be fit to the experimental data because it is computationally expensive to run, making it impractical for data fitting. On the other hand, deterministic models can often be solved analytically or numerically much more quickly, allowing for data fitting.

As with Chapter 2, a Gibson-Bruck Next Reaction Method was used to create a stochastic simulation of the chemical reactions so that our mathematical model could be compared with the stochastic model. The stochastic model was coded by Dr. Jun Allard (University of California, Irvine) and optimised by Dr. Samuel Isaacson (University of Boston) using MATLAB and C++. A short explanation of how the method works is given in Section 2.1.4.

3.2 Mean field approximation for protein phosphorylation by an enzyme

We can gather the reaction scheme from Section 3.1.3 into the following chemical reactions for our model:



Because F_T is constant, we can simplify the forward reaction rate in (3.1) to:

$$k_{\text{on}}^* = F_T k_{\text{on}}.$$

As done in Section 2.2.2, we intuitively derive the following set of reaction rate equations from a phenomenological point of view:

$$\begin{aligned} \frac{\partial n_A(\vec{x}, t)}{\partial t} &= -k_{\text{on}}^* n_A(\vec{x}, t) + k_{\text{off}} n_B(\vec{x}, t) + \lambda n_C(\vec{x}, t) \\ &\quad + k_{\text{cat}} \int \sigma(\vec{x} - \vec{x}') n_C(\vec{x}, t) n_B(\vec{x}', t) d^3 \vec{x}', \\ \frac{\partial n_B(\vec{x}, t)}{\partial t} &= k_{\text{on}}^* n_A(\vec{x}, t) - k_{\text{off}} n_B(\vec{x}, t), \\ \frac{\partial n_C(\vec{x}, t)}{\partial t} &= -\lambda n_C(\vec{x}, t) - k_{\text{cat}} \int \sigma(\vec{x} - \vec{x}') n_C(\vec{x}, t) n_B(\vec{x}', t) d^3 \vec{x}', \end{aligned} \quad (3.4)$$

where $\vec{x}$ is a three-dimensional position vector, and n_A , n_B , and n_C are the densities of the particles A , B , and C , respectively.

Because the initial conditions are spatially homogeneous (i.e. $n_C(t=0) = C_0/V$, with C_0 being the initial number of C molecules in the system and V the volume of the system, and $n_A(t=0) = n_B(t=0) = 0$), there is no diffusion, and the reactions

are symmetric, it follows that $n_A(\vec{x}, t)$, $n_B(\vec{x}, t)$, and $n_C(\vec{x}, t)$ should remain spatially homogeneous for all time. That is:

$$\begin{aligned} n_A(\vec{x}, t) &= a(t), \\ n_B(\vec{x}, t) &= b(t), \\ n_C(\vec{x}, t) &= c(t), \end{aligned} \tag{3.5}$$

where a , b , and c are the densities of A , B , and C , respectively, throughout space.

The initial conditions are:

$$\begin{aligned} a(t = 0) &= 0, \\ b(t = 0) &= 0, \\ c(t = 0) &= c_T, \end{aligned} \tag{3.6}$$

with c_T the mean concentration of C in the system initially. Making changes (3.5) to Equations (3.4) and rescaling by c_T , we get:

$$\begin{aligned} \frac{da(t)}{dt} &= -k_{\text{on}}^* a(t) + k_{\text{off}} b(t) + \lambda c(t) + k_{\text{cat}} c_T c(t) b(t) \int \sigma(\vec{x} - \vec{x}') d^3 \vec{x}', \\ \frac{db(t)}{dt} &= k_{\text{on}}^* a(t) - k_{\text{off}} b(t), \\ \frac{dc(t)}{dt} &= -\lambda c(t) - k_{\text{cat}} c_T c(t) b(t) \int \sigma(\vec{x} - \vec{x}') d^3 \vec{x}', \end{aligned} \tag{3.7}$$

with

$$\begin{aligned} a(t = 0) &= 0, \\ b(t = 0) &= 0, \\ c(t = 0) &= 1. \end{aligned} \tag{3.8}$$

We assume an infinite space, therefore there are no boundaries.

3.2.1 Interaction kernel $\sigma(r)$

We will be using the same probability for two protein tips to meet as we used in the previous chapter. In Section 2.2.1, Chapter 2, we used the work of Yamakawa [75] and obtained an expression for the probability distribution function of having the last point in the protein chain at a point $\vec{r}$ from the start of the chain:

$$P(\vec{r}) = \left(\frac{3}{2\pi L^2} \right)^{3/2} \exp \left(\frac{-3\vec{r} \cdot \vec{r}}{2L^2} \right). \quad (3.9)$$

For both proteins to end in the same point, regardless of where that point is, but having their tethered ends a distance r apart, we had $\sigma(r)$:

$$\sigma(r) = \left[\frac{3}{2\pi(L_B^2 + L_C^2)} \right]^{3/2} \exp \left[\frac{-3r^2}{2(L_B^2 + L_C^2)} \right], \quad (3.10)$$

where L_B is the average size of one of the proteins involved, and L_C is the average size of the other protein involved. Because L_B and L_C are indistinguishable in our calculations, we will simplify to one unknown parameter, L , with: $L^2 = L_B^2 + L_C^2$.

Therefore:

$$\sigma(r) = \left[\frac{3}{2\pi L^2} \right]^{3/2} \exp \left[\frac{-3r^2}{2L^2} \right], \quad (3.11)$$

3.2.2 Solving the Mean Field Approximation

As seen in Section 2.2.2, the three-dimensional integral of $\sigma(\vec{x} - \vec{x}')$ is:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} \sigma(|\vec{x} - \vec{x}'|) \, dr' \, d\theta \, d\phi = 1. \quad (3.12)$$

We will substitute this expression into Equations (3.7) to obtain the equations to solve for the Mean Field Approximation:

$$\frac{\partial a(t)}{\partial t} = -k_{\text{on}}^* a(t) + k_{\text{off}} b(t) + \lambda c(t) + k_{\text{cat}} c_{TC}(t) b(t),$$

$$\begin{aligned}
\frac{\partial b(t)}{\partial t} &= k_{\text{on}}^* a(t) - k_{\text{off}} b(t), \\
\frac{\partial c(t)}{\partial t} &= -\lambda c(t) - k_{\text{cat}} c_T c(t) b(t).
\end{aligned}
\tag{3.13}$$

Taking into account the conservation of molecules, that is $a(t) + b(t) + c(t) = 1$ $\forall t$, and that $b(t) \geq 0 \forall t$, the steady state for this system is:

$$\begin{aligned}
a_s &= \frac{k_{\text{off}}}{k_{\text{on}}^* + k_{\text{off}}}, \\
b_s &= \frac{k_{\text{on}}^*}{k_{\text{on}}^* + k_{\text{off}}}, \\
c_s &= 0.
\end{aligned}
\tag{3.14}$$

Several simulations of the computational stochastic model were run to compare with the numerical solution of Equation (3.13) and the comparison is shown in Figure 3.5. It can be observed that, whilst both models reach the same theoretical steady states in the long run, there is a significant difference between the stochastic simulations and the Mean Field Approximation numerical solution during the first 25 seconds. Whilst we can use these steady states to obtain $k_{\text{on}}^*/k_{\text{off}}$ from experimental data, our aim is to gain more information from the data. This discrepancy between the models is important because by fitting a model that takes into account stochasticity and location we will be able to obtain the value of five kinetic rate constants from the experiments. In the next section we focus on an improved deterministic model that does take into account stochasticity and location.

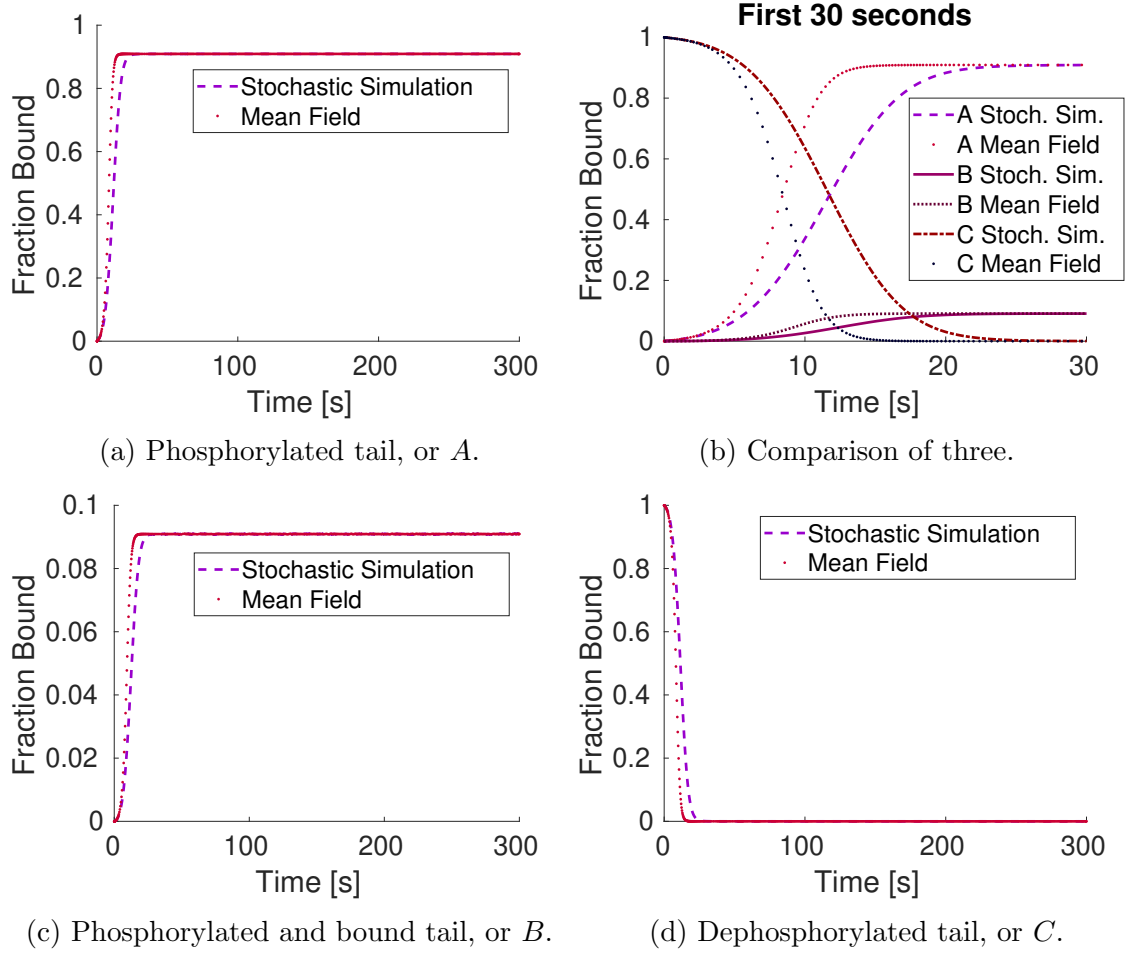


Figure 3.5: Comparison between the stochastic model and the numerical solution of the Mean Field approximation, Equations (3.13). The parameter values used were: $k_{\text{on}}^* = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{M}^{-1}\text{s}^{-1}$, $\lambda = 0.005 \text{ s}^{-1}$. The stochastic simulation used 300,000 initial C molecules in a volume of $1000 \times 1000 \times 1000 \text{ nm}^3$, which translates to $c_T = 498.1734 \mu\text{M}$ in the Mean Field model. Length of tail used for the stochastic simulation was $L = 20 \text{ nm}$, the simulation had periodic boundary conditions on all four sides, and a 100 repetitions were averaged. Run time of stochastic simulations was 3.5 days.

3.3 Multi-centre Particle Density Model of Phosphorylation (MPDPDE)

As discussed in the previous chapter, and verified in Section 3.2, the well-mixed model does not fit the *in silico* data well enough, because the peptides are immobile.

As before, we will explicitly model spatial effects and, in addition, include stochastic

effects because the enzyme is only able to reach a few substrates. In contrast to the case in Chapter 2, in this case the reactants (unphosphorylated peptides) and products (phosphorylated peptides) both play important roles. Therefore, we have three chemical species and a set of equations for them. As we will see, this will increase the model from five to nine integral partial-differential-equations.

As before, we propose B and C , situated at a distance r apart, react with a probability of interaction of $\sigma(r)$. This interaction kernel will be the same as the one used in the previous model (seen again in subsection 3.2.1). We base our model on Equations (3.1), (3.2), and (3.3), restated here:



where $k_{\text{on}}^* = F_T k_{\text{on}}$.

We will start with all molecules unphosphorylated, so the initial conditions will be:

$$A(t = 0) = 0,$$

$$B(t = 0) = 0,$$

$$C(t = 0) = C_T,$$

where C_T is the initial amount of tethered unphosphorylated peptide. As before, we have the following conservation equation, $C_T = A + B + C \ \forall t$.

3.3.1 Pair Correlation Functions

We will use k -point distribution functions [76, 78, 79], $\rho^{(k)}$ ($k = 1, 2, \dots$), as in Section 2.3.1, which quantify “how the particle of interest is surrounded by other particles” [99]. This $\rho^{(k)}$ can be seen as a probability distribution function with k centres. In other words, it is related to the density of k particles seen from the viewpoint of one of these particles.

In our case, because we have three species, the *many-point densities* will be written as $\rho^{(k)} = \rho_{m,m',m''}$, where $k = m + m' + m''$. We will use a definition similar to that given in Equation (2.23), following the work in Kotomin and Kuzovkov [82], but for three species:

$$\rho_{m,m',m''}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}; \{\vec{r}''\}_{m''}) = v^{-(m+m'+m'')} P(\{A_{\vec{r}}\}_m; \{B_{\vec{r}'}\}_{m'}; \{C_{\vec{r}''}\}_{m''}),$$

where v are small regions of the same volume such that each region has at most one particle, each identified by its vector $\vec{r}$, and $P(\{A_{\vec{r}}\}_m; \{B_{\vec{r}'}\}_{m'}; \{C_{\vec{r}''}\}_{m''})$ gives the probability of finding particles A in m regions with the coordinates $\vec{r}_i (i = 1, \dots, m)$, particles B in m' regions with the coordinates $\vec{r}_j (j = 1, \dots, m')$, and particles C in m'' regions with the coordinates $\vec{r}_k (k = 1, \dots, m'')$. $v^{-(m+m'+m'')}$ is the multiplication of the volume of each of these small regions. This definition gives us an idea of the density of probability of specific species.

We have, as in Section 2.3.1, that $\rho_{1,0,0}(\vec{r};) = P(A_{\vec{r}})/v$ is the density (concentration) of particles A at a region identified by $\vec{r}$ because there is only one particle per region. If the system is homogeneous ($V \rightarrow \infty$) then $\rho_{1,0,0}$, $\rho_{0,1,0}$, and $\rho_{0,0,1}$ are

independent of position, so we can write:

$$\rho_{1,0,0} \equiv n_A(t), \quad (3.18)$$

where we define $n_A(t)$ as the mean-field particle concentration of A . The same can be done for $\rho_{0,1,0} \equiv n_B(t)$ and $\rho_{0,0,1} \equiv n_C(t)$. This will help us define correlation functions for higher level distribution functions [81, 84]:

$$\rho_{m,m',m''}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}; \{\vec{r}''\}_{m''}) = n_A^m(t) n_B^{m'}(t) n_C^{m''}(t) \chi_{m,m',m''}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}; \{\vec{r}''\}_{m''}), \quad (3.19)$$

where $\chi_{m,m',m''}(\{\vec{r}\}_m; \{\vec{r}'\}_{m'}; \{\vec{r}''\}_{m''})$ (we simplify the notation by denoting this $\chi_{m,m',m''}$) will be the many-point correlation functions, which will help us keep track of the probability of finding centres of other particles a given distance from any one particle [80]. This function is not known *a priori* and cannot normally be calculated analytically, so we will obtain it numerically.

Now, we know the distribution function for just one particle, which is the mean-field density of particles. For two particles the distribution function depends on the particles' coordinates, and there are 6 combinations:

$$\rho_{2,0,0} \equiv n_A^2(t) X_A(r_1 - r_2, t), \quad (3.20)$$

$$\rho_{0,2,0} \equiv n_B^2(t) X_B(r'_1 - r'_2, t), \quad (3.21)$$

$$\rho_{0,0,2} \equiv n_C^2(t) X_C(r''_1 - r''_2, t), \quad (3.22)$$

$$\rho_{1,1,0} \equiv n_A(t) n_B(t) Y_{AB}(r_1 - r'_1, t), \quad (3.23)$$

$$\rho_{0,1,1} \equiv n_B(t) n_C(t) Y_{BC}(r'_1 - r''_1, t), \quad (3.24)$$

$$\rho_{1,0,1} \equiv n_A(t) n_C(t) Y_{AC}(r''_1 - r_1, t), \quad (3.25)$$

where $X_A(r, t)$, $X_B(r, t)$, $X_C(r, t)$, $Y_{AB}(r, t)$, $Y_{BC}(r, t)$, and $Y_{AC}(r, t)$ are the pair correlation functions. We note that these functions are equal to one when there is no correlation between two particles and their probability is independent. For example, if every A particle is independent of each other, then $\rho_{2,0,0} = \rho_{1,0,0} \times \rho_{1,0,0}$, and:

$$X_A = \frac{\rho_{2,0,0}}{\rho_{1,0,0}\rho_{1,0,0}} = 1.$$

Equation (3.20) states how likely you are to find another A particle a distance $|r_1 - r_2|$ away given that you are positioned at an A particle. Equations (3.21) and (3.22) are similar but with B and C particles, respectively, whereas Equation (3.23) refers to having an A particle and trying to find a B particle a distance $|r_1 - r'_1|$ away. Equation (3.23) is symmetric for particles A and B . The same logic applies to the remaining correlations.

Using the formalism introduced by Kuzovkov and Kotomin [81], we derive the general model, based on the reactions in equations (3.15), (3.16), and (3.17):

$$\begin{aligned} \frac{\partial \rho_{m,m',m''}}{\partial t} = & \left. \begin{aligned} & m k_{\text{off}} \rho_{m-1,m'+1,m''} + m' k_{\text{on}}^* \rho_{m+1,m'-1,m''} \\ & - m k_{\text{on}}^* \rho_{m,m',m''} - m' k_{\text{off}} \rho_{m,m',m''} \end{aligned} \right\} & \text{From Eq. (3.15)} \\ & \left. \begin{aligned} & + m \lambda \rho_{m-1,m',m''+1} - m'' \lambda \rho_{m,m',m''} \end{aligned} \right\} & \text{From Eq. (3.17)} \\ & \left. \begin{aligned} & - \sum_{i=1}^{m'} \sum_{j=1}^{m''} k_{\text{cat}} \sigma(r'_i - r''_j) \rho_{m,m',m''} \\ & - \sum_{i=1}^{m''} \int k_{\text{cat}} \sigma(r'_{m'+1} - r''_i) \rho_{m,m'+1,m''} d^d r'_{m'+1} \\ & + \sum_{i=1}^m \sum_{j=1}^{m'} k_{\text{cat}} \sigma(r'_j - r''_i) \rho_{m-1,m',m''+1} \\ & + \sum_{i=1}^m \int k_{\text{cat}} \sigma(r'_{m'+1} - r''_i) \rho_{m-1,m'+1,m''+1} d^d r'_{m'+1}. \end{aligned} \right\} & \text{From Eq. (3.16)} \end{aligned} \quad (3.26)$$

The first four terms on the right-hand side come from the reaction in Equation (3.15). The first of these terms captures the observation that any of the m A particles could have come from a B particle in that position that turned into A with a k_{off}

rate, necessarily coming from a state with one more B particle and one fewer A particle to reach the desired $\rho_{m,m',m''}$ state. The second term is, similarly, for any of the m' B particles that could come from an A particle turning into a B particle with a k_{on}^* rate. This is how we could have got into the $\rho_{m,m',m''}$ state. In contrast, the third and fourth terms refer to being in the $\rho_{m,m',m''}$ state and either losing any of the present m A particles with a rate k_{on}^* , or losing any of the m' B particles with a rate k_{off} .

Similarly, the fifth term comes from moving into the $\rho_{m,m',m''}$ state due to a C particle turning into one of the m A particles required with a rate λ , coming from a state with one fewer A particle and one more C particle. In turn, the sixth term assumes we start at the $\rho_{m,m',m''}$ state and one of the m'' C particles turns into an A particle with a rate λ .

Finally, the last four terms on the right-hand side stem from the reaction in Equation (3.16). We first start with the $\rho_{m,m',m''}$ state, which can be lost by a reaction between any of the m' B particles with any of the m'' C particles inside the $k = m + m' + m''$ group of particles with which we are specifically dealing. Because the reaction rate depends on k_{cat} , but also the distance between each pair of particles through the σ function, a check through every possible couple within the group is made with the summations. Secondly, we consider that any C particle in the $\rho_{m,m',m''}$ state could be lost by reacting with a B particle out of the group being considered, which is why the second of these terms has an integral with respect to “ $m' + 1$ ” B particles. The third of these terms considers a C particle within the group reacting with one of the B particles within the group to become the required

A particle for the $\rho_{m,m',m''}$ state, and the fourth term considers, again, a C particle within the group becoming the required A particle, but through interaction with a B particle outside the group, hence the integral over space.

This general formula, Equation (3.26), will give us the specific PDEs for our model variables. For example, to start with we can consider only one A particle:

$$\frac{\partial \rho_{1,0,0}}{\partial t} = k_{\text{off}} \rho_{0,1,0} - k_{\text{on}}^* \rho_{1,0,0} + \lambda \rho_{0,0,1} + \int k_{\text{cat}} \sigma(r'_1 - r_1) \rho_{0,1,1} d^d r'_1. \quad (3.27)$$

With the definitions in Equations 3.18, its equivalents for $\rho_{0,1,0}$ and $\rho_{0,0,1}$, and 3.24 we can substitute into Equation (3.27):

$$\frac{\partial n_A(t)}{\partial t} = k_{\text{off}} n_B(t) - k_{\text{on}}^* n_A(t) + \lambda n_C(t) + k_{\text{cat}} n_B(t) n_C(t) \int \sigma(r') Y_{BC}(r', t) d^d r'.$$

This gives us a PDE for the evolution of $n_A(t)$ over time.

In a similar way, we can use Equation (3.26) for the first 9 moments:

$$\frac{\partial \rho_{1,0,0}}{\partial t} = k_{\text{off}} \rho_{0,1,0} - k_{\text{on}}^* \rho_{1,0,0} + \lambda \rho_{0,0,1} + \int k_{\text{cat}} \sigma(r'_1 - r_1) \rho_{0,1,1} d^d r'_1, \quad (3.28)$$

$$\frac{\partial \rho_{0,1,0}}{\partial t} = k_{\text{on}}^* \rho_{1,0,0} - k_{\text{off}} \rho_{0,1,0}, \quad (3.29)$$

$$\frac{\partial \rho_{0,0,1}}{\partial t} = -\lambda \rho_{0,0,1} - \int k_{\text{cat}} \sigma(r'_1 - r''_1) \rho_{0,1,1} d^d r'_1, \quad (3.30)$$

$$\begin{aligned} \frac{\partial \rho_{2,0,0}}{\partial t} &= 2k_{\text{off}} \rho_{1,1,0} - 2k_{\text{on}}^* \rho_{2,0,0} + 2\lambda \rho_{1,0,1} \\ &+ \int k_{\text{cat}} \sigma(r'_1 - r''_1) \rho_{1,1,1} d^d r'_1 + \int k_{\text{cat}} \sigma(r'_1 - r''_2) \rho_{1,1,1} d^d r'_1, \end{aligned} \quad (3.31)$$

$$\frac{\partial \rho_{0,2,0}}{\partial t} = 2k_{\text{on}}^* \rho_{1,1,0} - 2k_{\text{off}} \rho_{0,2,0}, \quad (3.32)$$

$$\begin{aligned} \frac{\partial \rho_{0,0,2}}{\partial t} &= -2\lambda \rho_{0,0,2} \\ &- \int k_{\text{cat}} \sigma(r'_1 - r''_1) \rho_{0,1,2} d^d r'_1 - \int k_{\text{cat}} \sigma(r'_1 - r''_2) \rho_{0,1,2} d^d r'_1, \end{aligned} \quad (3.33)$$

$$\frac{\partial \rho_{1,1,0}}{\partial t} = k_{\text{off}} \rho_{0,2,0} + k_{\text{on}}^* \rho_{2,0,0} - k_{\text{on}}^* \rho_{1,1,0} - k_{\text{off}} \rho_{1,1,0} + \lambda \rho_{0,1,1}$$

$$+ k_{\text{cat}}\sigma(r'_1 - r''_1)\rho_{0,1,1} + \int k_{\text{cat}}\sigma(r'_2 - r''_1)\rho_{0,2,1} \, d^d r'_2, \quad (3.34)$$

$$\begin{aligned} \frac{\partial \rho_{0,1,1}}{\partial t} &= k_{\text{on}}^* \rho_{1,0,1} - k_{\text{off}} \rho_{0,1,1} - \lambda \rho_{0,1,1} \\ &\quad - k_{\text{cat}}\sigma(r'_1 - r''_1)\rho_{0,1,1} - \int k_{\text{cat}}\sigma(r'_2 - r''_1)\rho_{0,2,1} \, d^d r'_2, \end{aligned} \quad (3.35)$$

$$\begin{aligned} \frac{\partial \rho_{1,0,1}}{\partial t} &= k_{\text{off}} \rho_{0,1,1} - k_{\text{on}}^* \rho_{1,0,1} + \lambda \rho_{0,0,2} - \lambda \rho_{1,0,1} \\ &\quad - \int k_{\text{cat}}\sigma(r'_1 - r''_1)\rho_{1,1,1} \, d^d r'_1 + \int k_{\text{cat}}\sigma(r'_1 - r''_1)\rho_{0,1,2} \, d^d r'_1. \end{aligned} \quad (3.36)$$

We now need to substitute into the above all many-point densities (ρ) for particle concentrations or pair correlation functions. Although the derivative for $\rho^{(k)}$ for $k = 1$ is straightforward, the derivatives of the other 6 moments are:

$$\begin{aligned} \frac{\partial \rho_{2,0,0}}{\partial t} &= 2n_A X_A \frac{\partial n_A}{\partial t} + n_A^2 \frac{\partial X_A}{\partial t}, \\ \frac{\partial \rho_{0,2,0}}{\partial t} &= 2n_B X_B \frac{\partial n_B}{\partial t} + n_B^2 \frac{\partial X_B}{\partial t}, \\ \frac{\partial \rho_{0,0,2}}{\partial t} &= 2n_C X_C \frac{\partial n_C}{\partial t} + n_C^2 \frac{\partial X_C}{\partial t}, \\ \frac{\partial \rho_{1,1,0}}{\partial t} &= n_B Y_{AB} \frac{\partial n_A}{\partial t} + n_A Y_{AB} \frac{\partial n_B}{\partial t} + n_A n_B \frac{\partial Y_{AB}}{\partial t}, \\ \frac{\partial \rho_{0,1,1}}{\partial t} &= n_C Y_{BC} \frac{\partial n_B}{\partial t} + n_B Y_{BC} \frac{\partial n_C}{\partial t} + n_B n_C \frac{\partial Y_{BC}}{\partial t}, \\ \frac{\partial \rho_{1,0,1}}{\partial t} &= n_C Y_{AC} \frac{\partial n_A}{\partial t} + n_A Y_{AC} \frac{\partial n_C}{\partial t} + n_A n_C \frac{\partial Y_{AC}}{\partial t}. \end{aligned}$$

Some of the equations, like Equation (3.31), have three-point densities that will need three-point distribution functions, which we have not defined. Introducing a definition for the three-point density would require us to extend the general formula and the use of four-point densities. This, in turn, would generate a need for five-point densities and so on, creating an infinite system of nonlinear partial differential equations. As in Chapter 2, we close the system by using the Kirkwood Superposition approximation, detailed in Section 2.3.2. This closure consists of decoupling the

three-centre correlation functions into the product of three pair correlation functions [85], which leads to:

$$\rho_{1,1,1} \approx n_A(t)n_B(t)n_C(t)Y_{AB}(r_1 - r'_1, t)Y_{AC}(r''_1 - r_1, t)Y_{BC}(r'_1 - r''_1, t),$$

$$\rho_{0,2,1} \approx n_B^2(t)n_C(t)X_B(r'_1 - r'_2, t)Y_{BC}(r'_1 - r''_1, t)Y_{BC}(r'_2 - r''_1, t),$$

$$\rho_{0,1,2} \approx n_B(t)n_C^2(t)X_C(r''_1 - r''_2, t)Y_{BC}(r'_1 - r''_1, t)Y_{BC}(r'_1 - r''_2, t).$$

Taking all the above points into account, the complete integro-PDE system for d dimensions (Equation 3.26), assuming all functions are even because they mostly depend on the distance from each particle and $|-r| = |r|$, becomes:

$$\begin{aligned} \frac{\partial n_A(t)}{\partial t} &= k_{\text{off}}n_B(t) - k_{\text{on}}^*n_A(t) + \lambda n_C(t) \\ &\quad + k_{\text{cat}}n_B(t)n_C(t) \int \sigma(r')Y_{BC}(r', t) \, d^d r', \end{aligned} \quad (3.37)$$

$$\frac{\partial n_B(t)}{\partial t} = k_{\text{on}}^*n_A(t) - k_{\text{off}}n_B(t), \quad (3.38)$$

$$\frac{\partial n_C(t)}{\partial t} = -\lambda n_C(t) - k_{\text{cat}}n_B(t)n_C(t) \int \sigma(r')Y_{BC}(r', t) \, d^d r', \quad (3.39)$$

$$\begin{aligned} \frac{\partial X_A(r, t)}{\partial t} &= 2k_{\text{off}}\frac{n_B}{n_A} [Y_{AB}(r, t) - X_A(r, t)] + 2\lambda\frac{n_C}{n_A} [Y_{AC}(r, t) - X_A(r, t)] \\ &\quad - 2k_{\text{cat}}\frac{n_B n_C}{n_A} X_A(r, t) \int \sigma(r')Y_{BC}(r', t) \, d^d r' \\ &\quad + k_{\text{cat}}\frac{n_B n_C}{n_A} Y_{AC}(r, t) \\ &\quad \cdot \int \sigma(r')Y_{BC}(r', t) [Y_{AB}(r + r') + Y_{AB}(r - r')] \, d^d r', \end{aligned} \quad (3.40)$$

$$\frac{\partial X_B(r, t)}{\partial t} = 2k_{\text{on}}^*\frac{n_A}{n_B} [Y_{AB}(r, t) - X_B(r, t)], \quad (3.41)$$

$$\begin{aligned} \frac{\partial X_C(r, t)}{\partial t} &= k_{\text{cat}}n_B(t)X_C(r, t) \\ &\quad \cdot \int \sigma(r')Y_{BC}(r', t) [2 - Y_{BC}(r' + r, t) - Y_{BC}(r' - r, t)] \, d^d r', \end{aligned} \quad (3.42)$$

$$\begin{aligned}
\frac{\partial Y_{AB}(r, t)}{\partial t} &= k_{\text{off}} \frac{n_B}{n_A} [X_B(r, t) - Y_{AB}(r, t)] + k_{\text{on}}^* \frac{n_A}{n_B} [X_A(r, t) - Y_{AB}(r, t)] \\
&+ \lambda \frac{n_C}{n_A} [Y_{BC}(r, t) - Y_{AB}(r, t)] + k_{\text{cat}} \frac{n_C}{n_A} \sigma(r) Y_{BC}(r, t) \\
&- k_{\text{cat}} \frac{n_B n_C}{n_A} Y_{AB}(r, t) \int \sigma(r') Y_{BC}(r', t) \, d^d r' \\
&+ k_{\text{cat}} \frac{n_B n_C}{n_A} Y_{BC}(r, t) \int \sigma(r') Y_{BC}(r', t) X_B(r + r', t) \, d^d r', \tag{3.43}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Y_{BC}(r, t)}{\partial t} &= k_{\text{on}}^* \frac{n_A}{n_B} [Y_{AC}(r, t) - Y_{BC}(r, t)] - k_{\text{cat}} \sigma(r) Y_{BC}(r, t) \\
&+ k_{\text{cat}} n_B Y_{BC} \int \sigma(r') Y_{BC}(r', t) [1 - X_B(r - r', t)] \, d^d r', \tag{3.44}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Y_{AC}(r, t)}{\partial t} &= k_{\text{off}} \frac{n_B}{n_A} [Y_{BC}(r, t) - Y_{AC}(r, t)] + \lambda \frac{n_C}{n_A} [X_C(r, t) - Y_{AC}(r, t)] \\
&+ \left(1 - \frac{n_C}{n_A}\right) k_{\text{cat}} n_B Y_{AC}(r, t) \int \sigma(r') Y_{BC}(r', t) \, d^d r' \\
&- k_{\text{cat}} n_B Y_{AC}(r, t) \int \sigma(r') Y_{BC}(r', t) Y_{AB}(r + r') \, d^d r' \\
&+ k_{\text{cat}} \frac{n_B n_C}{n_A} X_C(r, t) \int \sigma(r') Y_{BC}(r', t) Y_{BC}(r' - r) \, d^d r'. \tag{3.45}
\end{aligned}$$

3.4 Solving the PDEs: 3D

We use spherical coordinates and restate Equations (3.37) to (3.45) for three dimensions, similarly to Section 2.3.3:

$$\begin{aligned}
\frac{\partial n_A(t)}{\partial t} &= k_{\text{off}} n_B(t) - k_{\text{on}}^* n_A(t) + \lambda n_C(t) \\
&+ 4\pi k_{\text{cat}} n_B(t) n_C(t) \int_0^\infty r'^2 \sigma(r') Y_{BC}(r', t) \, dr', \tag{3.46}
\end{aligned}$$

$$\frac{\partial n_B(t)}{\partial t} = k_{\text{on}}^* n_A(t) - k_{\text{off}} n_B(t), \tag{3.47}$$

$$\frac{\partial n_C(t)}{\partial t} = -\lambda n_C(t) - 4\pi k_{\text{cat}} n_B(t) n_C(t) \int_0^\infty r'^2 \sigma(r') Y_{BC}(r', t) \, dr', \tag{3.48}$$

$$\begin{aligned}
\frac{\partial X_A(r, t)}{\partial t} &= 2k_{\text{off}} \frac{n_B}{n_A} [Y_{AB}(r, t) - X_A(r, t)] + 2\lambda \frac{n_C}{n_A} [Y_{AC}(r, t) - X_A(r, t)] \\
&\quad - 8\pi k_{\text{cat}} \frac{n_B n_C}{n_A} X_A(r, t) \int_0^\infty r'^2 \sigma(r') Y_{BC}(r', t) \, dr' \\
&\quad + \frac{4\pi}{r} k_{\text{cat}} \frac{n_B n_C}{n_A} Y_{AC}(r, t) \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q Y_{AB}(q, t) \, dq \, dr',
\end{aligned} \tag{3.49}$$

$$\frac{\partial X_B(r, t)}{\partial t} = 2k_{\text{on}}^* \frac{n_A}{n_B} [Y_{AB}(r, t) - X_B(r, t)], \tag{3.50}$$

$$\frac{\partial X_C(r, t)}{\partial t} = \frac{4\pi}{r} k_{\text{cat}} n_B(t) X_C(r, t) \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q [1 - Y_{BC}(q, t)] \, dq \, dr', \tag{3.51}$$

$$\begin{aligned}
\frac{\partial Y_{AB}(r, t)}{\partial t} &= k_{\text{off}} \frac{n_B}{n_A} [X_B(r, t) - Y_{AB}(r, t)] + k_{\text{on}}^* \frac{n_A}{n_B} [X_A(r, t) - Y_{AB}(r, t)] \\
&\quad + \lambda \frac{n_C}{n_A} [Y_{BC}(r, t) - Y_{AB}(r, t)] + k_{\text{cat}} \frac{n_C}{n_A} \sigma(r) Y_{BC}(r, t) \\
&\quad - 4\pi k_{\text{cat}} \frac{n_B n_C}{n_A} Y_{AB}(r, t) \int_0^\infty r'^2 \sigma(r') Y_{BC}(r', t) \, dr' \\
&\quad + \frac{2\pi}{r} k_{\text{cat}} \frac{n_B n_C}{n_A} Y_{BC}(r, t) \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q X_B(q, t) \, dq \, dr',
\end{aligned} \tag{3.52}$$

$$\begin{aligned}
\frac{\partial Y_{BC}(r, t)}{\partial t} &= k_{\text{on}}^* \frac{n_A}{n_B} [Y_{AC}(r, t) - Y_{BC}(r, t)] - k_{\text{cat}} \sigma(r) Y_{BC}(r, t) \\
&\quad + \frac{2\pi}{r} k_{\text{cat}} n_B Y_{BC} \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q [1 - X_B(q, t)] \, dq \, dr',
\end{aligned} \tag{3.53}$$

$$\begin{aligned}
\frac{\partial Y_{AC}(r, t)}{\partial t} &= k_{\text{off}} \frac{n_B}{n_A} [Y_{BC}(r, t) - Y_{AC}(r, t)] + \lambda \frac{n_C}{n_A} [X_C(r, t) - Y_{AC}(r, t)] \\
&\quad + 4\pi \left(1 - \frac{n_C}{n_A}\right) k_{\text{cat}} n_B Y_{AC}(r, t) \int_0^\infty r'^2 \sigma(r') Y_{BC}(r', t) \, dr' \\
&\quad - \frac{2\pi}{r} k_{\text{cat}} n_B Y_{AC}(r, t) \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q Y_{AB}(q, t) \, dq \, dr' \\
&\quad + \frac{2\pi}{r} k_{\text{cat}} \frac{n_B n_C}{n_A} X_C(r, t) \int_0^\infty r' \sigma(r') Y_{BC}(r', t) \int_{|r-r'|}^{|r+r'|} q Y_{BC}(q, t) \, dq \, dr',
\end{aligned} \tag{3.54}$$

where r' and r are radial coordinates.

The initial conditions will be $n_A(t = 0) = 0$, $n_B(t = 0) = 0$, $n_C(t = 0) = C_T$,
 $X_A(r, t = 0) = X_B(r, t = 0) = X_C(r, t = 0) = Y_{AB}(r, t = 0) = Y_{BC}(r, t = 0) =$

$$Y_{AC}(r, t = 0) = 1.$$

We can next rescale the variables as follows: n_A , n_B , and n_C by C_T , r by L , and t by k_{off} , such that:

$$\hat{n}_A = n_A/C_T,$$

$$\hat{n}_B = n_B/C_T,$$

$$\hat{n}_C = n_C/C_T,$$

$$\hat{r} = r/L,$$

$$\hat{t} = t * k_{\text{off}}.$$

Dropping the hat notation leads to the following non-dimensional MPD-PDE system, which corresponds to our MPDPDE model (remembering this is different to the similarly named model of Chapter 2):

$$\begin{aligned} \frac{\partial n_A(t)}{\partial t} = & n_B(t) - \frac{k_{\text{on}}^*}{k_{\text{off}}} n_A(t) + \frac{\lambda}{k_{\text{off}}} n_C(t) \\ & + 4\pi \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B(t) n_C(t) \int_0^\infty r'^2 e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \, dr', \end{aligned} \quad (3.55)$$

$$\frac{\partial n_B(t)}{\partial t} = \frac{k_{\text{on}}^*}{k_{\text{off}}} n_A(t) - n_B(t), \quad (3.56)$$

$$\begin{aligned} \frac{\partial n_C(t)}{\partial t} = & -\frac{\lambda}{k_{\text{off}}} n_C(t) \\ & - 4\pi \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B(t) n_C(t) \int_0^\infty r'^2 e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \, dr', \end{aligned} \quad (3.57)$$

$$\begin{aligned} \frac{\partial X_A(r, t)}{\partial t} = & 2 \frac{n_B}{n_A} [Y_{AB}(r, t) - X_A(r, t)] + 2 \frac{\lambda}{k_{\text{off}}} \frac{n_C}{n_A} [Y_{AC}(r, t) - X_A(r, t)] \\ & - 8\pi \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} \frac{n_B n_C}{n_A} X_A(r, t) \int_0^\infty r'^2 e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \, dr' \\ & + \frac{4\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} \frac{n_B n_C}{n_A} Y_{AC}(r, t) \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q Y_{AB}(q, t) \, dq \, dr', \end{aligned} \quad (3.58)$$

$$\frac{\partial X_B(r, t)}{\partial t} = 2 \frac{k_{\text{on}}^*}{k_{\text{off}}} \frac{n_A}{n_B} [Y_{AB}(r, t) - X_B(r, t)], \quad (3.59)$$

$$\begin{aligned} \frac{\partial X_C(r, t)}{\partial t} = & \frac{4\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B(t) X_C(r, t) \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q [1 - Y_{BC}(q, t)] \, dq \, dr', \end{aligned} \quad (3.60)$$

$$\begin{aligned} \frac{\partial Y_{AB}(r, t)}{\partial t} = & \frac{n_B}{n_A} [X_B(r, t) - Y_{AB}(r, t)] + \frac{k_{\text{on}}^*}{k_{\text{off}}} \frac{n_A}{n_B} [X_A(r, t) - Y_{AB}(r, t)] \\ & + \frac{\lambda}{k_{\text{off}}} \frac{n_C}{n_A} [Y_{BC}(r, t) - Y_{AB}(r, t)] + \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}}}{k_{\text{off}} L^3} \frac{n_C}{n_A} e^{-\frac{3r^2}{2}} Y_{BC}(r, t) \\ & - 4\pi \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} \frac{n_B n_C}{n_A} Y_{AB}(r, t) \int_0^\infty r'^2 e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \, dr' \\ & + \frac{2\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} \frac{n_B n_C}{n_A} Y_{BC}(r, t) \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q X_B(q, t) \, dq \, dr', \end{aligned} \quad (3.61)$$

$$\begin{aligned} \frac{\partial Y_{BC}(r, t)}{\partial t} = & \frac{k_{\text{on}}^*}{k_{\text{off}}} \frac{n_A}{n_B} [Y_{AC}(r, t) - Y_{BC}(r, t)] - \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}}}{k_{\text{off}} L^3} e^{-\frac{3r^2}{2}} Y_{BC}(r, t) \\ & + \frac{2\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B Y_{BC} \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q [1 - X_B(q, t)] \, dq \, dr', \end{aligned} \quad (3.62)$$

$$\begin{aligned} \frac{\partial Y_{AC}(r, t)}{\partial t} = & \frac{n_B}{n_A} [Y_{BC}(r, t) - Y_{AC}(r, t)] + \frac{\lambda}{k_{\text{off}}} \frac{n_C}{n_A} [X_C(r, t) - Y_{AC}(r, t)] \\ & + 4\pi \left(\frac{3}{2\pi} \right)^{3/2} \left(1 - \frac{n_C}{n_A} \right) \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B Y_{AC}(r, t) \int_0^\infty r'^2 e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \, dr' \\ & - \frac{2\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} n_B Y_{AC}(r, t) \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q Y_{AB}(q, t) \, dq \, dr' \\ & + \frac{2\pi}{r} \left(\frac{3}{2\pi} \right)^{3/2} \frac{k_{\text{cat}} C_T}{k_{\text{off}}} \frac{n_B n_C}{n_A} X_C(r, t) \int_0^\infty r' e^{-\frac{3r'^2}{2}} Y_{BC}(r', t) \\ & \cdot \int_{|r-r'|}^{|r+r'|} q Y_{BC}(q, t) \, dq \, dr'. \end{aligned} \quad (3.63)$$

Bear in mind that to solve this system we require that neither n_A nor n_B are equal to 0, because of their presence in some of the denominators. Since the experimental

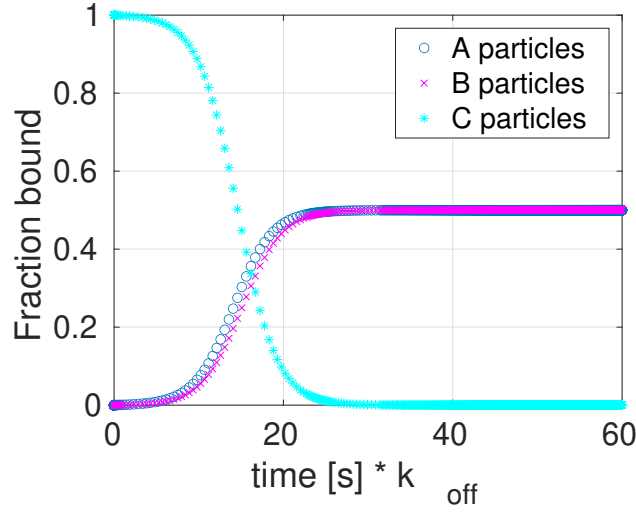


Figure 3.6: Example of numerical solution of three-dimensional MPDPDE model of Chapter 3, shown are Equations (3.55), (3.56), and (3.57). Solved using Matlab's ode23. Parameter values are: $k_{\text{on}}^* = 1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 1 \text{ } \mu\text{M/s}$, $\lambda = 0.001 \text{ s}^{-1}$, $L_A = L_B = 10 \text{ nm}$. The initial conditions were $n_A(0) = \text{eps}$, $n_B(0) = \text{eps}$ (where eps is the smallest possible number different from 0 in the computer, to avoid singularities, which is 2.2204×10^{-16}), $n_C(0) = 1$, and $X_A(r, 0) = X_B(r, 0) = X_C(r, 0) = Y_{AB}(r, 0) = Y_{BC}(r, 0) = Y_{AC}(r, 0) = 1$ (assuming there is no correlation to start with because of homogeneity).

starting point is $n_A = n_B = 0$, but the system leaves that state immediately, we assume at least one C particle gets phosphorylated by an external factor, and ZAP-70 will immediately start being recruited. That is, the system moves away from the initial zero state very quickly and we make the assumption that it is realistic to talk about densities immediately after the start of the experiment.

We can see an example of the numerical solution to Equations (3.55) to (3.63) in Figure 3.6. We also show the time evolution of the pair correlation functions in Figure 3.7, where the results are in a two-dimensional representation, and Figure 3.8, where results are in a three-dimensional representation because of the ease of understanding some functions in one representation or the other.

At the end of the experiment, we note that the steady state is the same as

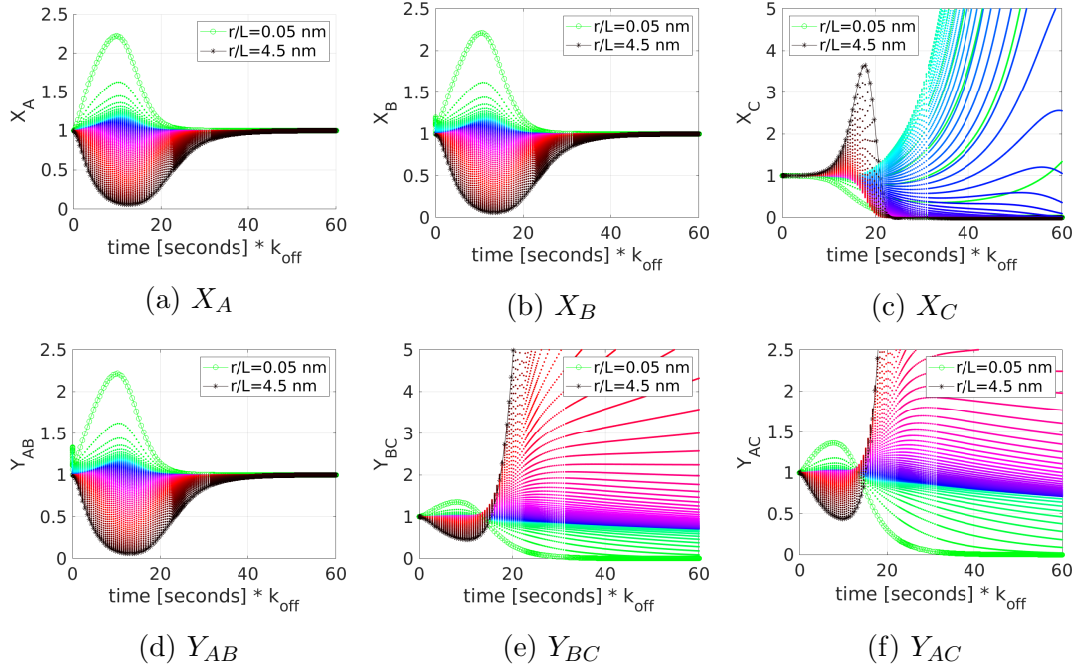


Figure 3.7: Pair correlation functions of three-dimensional MPDPDE model of Chapter 3, Equations (3.55) to (3.63). Shown are the solutions to each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in green, to the furthest $r/L = 4.5$ in black. Hence 90 solutions to each function. Numerically solved by MATLAB's ode23. Solved using Matlab's ode23. Same parameters and starting conditions as in Figure 3.6.

the one given by the Mean Field Approximation model, but this means $n_C = 0$. Although our Equations (3.55) to (3.63) seem to be well-defined in this case, the proximity of n_C to 0 makes C go into a region where it is no longer reasonable to talk about densities, because the number of C proteins becomes too small and stochastic effects become important. Equations (3.55) to (3.63) were solved numerically using MATLAB ode23 package, but some difficulties were found due to it taking a long time to numerically solve. We see in Appendix B that n_C going to 0 and the possible stiffness of the equations are most likely responsible for this long time taken to numerically solve. This means our MPDPDE model can be numerically solved between immediately after $t = 0$ and just before reaching steady state, although the

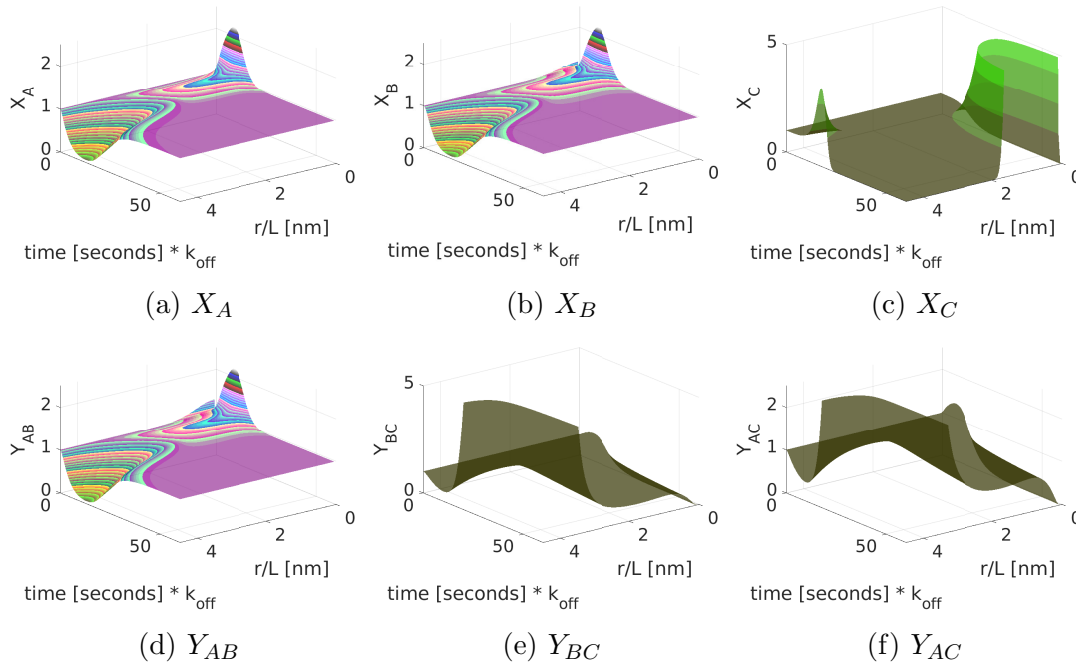


Figure 3.8: Three-dimensional representation of pair correlation functions of the three-dimensional MPDPDE model in Chapter 3, Equations (3.55) to (3.63). Same data points as in Figure 3.7 as well as parameters and starting conditions.

model is valid beyond this region, as long as $n_A, n_B > 0$. Within the numerically solvable interval it was found that the numerical solutions correspond closely with the stochastic model, as we now show.

3.4.1 Comparing models

The MPDPDE 3D model and the Mean Field model were solved with the same parameters used to generate the data from the stochastic simulations to which we want to compare them. After comparing our MPDPDE 3D solution, solved with Matlab's 'ode15s', to the stochastic simulations, we can see that an improvement was made relative to the well-mixed ODE of the Mean Field model. This can be seen in Figure 3.9. Even though our MPDPDE 3D model could not yet be fitted to the stochastic simulations because of long solution times, as explained in Appendix

B, an agreement of $R^2 = 0.998$ was generally obtained with our MPDPDE model, which is remarkable.

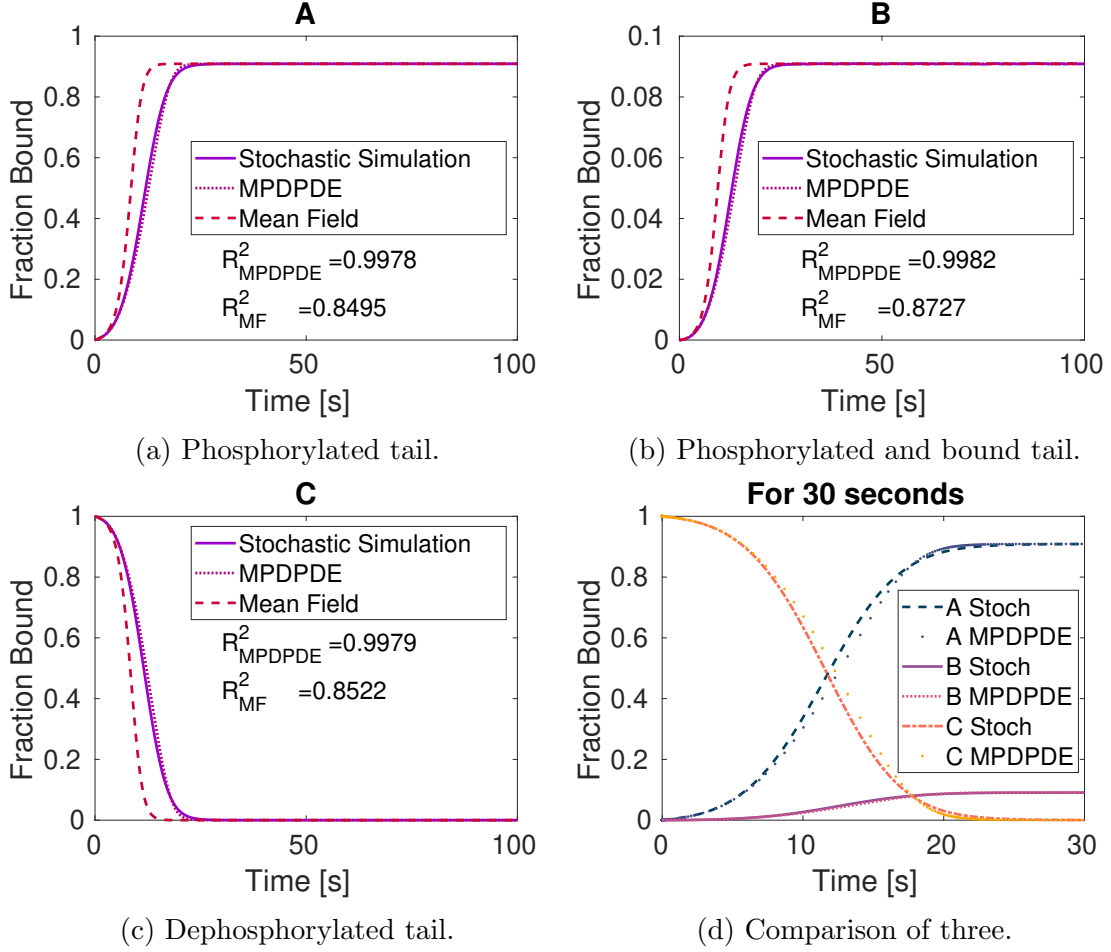


Figure 3.9: Comparing the MPDPDE model and the Mean Field model to the data produced by the stochastic simulation model. The same set of parameters were used for all models presented here, namely: $k_{\text{on}} = 0.1 \text{ s}^{-1}$, $k_{\text{off}} = 1 \text{ s}^{-1}$, $k_{\text{cat}} = 0.02 \mu\text{Ms}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$. The stochastic simulations used 300,000 initial C particles in a $1000 \times 1000 \times 1000 \text{ nm}^3$ area with periodic boundaries. 100 simulations were averaged for this solution. The stiff (ode15s) solver in MATLAB was used to numerically solve the MPDPDE model.

3.5 Conclusions

Signalling reactions through phosphorylation and tethered proteins are common in signalling by surface receptors, such as those found on the surface of T cells. To study

these kinds of reactions we developed mathematical models in the previous chapter which can directly fit SPR data to recover biophysical and biochemical reaction parameters. In this chapter, mathematical models have been developed that are applicable for kinase reactions. In particular, we found that our MPDPDE model compares favourably with the solutions generated from the underlying stochastic model. This MPDPDE model can guide future laboratory experiments by informing on parameter values of interest within those available for SPR experimentation, before the use of kinases in SPR experiments.

Agreement between the models (MPDPDE model and the stochastic simulations) and the SPR experimental data will be tested once experiments are available, but the success of the model in Chapter 2 is encouraging. Once tested, the hope is that the model developed in the present chapter will be of use to easily and quickly obtain the value of the four reaction rates at play and the estimated length of the proteins involved. Knowing whether the model fits the experiments will also work as a proof of concept, to assess whether the main interactions between the participating proteins have been accurately captured.

The use of the MPDPDE model would be preferred to the use of the stochastic simulations because of its shorter running time and because it uses less computational memory. Furthermore, as the MPDPDE model takes into account space, this model could also be used for reactions where the “activation” of a spatially fixed entity enhances the switch of other similar entities in the surrounding area.

There are several avenues for future research. For example, as in Chapter 2, the model could be solved for two-dimensional experiments, rather than three-

dimensional ones as done here, to accommodate the different geometry in which the experiment can be carried out. This would widen the model's use and take into account different experimental set-ups.

In conclusion, we have derived and solved a mathematical model that seeks to represent tethered phosphorylation reactions. This model takes into consideration both stochastic and spatial elements of tethered reactions when the tethers are immobile. The MPDPDE model, but not a well-mixed ODE model, agreed favourably with the corresponding stochastic model simulations. This work furthers our understanding of immobile proteins that interact with their surroundings using flexible tethers, and could be adapted to other similar interactions in a wider biological context.

Chapter 4

Hunter-Gatherer Friendships: Mathematical model of meeting

4.1 Introduction

In this chapter, we turn our attention to early-human communities, when nomadic movement took place. For early human communities to survive, they needed to share knowledge and responsibilities with other members of the species. The larger their social network, the bigger the pool of knowledge available to them. However, time is a limited resource and individuals have to fulfil a number of necessary functions, such as eating and sleeping. Since arranging meetings consumes precious time, we consider the possibility of optimising time use by socialising and bonding whilst fulfilling other needs. In this sense it is useful to know how many, and how often, humans can meet during a given period of time, whilst performing other necessary tasks, such as gathering food.

This project seeks to investigate and quantify the interaction rates of such populations given constraints on extrinsic group factors (such as time, food, environment, and climate). This goal will be achieved by initially using stochastic agent-based

models and then comparing our numerical results to mean-field approaches in order to scale the individual effects up to a population-wide understanding, giving the problem a multi-scale character. Ultimately, we endeavour to produce a mechanistic framework for understanding the current theoretical “friendship limit” of Dunbar [65] and determine which features contribute to such limits. To this end we will use archaeological and anthropological data to obtain insight for our theoretical models.

This chapter will study the problem from a mathematical perspective, by seeking to model and characterise the size of an individual’s social network. This approach has the potential to facilitate the testing of ecological hypotheses on how a social network size depends on an individual’s interactions, movement, memory, and population density. Understanding the socialization of hunter-gatherers could bring us closer to understanding general human socialization, along with present-day health issues related to loneliness, mental illnesses, and anti-social crime. Going beyond humans, understanding the mechanisms and needs for social connection may help in the conservation of animal habitats and the environment needed for their healthy existence.

4.1.1 Biological background

For social animals, larger interactive networks mean a higher rate of survival, because of the sharing of knowledge, information, and resources. It is a well known fact that social organization has been present throughout hominid evolution [100, 101, 102, 103, 104].

For example, childrearing is easier when shared amongst many, with increased

chances of offspring survival [105], whilst the knowledge of where and how to obtain food is of great value to the whole community. Also, there is evidence that larger social networks and stronger connections amongst individuals may play an important role in helping individuals fight illnesses and improve their physical and mental health [105, 106, 107, 108, 109]. Such unity has proved vital for our species' survival.

The organization of social animals has different levels of complexity that allow for different numbers of individuals in each network. In our case, although large networks are good, human social groups cannot grow infinitely large. Indeed, as human social groups grow, they tend to fracture for periods of time, forming bands (residential units [110] or, more broadly, groups of people), before periodically reuniting to renew social bonds. Critically, even though they are fractured, the bands still consider each other to be part of the same society. Layton et al. [111] hypothesized that band formation may have been due to the predatory nature of hunter-gatherers: The scarcity of prey, as opposed to the availability of plants in the case of chimpanzees, for example, could have forced the density of humans to be lower. The main difference between chimpanzees and hunter-gatherers is that the latter consume much more meat [103, 111, 112, 113], allowing for energy to be extracted quickly for use in the brain rather than in digestion [114]. The increase in brain size, in turn, allowed for more complex organizations, amongst other properties.

Band formation and splitting could have also been a response to the need for a greater collaboration amongst hunter-gatherers within a more dispersed community [111]. For example, to increase the robustness of hunting, the existence of a larger community means that the failure of a smaller group would not necessarily doom

each individual, who could then seek help from other groups. Moreover, a smaller group could move faster in order to hunt and would have to hunt fewer prey.

Band formation in humans created a hierarchical importance of individuals with respect of each other and, hence, a social stratification.

Anthropologists acknowledge up to 6 layers in human social organization [115]. The number of layers depends on the season of the year and on the precise definition used, although these definitions do not vary widely. These 6 layers can be seen in Figure 4.1. For the analysis of our models we will use ‘idealised’ layers [2, 115]. These layers are summarised in Table 4.1, taken from Pearce [2], which compares the different definitions across the area. The innermost layer comprises the domestic band or family, which tends to have 3 to 5 individuals. This group tends to stick together for daily activities. The second layer is the sympathy group, with around 12 to 15 individuals, who would always share sleeping space at the end of the day, even during the most dispersed time of the season. These two layers are the ones to which an individual would go to seek advice and help, making them the most important for an individual’s mental health [116]. The third level would be the aggregated band, composed of around 50 individuals, with which the members would share sleeping space at the end of the day during the most aggregated time of the season. Fourth is the active network of around 150 individuals. These are the people with whom an individual makes a conscious effort to maintain contact. In an analysis by Dunbar [65], it emerged that this active network layer is limited by the capacity of our brain to process relationships, also reaching the conclusion that it is consistently made up of approximately 150 individuals (Dunbar’s number). We mention the

importance of the time of the season in the second and third layers because for tribes the abundance of food at certain times of the year means bigger groups can come together [111]. The fifth and sixth layers do not revolve around the individual but mark the boundaries of the society in which that individual lives. With around 500 individuals we have the endogamous mating pool, and with around 1500 individuals we have the ethnolinguistic tribe.

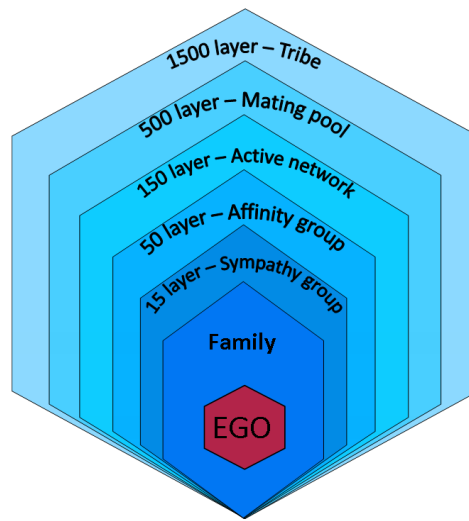


Figure 4.1: Network layers for an average hunter-gatherer. See text for details.

As hunter-gatherers moved out of Africa and into higher latitudes, social organization had to be maintained over greater geographic distances. Meeting everyone in the tribe would be harder due to higher travelling costs and lower human densities. It would become more challenging to maintain periodic aggregations, but the importance of these could be even greater than before to maintain the social network. We hypothesise that this could have been the reason for the evolution of different behaviours, like periodic aggregations or gift-giving, but verifying these hypotheses is difficult due to lack of data.

‘Idealised’ layers based on those identified by Zhou et al. [115]	Personal network layers	Hunter-gatherer grouping layers	Binford [117] hunter-gatherer groups
5	Clique	Nuclear family	Family
15	Sympathy group	Foraging party	Group 1 (band during the dispersed phase of the season)
50	Affiliation group	Residential band	Group 2 (band during the aggregated phase of the season)
150	Active network		Group 3 (periodic ag- gregations of bands)
500		Megaband / endogamous mating pool	
1500		Ethnolinguistic tribe	Population

Table 4.1: The assumed equivalences used in this project for social organization in hunter-gatherers as given in Pearce [2].

Because experiments are hard to perform on hunter-gatherers, mathematical models can be used to investigate possible case-scenarios. Pearce [2] uses a gas model (see below) to suggest that, at lower latitudes, the walks made daily to fulfil basic needs alone are sufficient for tribes to maintain such social networks, whereas in Arctic latitudes this would not be sufficient and further mechanisms for social cohesion would be needed, such as token gifts from tribe to tribe. This can help us explain the geographic patterning of certain elements of the archaeological record. This study was done from the archaeological point of view, tracking maximum obsidian transfer distances from source to burial, and relating these to the possible existing social links between groups [118].

As the groups spread over the globe, regions they moved into had different char-

acteristics, such as: weather, flora, fauna, carrying capacities, and landscape. These differences affected the individuals' ability to interact with each other. Some combinations of characteristics would have had increased the likelihood of interactions, whereas some of them would have had decreased it. For example, places with a food surplus and a flat landscape could have a higher concentration of groups [119], making it easier for them to find each other, whilst places with little flora and harsh weather could mean fewer groups near each other for possible interactions. Based on the idea that similar conditions will provide similar amounts of interactions, we initially associate each community with its biome. A biome, as seen in Figure 4.2, is an area of land with similar characteristics.

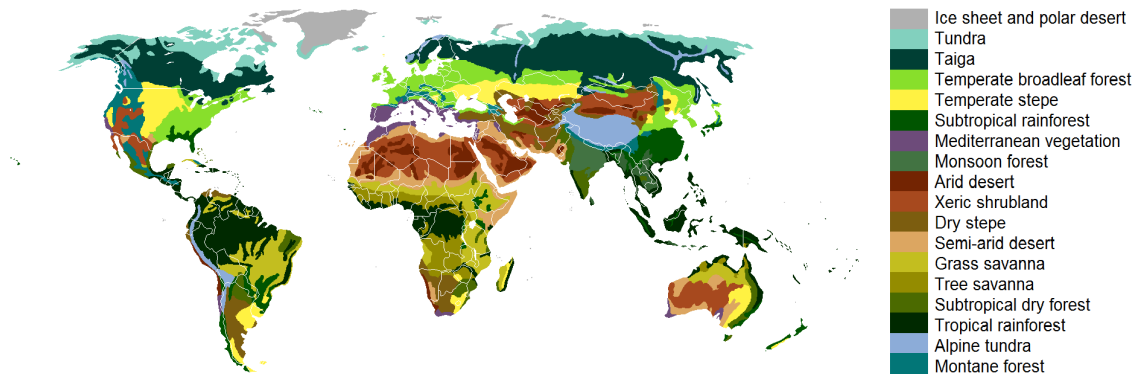


Figure 4.2: Biomes: regions of land with similar weather, flora and fauna conditions. Different biomes are represented by different colours. Image by Ville Koistinen, taken from:

<https://commons.wikimedia.org/wiki/File:Vegetation.png>,

reproduced as permitted by the CC-BY-SA 3.0 licence.

4.1.2 Mathematical background

Due to the seemingly somewhat erratic movement in animals, a vast range of studies have been carried out based on gas models and random walks [120, 121, 122].

However, less attention has been given to the study of the effects of small-number

stochasticity in animal models. This chapter focuses on studying the effect of small-number dynamics in random walks with memory. Studies on animal movement over long periods of time have revealed non-Markovian effects [123, 124, 125], which we intend to consider as part of our work in the future. However, we presently begin with the most basic set of dynamics.

4.1.2.1 Gas model and applications

The ideal gas model [126] refers to the idea of randomly moving particles colliding with each other. It is a simple model that has been used in biology to aid in the modelling of events in which randomness seems to play an important role [127, 128, 129, 130, 131], for example, to estimate encounter rates within and between species [129, 132]. It has also been used to prove that the movement of some animals is far from random by showing clear differences between the predicted results of the gas model and real measurements [132, 133, 134, 135, 136]. A good overview of this kind of model, and its shortcomings, can be found in Hutchinson and Waser [126].

We start with the gas model as a way of coarsely approximating the motion of hunter-gatherers as random walkers. It should be noted that we will be dealing with relatively small numbers of groups moving over large areas, therefore the usual population-level deterministic framework may need to incorporate stochastic effects.

4.1.3 Computational background

We will be using a continuous-time, discrete space stochastic simulation algorithm. These are good at giving insight into the overall effect of stochasticity arising from low numbers of agents. The simulations will be compared with the analytical and

numerical representations of our models.

Our stochastic compartmental model stems from the framework of chemical reactions, as in Erban et al. [73], and it is called a Gillespie Stochastic Simulation Algorithm (SSA). Whilst agent-based simulations can keep track of individual entities (agents), be it molecules, animals, or families, the Gillespie SSA does not naturally do this. The Gillespie SSA is capable, though, of “labelling” each entity into different groups, called *species*, which we will use to keep track of agents (or families) separately. The implied stochastic kinetics take into account the fact that populations have discrete integer values and have a randomness associated with their movements and reactions (in our context, meetings). These computer simulations are useful for low numbers of molecules [137, 138]. The discrete and stochastic simulations we will be working with have the benefit of providing statistically exact sample paths, which we can analyse to gain knowledge on the distribution of the agents at a given time. This is helpful when modelling human dispersal since the population numbers we deal with are much smaller than those traditionally taken into account in deterministic models, making the system’s discreteness and stochasticity important for its behaviour.

Each simulation run gives only a single history for each family and a possible number of encounters which is unlikely to be repeated in multiple simulation runs, thus we repeat the simulation several times to better estimate the distribution of results and their deviations from the mean. The individual solutions we obtain from each simulation may differ from each other, but their ensemble converges to a more useful distribution. Computer simulations can be very useful, but they also have an

important drawback: the more accurate the simulation is, the longer it takes to run.

For our simulations we used a spatially-extended Gillespie SSA [73, 139]. This Gillespie SSA is a lattice-based simulation that tracks the number of molecules or individuals in each compartment for each species. Hence, it is very useful for spatial distributions, but not for tracking the path of each individual throughout the domain (once a molecule joins other molecules of the same species in any compartment its individuality is lost and we can not tell it apart from others). It is also meant to emulate a continuous time by generating two random numbers, r_1 and r_2 , uniformly distributed in $[0, 1)$, using the first one to generate the time of the next reaction, and the second one to choose which of the available reactions will take place. This is achieved with the use of propensity functions: if α_j is the propensity function of reaction j , then $\alpha_j \Delta t$ will be the probability of the reaction j occurring in a small time interval $[t, t + \Delta t]$.

If in our simulations we are at time t , then the time when the next reaction takes place will be $t + \tau$, with

$$\tau = \frac{1}{\alpha_0} \ln \left[\frac{1}{r_1} \right], \quad (4.1)$$

where r_1 is the first random number and α_0 is the sum of all the Q possible propensity functions:

$$\alpha_0 = \sum_{j=1}^Q \alpha_j(t). \quad (4.2)$$

After obtaining the next time, the Gillespie SSA searches for the reaction that will occur. This is done with the second random number, r_2 , and finding q such

that:

$$r_2 \geq \frac{1}{\alpha_0} \sum_{j=1}^{q-1} \alpha_j \quad \text{and} \quad r_2 < \frac{1}{\alpha_0} \sum_{j=1}^q \alpha_j.$$

This allows us to know that the reaction that takes place is the q^{th} reaction. This is equivalent to normalizing the sum of the Q reactions to lie within the interval $[0, 1]$ and giving each reaction a compartment of length proportional to the probability of it happening, then randomly choosing a point in the segment which corresponds to a given reaction.

There are other models of stochastic simulation, such as the velocity-jump process [140] and the Smoluchowski algorithm [141]. The latter has a set time-step and generates a normally distributed random number that dictates the distance a particular molecule or individual will move. This algorithm is better at following individuals across the domain and obtaining their trajectories. A velocity-jump process, on the other hand, changes the velocity of a molecule or individual stochastically, instead of its position, as in the Smoluchowski algorithm.

We considered two main methods for our stochastic simulations, the Gillespie SSA and the Smoluchowski algorithm. A quick comparison can be seen in Appendix C. We chose the spatially-extended Gillespie SSA because the spatial discretization allows us to explicitly control the separation that defines interacting and non-interacting families, as well as thinking ahead for our plans of including the environment into the simulations.

4.1.4 Data on Hunter-Gatherers

We were kindly pointed to available data by Dr. Eiluned Pearce. Thus we have data on biomes, and family sizes of different hunter-gatherer societies, as well as their area of residence, total population in said area, number of home reallocations in a year, and total displacement in a year. The data we are using have been put together by Binford [117], and are available in Grove et al. [3, Table 5.01]. An extract of them can be seen in Appendix D. Many of the parameter values for the simulations were derived based on these data.

4.1.5 Presented research

The work presented in this chapter is divided into two general segments: model building and then its applications. The segments explore and systematically build towards a model aimed at answering the question concerning the number of people a hunter-gatherer could meet by essential movement, whilst performing necessary tasks like collecting resources. For this it is necessary to study the diffusive movement of individuals, the different ways of incorporating ‘meeting’ events into the models, and the impact of memory inclusion, for the final integration of all these proposed attributes.

Section 4.2 will be divided into three subsections: 4.2.1 understanding simple diffusion, 4.2.2 the Immediate Meeting Model (IMM), and 4.2.3 the Meeting and Forgetting as a Reaction Model (MFRM). In Section 4.2.1 we explain the implementation of diffusion as an agent-based stochastic simulation, then we show the mean-field equivalent of diffusion, and finally we develop the analytical equivalent to

one agent diffusing in the computational compartment-based model. Section 4.2.2 explains the basic model of how we initially consider agent meeting. Section 4.2.3 studies a different approach to the meeting and forgetting events, treating them as possible reactions. The classical deterministic approximation of the MFRM is obtained as well as its steady state.

Section 4.3 provides a parallel comparison between the computational simulations and the mean-field equations. We show that the differences present between these two suggest the importance of stochastic dynamics in the system.

Section 4.4 makes use of the developed models to generate data whilst exploring the best way to classify these data. The input parameter values are based on available data from hunter-gatherers. Section 4.4.1 bases the IMM on the data we have available, firstly grouped into biomes and then individually as societies. In Section 4.4.2 we compute the outcome of the MFRM for a vast range of parameter values so we can create a look-up table to know which combinations of parameter values are likely to meet the desired social tier and which are not. We then analyse the results, realising there is a power-law relation to be exploited.

4.2 Mathematical models of meeting

We seek to provide a mathematical answer to the question: “given numerous environmental factors, how do animals, and especially humans, maintain large social networks throughout a year.” The hypothesis is that in some biomes, conditions are good enough for hunter-gatherers to meet everyone in their active network (estimated to be 150 individuals), mating pool (500 individuals), or even tribe (1500

individuals) within the year, whereas in environments where movement is inhibited, additional cohesion mechanisms, such as exchange of symbolic artefacts, may be required.

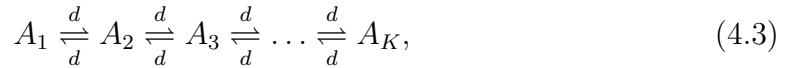
In this section we first describe the diffusive motion that our families will undergo in the simplest simulations, as well as obtaining an analytical solution to the discrete, small numbers case. Afterwards we describe the initial model, the “Immediate Meeting Model” (IMM), and run simulations for one and two spatial dimensions, before using the model to obtain information on different biomes and societies. Finally, we adapt the IMM to make it tractable to the classical deterministic description, hence treating each meeting and forgetting event as an event with reaction probability, rather than meeting immediately as soon as two hunter-gatherers are within detection distance, or forgetting families every 24 hours exactly. We finally compare this “Meeting and Forgetting as a Reaction Model” (MFRM) with the IMM and with a deterministic equivalent.

4.2.1 Diffusive motion

Although a strong assumption, random-walk-like movement is not completely unreasonable since hunter-gatherers did not always know the exact position of desired food resources (such as edible plants or prey). Fundamentally, we see this diffusive motion as a worst-case scenario in that if we could maintain a social network through random motion then it would be even easier in the case in which the agents were directed to one another or actively trying to meet up. The computational and analytical meanings of this diffusive motion are explained below.

4.2.1.1 Computational simulation of diffusion

Diffusion is simulated by considering a space-jump process over the discretized domain, which will only be a line section as we start with the one-dimensional case. Imagine a hypothetical species A (composed of families, instead of individuals) and we want it to diffuse across a domain of length L . Since the spatially extended SSA works on compartments, this domain will be divided into K compartments of equal length, such that each compartment has a length $h = L/K$. Here, x_k , with $k = 1, \dots, K$, will denote the compartment between $x = h(k - 1)$ and $x = hk$, or the k^{th} compartment. The diffusive movement can be seen as families transitioning from one compartment to a neighbouring compartment. To keep track of these movements, we assign a ‘population’ to each compartment, as can be seen in Figure 4.3. For example, if a family moves from compartment 3 to 2, then population A_3 loses a family at the same time as population A_2 gains it. Thus, the diffusion reaction is defined as



with $d = D/h^2$ being the diffusion rate constant, where D is the deterministic diffusion coefficient, and h the compartment length [73, 142]. Thus each population will have a location index $k = 1, \dots, K$.

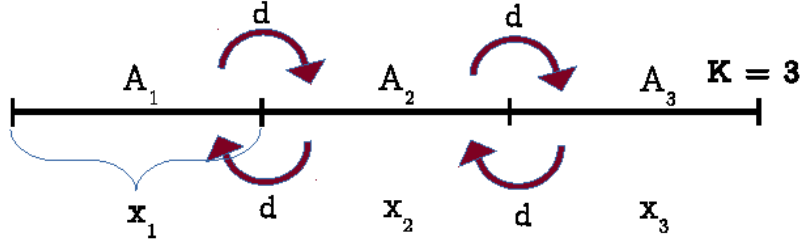


Figure 4.3: Schematic diagram of stochastic diffusion illustrating parameter and variable definitions.

The traditional approach now would be to obtain a continuous-time master equation and then its mean, as done in Appendix E. This would give us the classical deterministic equation for diffusion. Although this mean-field approach fits sufficiently well for one species in simple diffusion, as seen in Figure 4.7, we want a compartmentalised solution that maintains its stochasticity for use in further models with more types of interactions. We develop this in the next subsection.

4.2.1.2 Gambler's ruin

The mean-field approach may miss the stochasticity we are interested in observing. In our case the formulation also quickly becomes analytically intractable by keeping track of too many equations. Thus, we take a different approach. Instead of following changes in vacancy, we choose to follow the development of one family. This is similar to the so-called Gambler's ruin problem [143]. This formulation will deal with fewer equations and is more likely to be tractable analytically.

The problem refers to a gambler who starts gambling with a finite and discrete amount of money, deciding they will retire when they have either earned a chosen amount of, or lost all of, their money. We wonder about the probability of them winning rather than losing all their money given the amount they started with.

This relates to our problem, because, given the discretization on the domain in our algorithm, the steps taken by each family will be discrete, and it would be useful to know the probability of their whereabouts given a known starting point. The difference with our problem is that the Gambler's ruin problem is not concerned with the time taken between gambling events, and so can be thought of as evolving in discrete time, whereas our problem should evolve in continuous time.

We will solve the Gambler's ruin problem as a proxy for the diffusion of a particle. If we take $P(x_k, t)$ as the probability of a particle being at discrete site x_k at time t , then:

$$P(x_k, t + \Delta t) = P(x_{k-1}, t)d\Delta t + P(x_{k+1}, t)d\Delta t + (1 - 2d\Delta t)P(x_k, t), \quad (4.4)$$

for $k = 2, \dots, K - 1$. The first term on the right-hand side is defined as the probability of a particle being one position to the left at the previous time and jumping right during the time interval Δt with probability $d\Delta t$; the second term is the probability of the particle being one position to the right and jumping left; whereas the third term is the probability of the particle having been in the same position for the previous time and just staying there. We will also have reflective boundary conditions:

$$P(x_1, t + \Delta t) = P(x_2, t)d\Delta t + P(x_1, t)(1 - d\Delta t), \quad (4.5)$$

$$P(x_K, t + \Delta t) = P(x_{K-1}, t)d\Delta t + P(x_K, t)(1 - d\Delta t), \quad (4.6)$$

meaning that the probability of moving left is zero if the particle is at x_1 and the probability of moving right is zero if the particle is at x_K . We have initial condition:

$$P(x_k, t = 0) = \delta_{[x_k, x_i]}, \quad (4.7)$$

using the Kronecker delta for the particle starting at a particular position x_i .

Akin to the solution of the continuous deterministic diffusion equation, we assume that $P(x_k, t)$ can be represented as the product of two functions, each independently dependent on one of the variables: $P(x_k, t) = g(x)f(t)$. For simplicity, we assume the x in $g(x)$ takes the integer value that k in $P(x_k, t)$ had, hence $P(x_j, t) = g(j)f(t)$ for any $j = 1, \dots, K$. Upon separation we obtain:

$$\frac{f(t + \Delta t) - f(t)}{\Delta t f(t)} = d \frac{g(x - 1) - 2g(x) + g(x + 1)}{g(x)}, \quad (4.8)$$

and, in the usual way, we can set each side of Equation (4.8) to $-C$, where C is a positive constant.

In the case of time: $f(t + \Delta t) = (1 - C\Delta t)f(t)$, so, by induction:

$$f(m\Delta t) = (1 - C\Delta t)^m f(0). \quad (4.9)$$

From Equation (4.7), $g(x)f(0) = \delta_{[x_k, x_i]}$. Because the Kronecker delta does not depend on time, we can say $g(x) = \delta_{[x_k, x_i]}$ at initial time, and since the probability for the particle to be anywhere in the domain must equal one, $\int_0^L g(x)f(0) dx = 1$, therefore

$$f(0) \int_0^L g(x) dx = 1, \quad (4.10)$$

$$f(0) \int_0^L \delta_{[x_k, x_i]} dx = 1, \quad (4.11)$$

$$f(0) = 1. \quad (4.12)$$

We conclude that:

$$f(m\Delta t) = (1 - C\Delta t)^m. \quad (4.13)$$

By considering the continuous version of the diffusion equation we may anticipate that there will be several solutions for C , each denoted by C_q , and we will have a different $f_q(t)$ solution depending on the C_q :

$$f_q(t) = (1 - C_q \Delta t)^{t/\Delta t}. \quad (4.14)$$

If we apply the separation of variables to Equation (4.4) and the boundary conditions, we obtain a finite set of equations:

$$\begin{aligned} \left(\frac{C}{d} - 1\right) g(1) + g(2) &= 0, \\ g(x-1) + \left(\frac{C}{d} - 2\right) g(x) + g(x+1) &= 0, \\ &\dots \\ g(K-1) + \left(\frac{C}{d} - 1\right) g(K) &= 0. \end{aligned}$$

We can write this problem in matrix form $Ax = b$:

$$\begin{pmatrix} \frac{C}{d} - 1 & 1 & 0 & \dots & 0 \\ 1 & \frac{C}{d} - 2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{C}{d} - 1 \end{pmatrix}_{K \times K} \begin{pmatrix} g(1) \\ g(2) \\ \vdots \\ g(K) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (4.15)$$

We can split the matrix such that:

$$\begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}_{K \times K} \begin{pmatrix} g(1) \\ g(2) \\ \vdots \\ g(K) \end{pmatrix} = -\frac{C}{d} \begin{pmatrix} g(1) \\ g(2) \\ \vdots \\ g(K) \end{pmatrix}. \quad (4.16)$$

We want the eigenvalues and eigenvectors of this system. Each eigenvalue must be equal to a $-C_q/d$, because of the variable separation in Equation (4.8). From [144], Theorem 5, the eigenvalues for this matrix are:

$$\lambda_q = -2 + 2 \cos \left[\frac{(q-1)\pi}{K} \right], \quad \text{for } 1 \leq q \leq K. \quad (4.17)$$

Making $\lambda_q = -\frac{C_q}{d}$, this system will have K solutions for C_q , giving:

$$C_q = 2d \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K} \right] \right\}, \quad \text{for } 1 \leq q \leq K. \quad (4.18)$$

We also know from [144] that there is a different eigenvector for each eigenvalue, therefore we have K eigenvectors. If $\vec{u}^{(q)} = (u_1^{(q)}, \dots, u_K^{(q)})$ then, using [144], we know that for our case the eigenvectors have the form:

$$u_j^{(q)} = 1, \quad \text{for } j = 1, \dots, K, \text{ and } q = 1,$$

and:

$$u_j^{(q)} = \cos \left[\frac{(q-1)(2j-1)\pi}{2K} \right], \quad \text{for } j = 1, \dots, K, \text{ and } q = 2, \dots, K.$$

Then solutions for the eigenvectors are, given a particular $q = \hat{q}$, with $\hat{q} = 1, \dots, K$:

$$\begin{pmatrix} g_{\hat{q}}(x=1) \\ g_{\hat{q}}(x=2) \\ \vdots \\ g_{\hat{q}}(x=K) \end{pmatrix} = \begin{pmatrix} \cos \left[\frac{(\hat{q}-1)\pi}{2K} \right] \\ \cos \left[\frac{(\hat{q}-1)3\pi}{2K} \right] \\ \vdots \\ \cos \left[\frac{(\hat{q}-1)(2K-1)\pi}{2K} \right] \end{pmatrix}. \quad (4.19)$$

Since the solution will be a linear combination of the eigenvectors, we can now combine Equations (4.14) and (4.19) to obtain the solution:

$$\begin{aligned} P(x_k, t) &= \sum_{q=1}^K \gamma_q g_q(x) f_q(t) \\ &= \sum_{q=1}^K \gamma_q \cos \left[\frac{(q-1)(2k-1)\pi}{2K} \right] (1 - C_q \Delta t)^{t/\Delta t} \\ &= \sum_{q=1}^K \gamma_q \cos \left[\frac{(q-1)(2k-1)\pi}{2K} \right] \left(1 - 2d \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K} \right] \right\} \Delta t \right)^{t/\Delta t}, \end{aligned}$$

with each γ_q a constant.

To solve for γ_q , we must make use of the initial condition (4.7).

$$P(x_k, 0) = \delta_{[x_k, x_i]} = \sum_{q=1}^K \gamma_q \cos \left[\frac{(q-1)(2k-1)\pi}{2K} \right]. \quad (4.20)$$

Using the fact that the set of functions $\{\cos(n\pi x/L)_{n=0}^\infty\}$ are mutually orthogonal on $-L \leq x \leq L$, we deduce that:

$$\gamma_1 = \frac{1}{K}, \quad \gamma_q = \frac{2}{K} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K} \right), \quad \text{for } q = 2, \dots, K.$$

Therefore:

$$P(x_k, t) = \frac{1}{K} + \sum_{q=2}^K \frac{2}{K} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K} \right) \cos \left[\frac{(q-1)(2k-1)\pi}{2K} \right] \left(1 - 2d \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K} \right] \right\} \Delta t \right)^{t/\Delta t}. \quad (4.21)$$

In Figure 4.4 we can see the comparison over time between this equation and stochastic simulations, noting they agree.

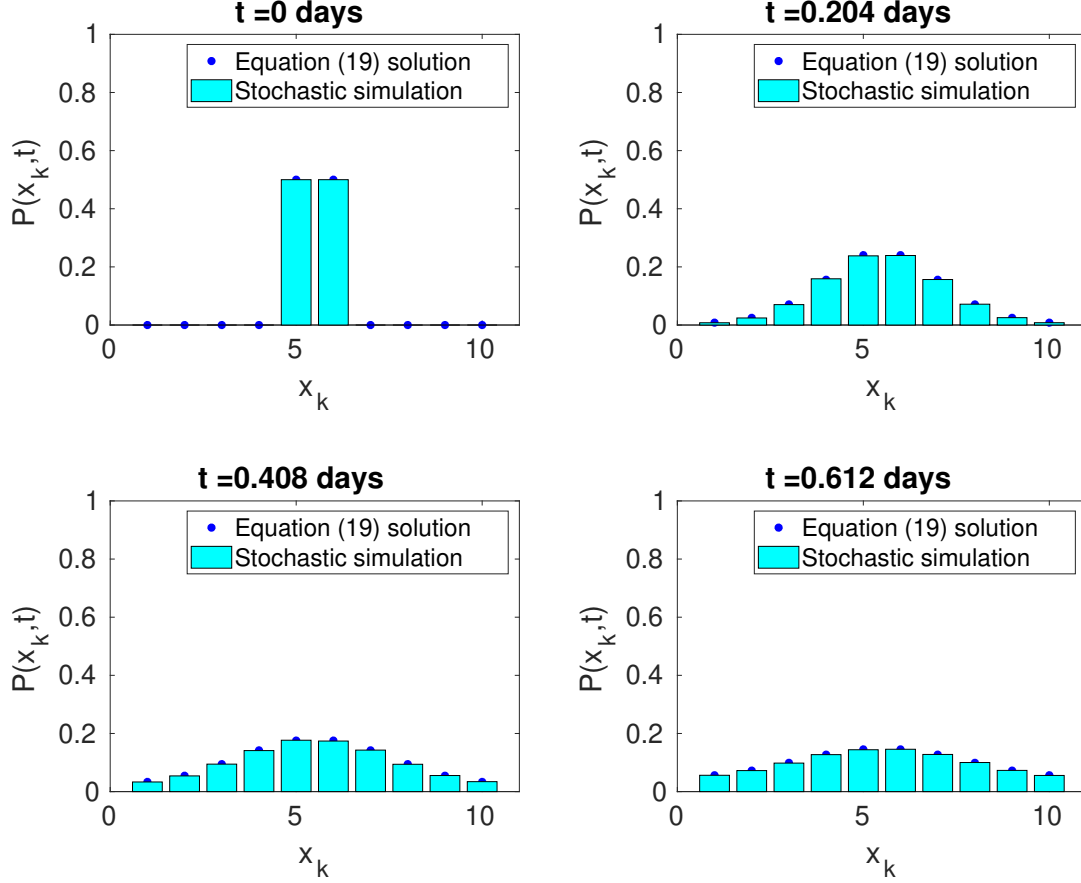


Figure 4.4: Comparison of the solution to our particular case of discrete space diffusion, Equation (4.21), in dark blue dots and the average of simulating 10^6 particles in light blue bars. Solution for $L = 10$, and an arbitrary $d = 6$ km/day². The independent variable, x_k , represents the k^{th} compartment, and the dependent variable, $P(x_k, t)$, is the probability for the particles to be in that compartment at the time t indicated at the top of each figure.

Since we are using continuous time in our model, we can use the fact that $(1 - x)^y \approx e^{-xy}$ for small x (remember that we can make Δt as small as needed) and approximate the time dependent term in Equation (4.21) as:

$$\left(1 - 2d \left\{1 - \cos \left[\frac{(q-1)\pi}{K} \right] \right\} \Delta t \right)^{t/\Delta t} \approx e^{(-2d \{1 - \cos [\frac{(q-1)\pi}{K}] \} t)} \quad (4.22)$$

Extending this to two dimensions in space, we have:

$$\begin{aligned}
P(x_k, y_l, t) = & \frac{1}{K_x K_y} \\
& + \sum_{q=2}^{K_x} \frac{2}{K_x K_y} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K_x} \right) \cos \left[\frac{(q-1)(2x_k-1)\pi}{2K_x} \right] \\
& \left(1 - 2d_x \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K_x} \right] \right\} \Delta t \right)^{t/\Delta t} \\
& + \sum_{p=2}^{K_y} \frac{2}{K_x K_y} \cos \left(\frac{(p-1)(2y_i-1)\pi}{2K_y} \right) \cos \left[\frac{(p-1)(2y_l-1)\pi}{2K_y} \right] \\
& \left(1 - 2d_y \left\{ 1 - \cos \left[\frac{(p-1)\pi}{K_y} \right] \right\} \Delta t \right)^{t/\Delta t} \\
& + \sum_{q=2}^{K_x} \sum_{p=2}^{K_y} \frac{4}{K_x K_y} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K_x} \right) \cos \left[\frac{(q-1)(2x_k-1)\pi}{2K_x} \right] \\
& \cos \left(\frac{(p-1)(2y_i-1)\pi}{2K_y} \right) \cos \left[\frac{(p-1)(2y_l-1)\pi}{2K_y} \right] \\
& \left(1 - 2d_x \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K_x} \right] \right\} \Delta t - 2d_y \left\{ 1 - \cos \left[\frac{(p-1)\pi}{K_y} \right] \right\} \Delta t \right)^{t/\Delta t},
\end{aligned} \tag{4.23}$$

where K_x is the number of compartments in the x direction, K_y the number of compartments in the y direction, x_k means the particle is in the k^{th} compartment in the x direction, y_l means it is in the l^{th} compartment in the y direction, d_x is the diffusion rate constant in the x direction, and d_y is the diffusion rate constant in the y direction.

We can use the same approximation as in Equation (4.22) to obtain the two-dimensional, discrete space, continuous time solution:

$$\begin{aligned}
P(x_k, y_l, t) = & \frac{1}{K_x K_y} \\
& + \sum_{q=2}^{K_x} \frac{2}{K_x K_y} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K_x} \right) \cos \left[\frac{(q-1)(2x_k-1)\pi}{2K_x} \right] \\
& \exp \left(-2d_x \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K_x} \right] \right\} t \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{p=2}^{K_y} \frac{2}{K_x K_y} \cos \left(\frac{(p-1)(2y_i-1)\pi}{2K_y} \right) \cos \left[\frac{(p-1)(2y_l-1)\pi}{2K_y} \right] \\
& \exp \left(-2d_y \left\{ 1 - \cos \left[\frac{(p-1)\pi}{K_y} \right] \right\} t \right) \\
& + \sum_{q=2}^{K_x} \sum_{p=2}^{K_y} \frac{4}{K_x K_y} \cos \left(\frac{(q-1)(2x_i-1)\pi}{2K_x} \right) \cos \left[\frac{(q-1)(2x_k-1)\pi}{2K_x} \right] \\
& \cos \left(\frac{(p-1)(2y_i-1)\pi}{2K_y} \right) \cos \left[\frac{(p-1)(2y_l-1)\pi}{2K_y} \right] \\
& \exp \left(-2d_x \left\{ 1 - \cos \left[\frac{(q-1)\pi}{K_x} \right] \right\} t - 2d_y \left\{ 1 - \cos \left[\frac{(p-1)\pi}{K_y} \right] \right\} t \right). \quad (4.24)
\end{aligned}$$

We can see then that $P(x_k, y_l, t) = P(x_k, t)P(y_l, t)$ as expected, as the probability of a particle moving in one direction is independent of it moving in the other direction.

A solution to this type of problem has been attempted several times, but with subtle differences. For example, since 1943 the famous astrophysicist and Nobel winner Chandrasekhar, worked on a similar problem [145], but provided only one boundary, as opposed to our solution with two or four boundaries, depending on the dimension. Lindeberg [146] also solved the discrete random walk problem, but they assumed an infinite domain and do not take into account boundary conditions. This means that they use an infinite number of Bessel functions, which then have to be truncated for numerical solutions. Our solutions is exact, no truncation was necessary. The random walk problem, as a discrete diffusion equation, was also reviewed by Campbell [147], where they used modified Bessel functions of the first kind to solve the problem. Firstly, our work extends to two dimensions, whereas Campbell only solved the one-dimensional problem. Further, although it may be possible to extend Campbell's work to two dimension, we supply a direct trigonometric relationship for the solution, whereas Campbell used a form of discrete Fourier transform

providing a set of difference equations, which can be used to generate an infinite sum of characteristic functions. Solutions found by the author seem different to the one here presented, but the ongoing interest in this problem throughout time shows its importance for many fields.

4.2.2 Immediate Meeting Model

Having derived an equation for movement we will now start to consider more interactions. We begin by adding the meeting interaction. The study of the one-dimensional interaction is presented in Appendix F, while here we show the two-dimensional extension. We will assume that a group of humans (which we will assume as five when not known) form a cohesive family that moves as a whole, as we assume the clique (a small group of people, Figure 4.1) always travels together. These families will have free, random movement. The domain in which these families move will have reflective boundaries, assuming that families will not fall off the domain or be trapped by its boundaries. As face-to-face interaction seems to be essential for social bonding, each encounter, or meeting, strengthens a link. Each family will have a detection distance, presently of $1/2$ km, as used in Pearce [2], such that a person may see other people half a kilometre away and, since strengthening the bond is desirable, once they see each other we consider these families to immediately have met. Finally, a memory function will be incorporated, allowing families to see each other several times within a particular time-frame without adding to the number of times they are considered to have met. For example, families will only be considered to have met once within any single day, no matter how many times

they interact during that day. We call this the Immediate Meeting Model (IMM).

We make the distinction that an encounter will be assumed to have occurred whenever two families come within the detection distance, whilst a meeting is assumed to occur only when encounters happen after the families have “forgotten” each other. That is, each meeting is an encounter, but an encounter will be a meeting only after a “forgetting time” has passed from the previous meeting. This “forgetting time” is initially set to 1 day (24 hours).

Because computationally simulating a large number of individuals takes a long time, we used the assumption that humans would not live or hunt alone, and would therefore always travel in cliques or families. This reduces the need to simulate every person in a domain and reduces the running time whilst being plausible for hunter-gatherers.

For quantification purposes, we will focus on one ‘main’ family and study its encounters with every other family in the domain. Our model will have three different types of families: Y (for the subject family), U (for the unmet families), and M (for the met families).

For our simulations, initially there will only be one Y family and N unmet families that can be met. We will enumerate each of the families of type U and M so that we can maintain their individuality and keep count of where they are and how many times each has met with Y , therefore the ‘unmet’ families will be denoted U_i , $i = 1, \dots, N$, and the ‘met’ families will be denoted M_i , $i = 1, \dots, N$.

These dynamics are:

$$Y + U_i \longrightarrow Y + M_i, \quad i = 1, \dots, N. \quad (4.25)$$

This means that whenever family Y meets one of the other families, that family turns into a met family, but the Y family remains unchanged. Since families only convert between met and unmet species, the total number of both met and unmet families is constant. Family Y is initiated at the middle of the domain whilst all the unmet families are uniformly spread out across the domain. The definitions used throughout the chapter can be seen in Table 4.2.

We expect these assumptions to give us a lower bound to the number of meetings a family could have for a constant environment. This is because hunter-gatherers may share a sleeping space with their sympathy group (15 people) or even affiliation group (50 people), as seen in Section 4.1.1, making the meeting of those layers more frequent than predicted here, or, their will to meet other members in their active network (150 people) may take them to agreed meeting spaces. On the other hand, once other environmental factors, like resources, start varying, the number of meetings could be even lower, because the lack of resources in certain areas could, essentially, “trap” the families in isolated spots. We will start considering such complications in Section 4.6.1, Future Work.

Variable	Definition	Unit
L	Length of one side of domain.	km
K	Number of compartments in one direction.	
h	Length of compartments or detection distance, $h = L/K$.	km
D	Deterministic diffusion coefficient.	km ² /day
d	Diffusion rate constant, $d = D/h^2$.	1/day
x_k	k^{th} compartment.	
A	Species A .	
A_k	Population of species A in compartment k .	
n_k	Number of families of species A in compartment k .	
t	Time.	day
$p(\vec{n}, t)$	Joint probability of population A_k having n_k families at time t .	
ν	Volume of system where reaction takes place.	km or km ²
$P(x_k, t)$	Probability of a particle being at discrete site x_k at time t .	
N	Number of “other” families.	
Y	Main family, to be followed and whose interactions will be quantified.	
U_i	Unmet family i . Can turn into a met family and back.	
M_i	Met family i . Has been met by family Y .	
μ	Meeting rate.	families/day
f	Forgetting rate.	families/day
$Y(x, t)$	Density of family Y at point x and time t .	
$U(x, t)$	Density of family U at point x and time t .	
$M(x, t)$	Density of family M at point x and time t .	

Table 4.2: Definition of symbols/abbreviations used throughout this chapter.

4.2.2.1 Particle movement in two dimensions

As seen in Section 4.2.1, and following from Appendix F.2 (where we focus on the one-dimensional case), not only will each family have an index identifying it, $k = 1, \dots, N$, but they will also have a location index for each dimension: (x, y) . Here, $x = 1, \dots, K$ specifies the i^{th} compartment in the horizontal location, whilst

$y = 1, \dots, K$ specifies the j^{th} compartment in the longitudinal location. The domain will be divided into $K \times K$ compartments of 1 km^2 each, for detection distance. Therefore, $M_{f,(x,y)}$, for example, will be met family f in the compartment (x, y) . Throughout the simulation all of these families diffuse. The position of any family in one box does not exclude another family occupying that box as well. The unmet species can convert to and from the met species, which means the unmet families have been met and then forgotten, turning back to unmet. Thus we have $N + 1$ families, but $2N + 1$ species, spread over K^2 spatial compartments in two dimensions. The algorithm will keep track of $K^2 \times (2N + 1)$ different “populations”. In total, the reactions will be:

$$Y_{(1,y)} \xrightleftharpoons[d]{d} Y_{(2,y)} \xrightleftharpoons[d]{d} Y_{(3,y)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} Y_{(K,y)}, \quad (4.26)$$

$$Y_{(x,1)} \xrightleftharpoons[d]{d} Y_{(x,2)} \xrightleftharpoons[d]{d} Y_{(x,3)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} Y_{(x,K)}, \quad (4.27)$$

$$U_{f,(1,y)} \xrightleftharpoons[d]{d} U_{f,(2,y)} \xrightleftharpoons[d]{d} U_{f,(3,y)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} U_{f,(K,y)}, \quad (4.28)$$

$$U_{f,(x,1)} \xrightleftharpoons[d]{d} U_{f,(x,2)} \xrightleftharpoons[d]{d} U_{f,(x,3)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} U_{f,(x,K)}, \quad (4.29)$$

$$M_{f,(1,y)} \xrightleftharpoons[d]{d} M_{f,(2,y)} \xrightleftharpoons[d]{d} M_{f,(3,y)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} M_{f,(K,y)}, \quad (4.30)$$

$$M_{f,(x,1)} \xrightleftharpoons[d]{d} M_{f,(x,2)} \xrightleftharpoons[d]{d} M_{f,(x,3)} \xrightleftharpoons[d]{d} \dots \xrightleftharpoons[d]{d} M_{f,(x,K)}, \quad (4.31)$$

$$Y_{(x,y)} + U_{f,(x,y)} \longrightarrow Y_{(x,y)} + M_{f,(x,y)}, \quad (4.32)$$

$$M_{f,(x,y)} \longrightarrow U_{f,(x,y)}, \quad (4.33)$$

where Y is for the main family we are following, M is for met families, and U is for unmet families, all for $f = 1, 2, \dots, N$, $x = 1, 2, \dots, K$, and $y = 1, 2, \dots, K$. d is the diffusion rate. We coded these reactions into the Gillespie SSA and obtained

position and number of meetings for each family every day for a year.

Since we assume that a family is composed of 5 people, we satisfy the clique social layer meetings on a daily basis. To achieve the number of meetings needed for the sympathy group, the 15-individual layer, family Y has to meet 2 other families. This would mean each member of family Y comes into contact with each member of the other two families, each achieving their sympathy group connections. The affiliation group (or residential band) is covered when family Y meets 9 other families. The next layer corresponds to the active network, of 150 individuals, for which family Y would have to meet 29 other families.

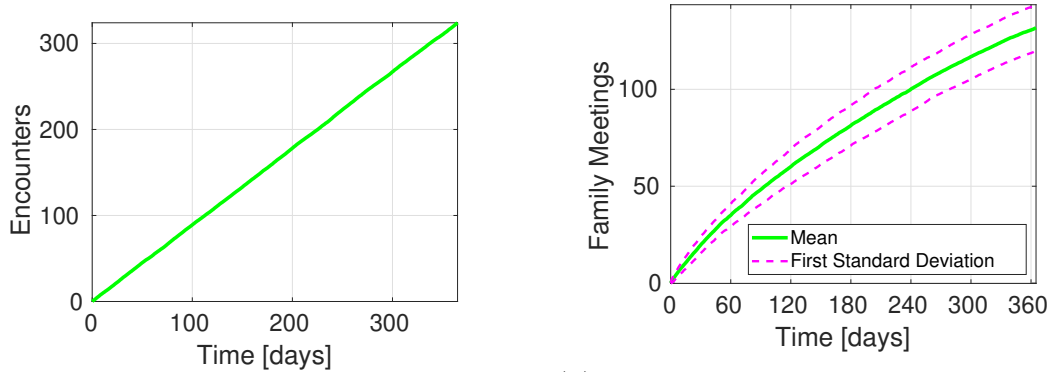
The way we will visualise our results is via three main quantities of interest:

1. The total number of encounters throughout the year.
2. The total number of new meetings throughout the year.
3. The number of times each family was met by Y .

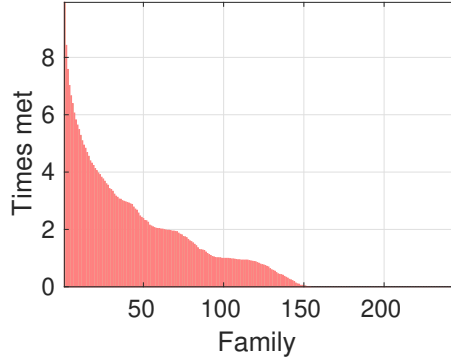
The first quantity is always a straight line as long as the forgetting time is small compared to a year (see Figure 4.5a). This is as expected because, according to the ideal gas model [126], the number of encounters an individual experiences when moving randomly is directly proportional to the individual's speed, their detection distance, the density of individuals, and time. Since the former three entities are constant in our model, the number of total meetings will be directly proportional to time only, assuming memory plays a proportionally small role.

The second quantity is a measure of how many families have been met throughout the year, counting only the first time Y met them. This graph tells us if our Y family

has met 150 people and how long it took for them to meet (see Figure 4.5b). This quantity can initially be approximated with a line, much like the ideal gas model, but its slope starts decreasing as soon as all nearby families have been met and family Y has to go further to find new families. This curve has a maximum where all available families have been met, and this will happen sooner or later depending on the combination of all parameter values. We note that the total number of families met must be bounded because of the finite number of families available to meet.



(a) Total number of encounters by family Y . (b) Total number of new meetings by family Y .



(c) Number of times each family was met by Y at $t = 365$ days.

Figure 4.5: Example solution from the Gillespie SSA of the model described in Section 4.2.2, for 2D. Average of 100 runs. Parameter values used were: 247 unmet families initially, forgetting time of a day, length of $71 \times 71 \text{ km}^2$, detection distance of 1 km, so each compartment is of 1 km^2 , and diffusion coefficient of $4.7 \text{ km}^2/\text{day}$. As this was taken from the North America: California and Northern Mexico biome, as seen in Section 4.4.1.1, the parameter values used here were the ones used for the BiomeMid case.

For the third quantity we order the met families from most met to least met in each run and then average over all the runs, obtaining an indicator of how many times a person in family Y will meet with each other person (see Figure 4.5c). This graph shows that a small number of families will be met several times, whilst most families will only be met a few times.

This model was applied to data from known hunter-gatherer societies, as can be seen in Section 4.4.1. The IMM was useful for computational simulations, but was difficult to translate fully to analytical equations because of the location-specific immediate meetings and the time-dependent forgetting actions. This is why we decided to build on this model by making meetings and forgetting actions dependent on a probability of reaction, rather than being immediate or time-dependent, which would better resemble a reaction-diffusion framework, as used in Baker [7]. This is done in the next subsection.

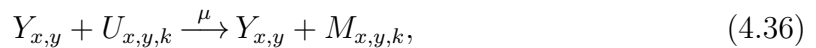
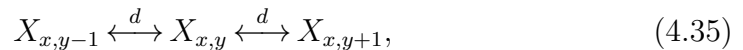
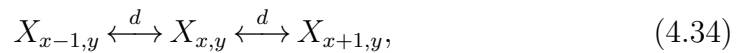
4.2.3 Meeting and Forgetting as a Reaction Model

As stated in Section 4.2.2, the original model has two events determined precisely by distance or time. The IMM has a detection distance for each family within which families Y and U would immediately meet. It also has a set forgetting time after which a met family will be forgotten. In order to translate this behaviour into reaction-diffusion equations, and following the molecular dynamics approach, we will treat both events as reactions with a reaction probability μ for meeting and f for forgetting. This is what we call the Meeting and Forgetting as a Reaction Model (MFRM).

This new model may also be closer to a real-life scenario. Treating the meeting as a reaction means there is a probability that two families will meet once they come into the same compartment, but there is also a possibility that the families will go in and out of the compartment without noticing each other. An immediate meeting would mean infinite μ , so, in the future, we can approximate the IMM by making μ very large. In simulations we will choose a high meeting rate such that meeting is much more probable than moving out.

Treating the forgetting event as a probability as well may also more closely resemble human behaviour. We assume an encounter counts towards strengthening the bond between two people but a certain amount of time has to pass between encounters for them to count towards the strengthening of the bond, or for them to be called “a meeting” in our paradigm. In the IMM we assume that an encounter 24 hours after the first encounter will count as a meeting. But humans are not exact in this sense, and two encounters with 18 hours between them can strengthen the bond between the people, or it can take longer than a day for these encounters to count as “meetings”. In this way, making the forgetting event a rate where the average is 24 hours, but leeway exists for the natural randomness of human interaction, may be closer to reality.

The reaction model in two dimensions is as follows:



$$M_{x,y,k} \xrightarrow{f} U_{x,y,k}, \quad (4.37)$$

where X can be family Y , U or M , d is the diffusion rate, with units of day^{-1} , μ is the meeting rate, with units of families/day, and f is the forgetting rate, with units of families/day as well. We assume the domain is isolated as in Equations (4.26) to (4.31) above. Thus, at each boundary the populations only have the ability to jump back into the domain. Equation (4.36) is the meeting action and Equation (4.37) is the forgetting event each family goes through after being met. In both, k refers to the specific U (unmet) or M (met) family.

The propensity function for meeting will depend on the volume, ν , of the compartments in which the bimolecular reaction happens:

$$\alpha_m(t) = Y_{x,y}(t)U_{x,y,k}(t)\mu/\nu, \quad k = 1, \dots, N, \quad (4.38)$$

with the units of ν being km when in one dimension, and km^2 when in two dimensions.

The propensity function for forgetting will depend on both families being close to each other, which is similar to what was done for the proteins in the previous two chapters.

$$\alpha_f(t) = M_{x,y}(t)f/\nu, \quad k = 1, \dots, N, \quad (4.39)$$

with the units of f being families per day.

4.2.3.1 Continuous time and space: mean-field approximation

We will obtain the mean-field approximation for future comparison. We will work in one dimension for simplicity. Assuming $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$ with d such that

$d\Delta x^2 = D$, D a positive constant, then $\mathcal{O}(\Delta x^3)$ in the Taylor expansion of the master equation will be negligible compared to the remaining terms and we get the classical deterministic description:

$$\frac{\partial Y(x, t)}{\partial t} = D \frac{\partial^2 Y(x, t)}{\partial x^2}, \quad (4.40)$$

where $Y(x, t)$ is the density of family Y at point x and time t .

Using the law of mass action, we obtain the mean-field approximation for the density of families U and M , respectively:

$$\frac{\partial U(x, t)}{\partial t} = D \frac{\partial^2 U(x, t)}{\partial x^2} - \mu Y(x, t)U(x, t) + fM(x, t), \quad (4.41)$$

$$\frac{\partial M(x, t)}{\partial t} = D \frac{\partial^2 M(x, t)}{\partial x^2} + \mu Y(x, t)U(x, t) - fM(x, t). \quad (4.42)$$

These three equations above are our system of deterministic PDEs for the MFRM.

We will use reflective boundary conditions, also known as Neumann boundary conditions:

$$\left. \frac{\partial X(x, t)}{\partial x} \right|_{x=0} = 0, \quad (4.43)$$

$$\left. \frac{\partial X(x, t)}{\partial x} \right|_{x=L} = 0, \quad (4.44)$$

where X stands for family Y , M , or U . We will also have family Y at the middle of the domain and the other N unmet families spread homogeneously across the domain of length L as initial conditions.

We can solve Equation (4.40) analytically, for which the general solution is:

$$Y(x, t) = \sum_{n=0}^{\infty} \alpha_n \cos\left(\frac{n\pi x}{L}\right) e^{-D(n\pi/L)^2 t}, \quad (4.45)$$

with α_n being constants to be determined from the initial conditions.

For the initial condition:

$$Y(x, t = 0) = \delta(x - L/2), \quad (4.46)$$

which is our initial configuration for family Y , we obtain:

$$Y(x, t) = \frac{1}{L} + \sum_{m=1}^{\infty} \frac{2}{L} (-1)^m \cos\left(\frac{2m\pi x}{L}\right) e^{-D(2m\pi/L)^2 t}. \quad (4.47)$$

We substitute this expression for Y into Equations (4.41) and (4.42) to obtain a pair of PDEs for U and M which we solve numerically.

The steady state of these coupled PDEs is:

$$Y(x) = 1/L, \quad (4.48)$$

$$U(x) = N \frac{f}{\mu + Lf}, \quad (4.49)$$

$$M(x) = \frac{N}{L} - U(x). \quad (4.50)$$

We can see that varying f from 0 to ∞ , any number of unmet people can be the steady state, from 0 to the N families across the domain. This final result does not mean they were not met, it just means they can be forgotten quicker than met.

4.3 Comparing Models

We now want to compare the three models:

- the IMM stochastic model,
- a MFRM stochastic model, and
- the expression for Y in (4.40) and the numerical solution to Equations (4.41) and (4.42) (the mean-field approximation of the MFRM).

As we are initially interested in the total number of new families met, counting only the first meeting between family Y and any family U should suffice. Therefore we only track family U up until the first meeting with family Y , as if it had never been forgotten. This means making $f = 0$ for Equations (4.37), (4.41), and (4.42). The evolution in time for the U families can be seen in Figure 4.6 and the evolution for Y can be seen in Figure 4.7. The M families are not shown as they are given by: $M(x, t) = N/L - U(x, t)$. The standard deviations of the simulations for U , from Figure 4.6, can be seen in Figure 4.8. Although $f = 1$ was used for the MFRM simulations, we counted only the first time a U family was met, so f is effectively zero in the comparison plots.

We can see in the solutions for U that, whilst the numerical solution to the PDEs seems to agree with the MFRM simulation at the beginning, by $t = 120$ days it diverges from the stochastic simulation results and drops faster, almost agreeing with the stochastic simulations for the IMM at the end. In fact, if we compare newly met families, as seen in Figure 4.9, the numerical solution crosses over both stochastic simulations, but does not converge with either. The fact that the meetings from the PDEs seem to be faster than the meetings from the MFRM simulations could be due to the fact that the PDEs take into account a probability of meeting with all the U families in the domain, however small that probability might be, whilst in the MRM simulations, family Y can only meet with the U families within its detection distance. That is, the meeting rate is the same in both models, but one considers only the families in Y 's compartment and the other considers the whole domain. For large numbers of Y and U families both models should be in

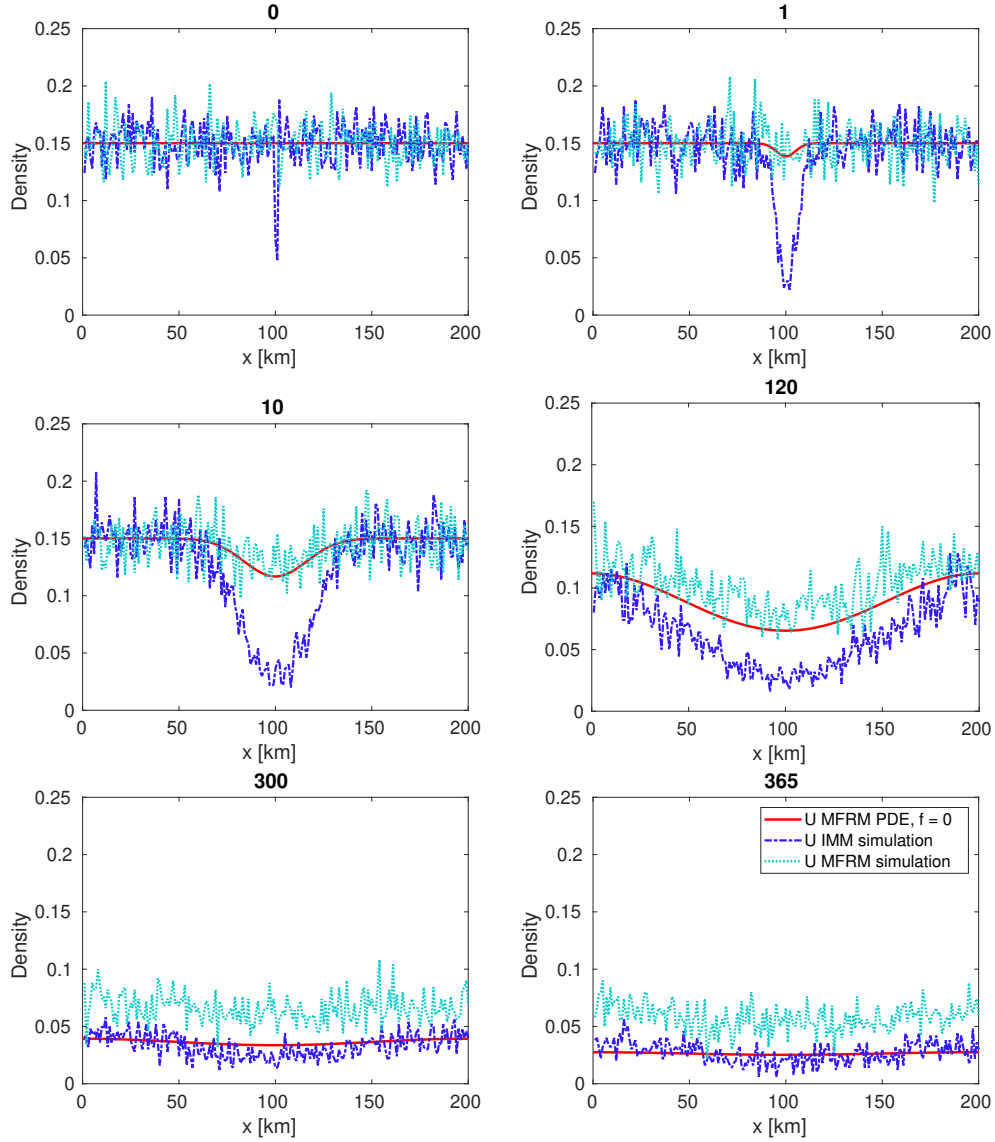


Figure 4.6: Density of families that were not yet met by the tracked family at different time points for three different models. A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200$ km, with diffusion constant, D , of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are the numerical solutions of the PDE Equation (4.41) for U and the stochastic simulations that used the IMM and MFRM models. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MFRM simulations a forgetting rate of $f = 1$ per day was used, although $f = 0$ was used for the numerical solution of Equation (4.41). Results shown are an average of 500 simulations each, where applicable.

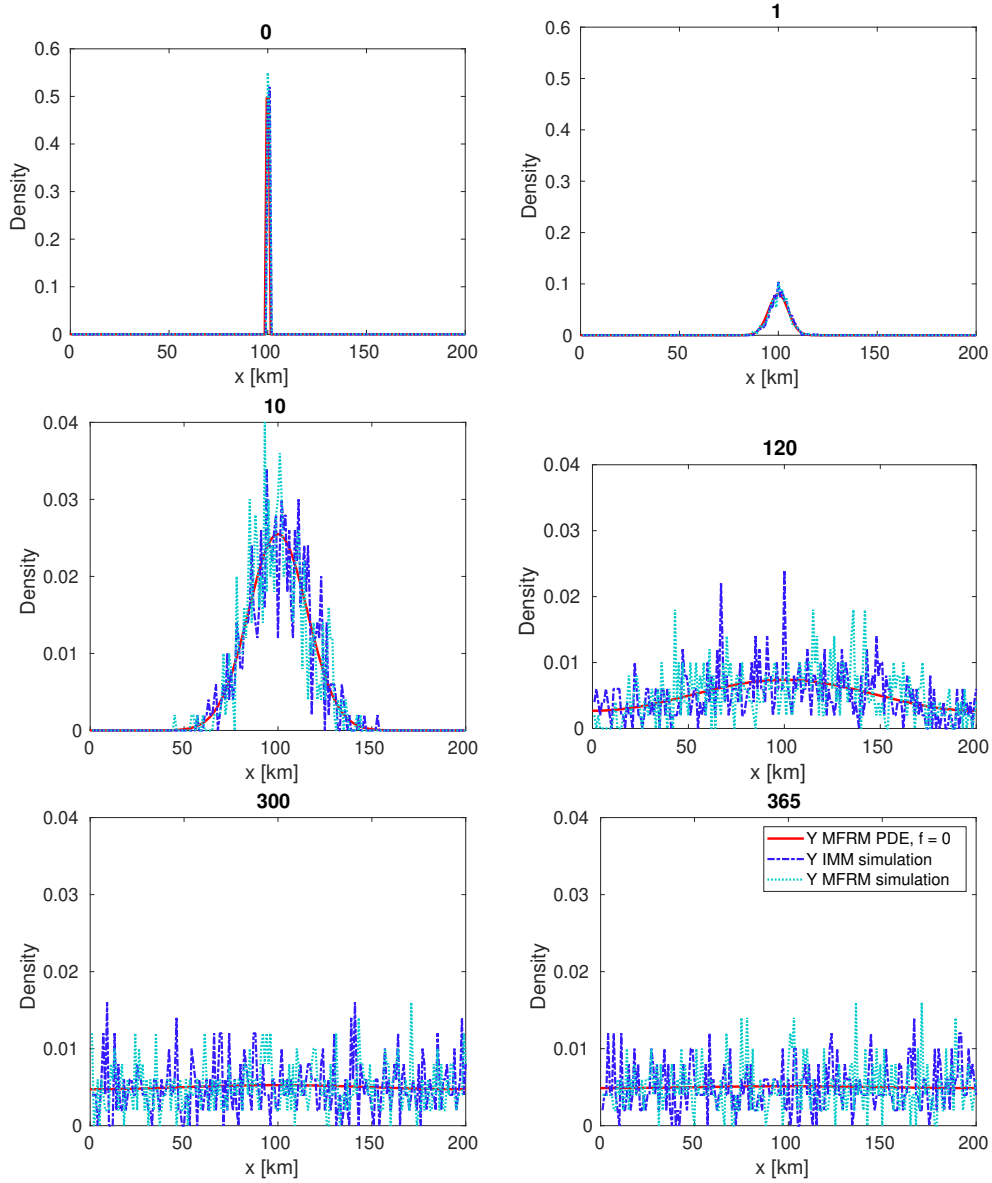


Figure 4.7: Density of tracked family at different time points for three different models. A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200$ km, with diffusion constant D of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are the numerical solutions of the PDEs Equations (4.40) for Y and the stochastic simulations that used the IMM and MFRM models. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MRM a forgetting time of one day is used, for the MFRM a forgetting rate of $f = 1$ per day was used. Results shown are an average of 500 simulations each, where applicable.

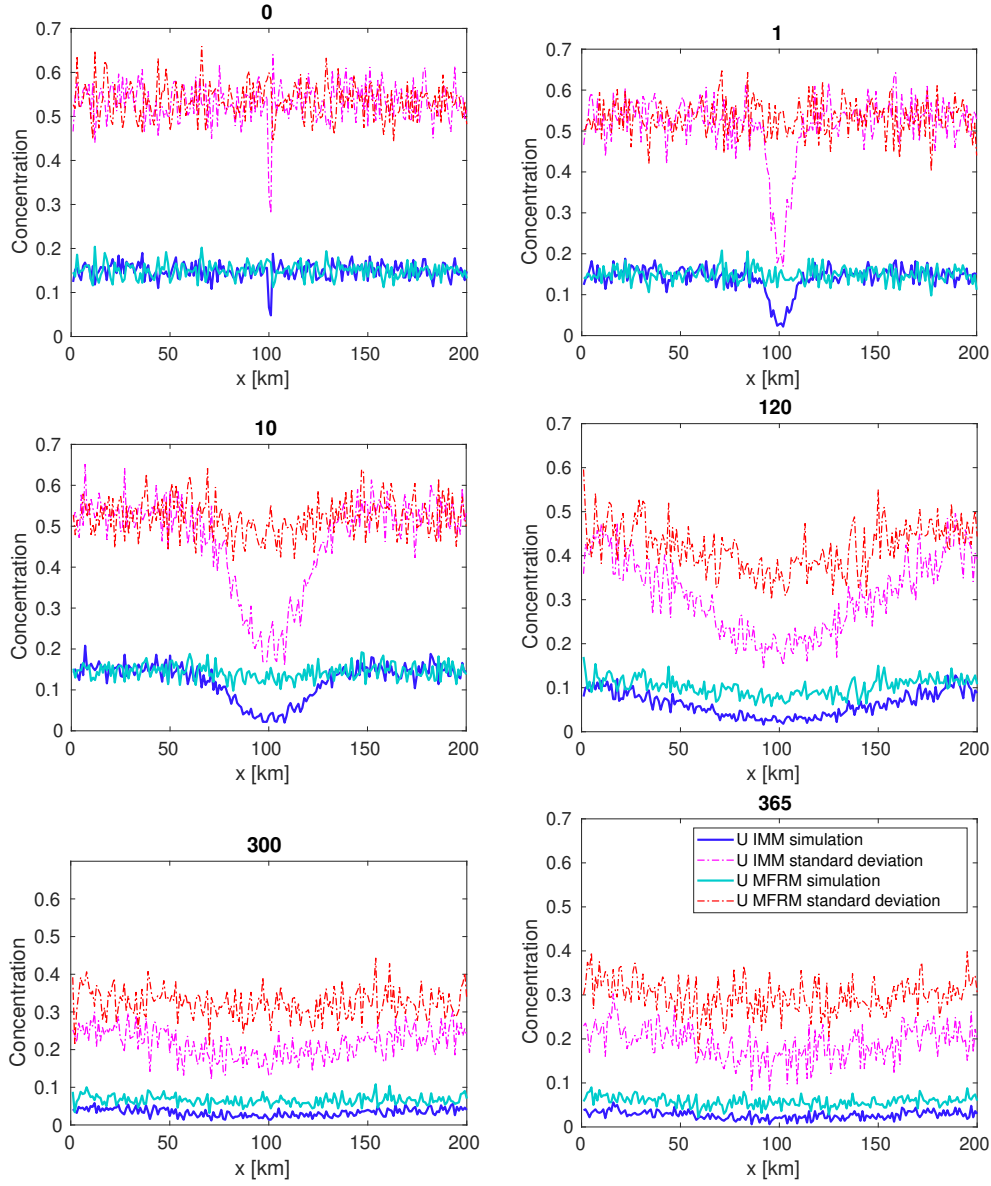


Figure 4.8: A family Y and 30 other unmet families moving in a one-dimensional domain of length $L = 200$ km, with diffusion constant, D , of $12.25 \text{ km}^2/\text{day}$, day specified at the top of each graph. Shown are stochastic simulations of the IMM and MFRM models. The average of 500 simulations for each are shown in blue, whilst their respective standard deviations are shown in red and magenta. A detection distance of 1 km is used. Meeting rates of $\mu = 1$ per day were used for the corresponding models. For the MFRM simulations a forgetting rate of $f = 1$ per day was used.

agreement, but verifying this is computationally expensive, so it is left for future work. However, as we only have one Y family, stochasticity seems important and this could be playing an essential role in the slower meetings.

We note that the standard deviation in Figure 4.8 does seem large compared to the difference between the simulation models, but it follows the same pattern as each corresponding mean line, making us think that the overarching difference in behaviour between both models is captured.

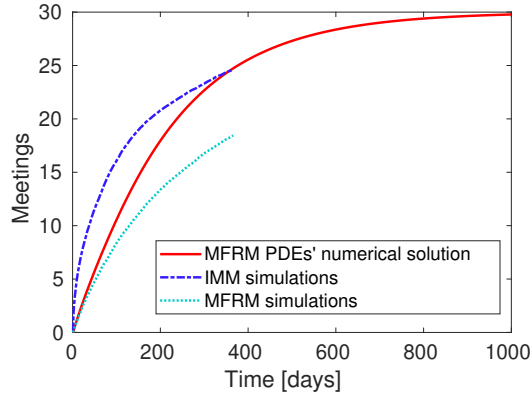


Figure 4.9: The number of newly met families for: the numerical solutions of the PDEs, Equations (4.40) and (4.41); the IMM stochastic; and the MFRM stochastic simulations. Parameter values as in Figure 4.6. The stochastic simulations (blue, cyan) were only run until day 365 because that is the range of interest. The numerical solution of the PDEs (red) is not as computationally intensive to run, so can be run for a longer end time. Therefore, it is shown until it reaches the maximum (30 families in the domain) to show this limit is reached. Results shown are an average of 500 simulations each, where applicable.

Another difference between the IMM and the MFRM is that the MFRM takes longer to simulate. This is because the extra number of reactions of around the same time-scale as the diffusion, or smaller, means that more events will happen for the MFRM than for the IMM in the same time interval. The algorithm will have to update the whole system and generate new random numbers after each event,

instead of just being based on the time that has passed or the family's position. Despite this disadvantage for the MFRM, its closer similarity to human behaviour and the possibility of translating it into our diffusion-reaction scheme makes it our chosen model for future application and development.

4.4 Model Application

Subsequently, we applied the IMM and the MFRM to study the relationship between biome properties and hunter-gatherer meetings. Specifically, we sought to understand under which conditions does an observed family meet the Dunbar number of 150 meetings per year in a hunter-gatherer society. We informed the models using anthropological data on societies (their location, population size, movement velocity, area of roaming, and mean family size) as provided by our collaborators Prof. Robin Dunbar and Dr. Eiluned Pearce. The data are taken from Binford's work [117] and can be seen in the Appendix, Table D.1.

The application of the IMM in Section 4.4.1 gives us insight into realistic parameter values and whether biomes are a good-enough split. With this information, and having developed a more complete model in parallel (the MFRM), we then applied the MFRM to a range of plausible parameter values in Section 4.4.2. We here present these applications and, in the end, a table for sets of parameter values that could, or could not, reach the 150-people threshold.

4.4.1 Immediate Meeting Model

4.4.1.1 Application of IMM to biomes

We want to know in which cases our family could reach the Dunbar number of 150 meetings in a year for a society of hunter-gatherers. We have the data from Binford’s work [117] (Table D.1) to inform the parameter values used for our model. The data contain, for 178 different societies across the world, data on: location, absolute latitude, general biome, population area (km^2), number of times the society moved base-camp in a year, distance moved in the movement of base-camp, population, and mean size of family. But because our model has to be solved on a case-by-case scenario and the simulations are time consuming, we speculated that there could be a way of grouping cases such that the results for one case would fit several societies, so we decided to group the societies by biomes: areas of land with similar characteristics as divided by Binford [117].

We used three distinct values for each biome to cover uncertainty about their precise values: “BiomeMin” (using the parameter values that would achieve the least amount of meetings for family Y), “BiomeMax” (using the parameter values that would achieve the most meetings for family Y), and “BiomeMid” (using the midrange of each parameter to create a plausible mid-range amount of meetings). We take the information from Table D.1 to determine where the biologically realistic boundaries in family size, people density, speed of movement, and area of domain are.

For the number of initially unmet families, we divided the population of each data entry in Table D.1 over its mean family size (or over 5 when the mean family

size was not available), then rounded to the closest integer. The maximum for each biome was used for the BiomeMax case, the minimum of each biome was used for the BiomeMin case, and an average was taken for the BiomeMid case.

The length of each side of the area was calculated as the square root of the total area, then rounded to the closest integer. The maximum for each biome was used for the BiomeMin case, the minimum of each biome was used for the BiomeMax case, and an average was taken for the BiomeMid case.

For the speed of movement of each family the distance walked in a day was calculated using a combination of Layton et al. [111, p 1230] and Table D.1. This gave us the diffusion coefficient for our model. In most cases, displacement in a day was taken from Layton et al. [111] where available or, if not, we took from Table D.1 the distance moved in a year over the number of displacements in a year divided by 3 to fall within acceptable walking ranges. This is all more fully explained in Appendix G.

We can see in Tables 4.3, 4.4, and 4.5 a summary of the number of families met (and their timing) for the BiomeMin, BiomeMid, and BiomeMax cases, respectively. Each result shown is the average of 100 simulations. In each table we see the total number of meetings family Y had with other families by the end of the year, the total number of distinct people met by the end of the year, the time family Y took before meeting each of its social layers, and the number of times the most frequently met family was met by the end of the year. Figures for more detailed results can be seen in Appendix H.

In general, none of the BiomeMin cases reached the 2 family meetings that would

Biome	Times other families were met	Number of distinct people met	Days to meet 15 people	Days to meet 50 people	Days to meet 150 people	Days to meet 500 people	Times most frequently met family was met
Arctic	0.6	1.9	Not met	Not met	Not met	Not met	0.5
Australia	0.1	0.3	Not met	Not met	Not met	Not met	0.1
North America: California and Northern Mexico	2.6	5.5	Not met	Not met	Not met	Not met	1.9
North America: desert and desert scrub	0.2	0.4	Not met	Not met	Not met	Not met	0.2
Subarctic and continental midlatitude forest in North America and Asia	0.3	0.7	Not met	Not met	Not met	Not met	0.3
Tropical and subtropical Africa	2.6	5.4	Not met	Not met	Not met	Not met	1.8
Tropical and subtropical Asia	1.1	2.0	Not met	Not met	Not met	Not met	1.0
Tropical and subtropical South America	0.0	0.0	Not met	Not met	Not met	Not met	0.0

Table 4.3: Results for the BiomeMin sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged. The observed family did not meet any of its network layers in a year in any of the tested biomes.

Biome	Times other families were met	Number of distinct people met	Days to meet 15 people	Days to meet 50 people	Days to meet 150 people	Days to meet 500 people	Times most frequently met family was met
Arctic	115.5	280.2	15	52	175	Not met	6.5
Australia	90.9	274.9	14	48	166	Not met	6.7
North America: California and Northern Mexico	324.0	619.2	5	17	54	262	9.9
North America: desert and desert scrub	168.7	292.8	11	38	143	Not met	8.3
Subarctic and continental midlatitude forest in North America and Asia	69.6	174.4	25	91	306	Not met	5.7
Tropical and subtropical Africa	777.9	377.5	3	7	26	Not met	19.1
Tropical and subtropical Asia	578.9	599.7	4	12	41	236	12.5
Tropical and subtropical South America	329.1	711.6	6	16	55	234	8.5

Table 4.4: Results for the BiomeMid sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; the number of days taken for family Y to meet its 15-, 50-, 150-, and 500-individual layer; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged.

Biome	Times other families were met	Number of distinct people met	Days to meet 15 people	Days to meet 50 people	Days to meet 150 people	Days to meet 500 people	Times most frequently met family was met
Arctic	10,021.0	6,957.7	2	2	3	6	20.4
Australia	190,160.0	4,517.1	1	1	2	2	279.4
North America: California and Northern Mexico	111,450.0	2,815.3	2	2	2	2	206.5
North America: desert and desert scrub	100,590.0	3,175.2	2	2	2	2	156.6
Subarctic and continental midlatitude forests of North America and Asia	10,668.0	4,687.8	2	2	2	6	25.7
Tropical and subtropical Africa	7,500.5	756.0	2	2	3	10	60.0
Tropical and subtropical Asia	53,634.0	2,236.5	2	2	2	2	106.8
Tropical and subtropical South America	234,720.0	7,735.7	1	2	2	2	156.1

Table 4.5: Results for the BiomeMax sets of parameter values (see text, or as explained in Appendix G). Average of 100 simulations. Shown are: the total number of times another family was met in a year, regardless of having met before or not; the total number of new people met by the end of the year, given by the number of distinct families met times the average size of the families in that biome; the number of days taken for family Y to meet its 15-, 50-, 150-, and 500-individual layer; and the number of times the most frequently met family was met. For this last measurement, families in each simulation were sorted from most met to least, and then all simulations were averaged.

correspond to the 15-layer foraging party, as can be seen in Table 4.3. This means 150 individuals were not met in very harsh conditions. In contrast, in the BiomeMax cases almost all of the biomes reached the 1500-layer, the ethnolinguistic tribe, as can be seen in Table 4.5, except for tropical and subtropical Africa, which did not have enough families in the domain for this layer. Still, in the BiomeMax cases all of the cases met more than the 500 individuals of the endogamous mating pool. Moreover, people in 5 of the 8 biomes met everyone in their domain within the first month for the BiomeMax cases.

Picking one aleatory biome to show as an example, the results for the ‘California and Northern Mexico’ biome can be seen in Figure 4.10. For this biome, family *Y* got to meet a little fewer than 170 different families in a year, which corresponds to 850 individuals. Emerging properties can be seen in Figure 4.10b, where small groups of families seem to be met roughly the same amount of times, forming “steps” in the graph. These “steps” or “oscillations” could relate to the discrete property of the number of families being met, although it could also relate to the ‘layering’ of social networks, making the network layers more of a mathematical inevitability than a social decision. Figure 4.10b shows the most met family was met 11 times by family *Y*.

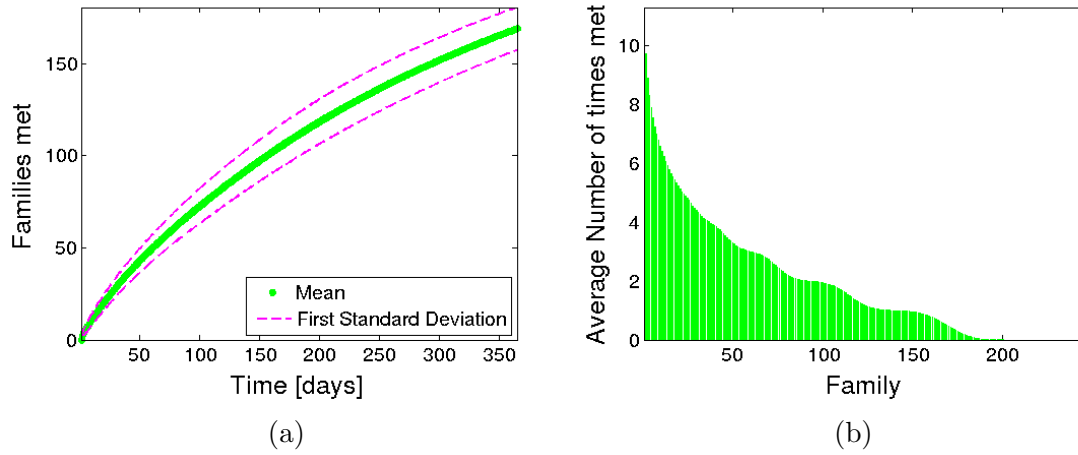


Figure 4.10: Results for the two-dimensional IMM for data from hunter-gatherers in California and Northern Mexico. (a) The number of families met by Y as they were being met throughout the year. (b) The number of times each of the other families was met by family Y on average. BiomeMid parameter values were used, with a domain of 6006 km^2 with 245 “families” of 5 individuals each, an encounter distance of 1 km, and a diffusion coefficient of $9 \text{ km}^2/\text{day}$. Simulation of a year, repeated 2000 times.

For the number of times each family was met, we find that the oscillations observed in Figure 4.10b emerged in all of the biomes for the BiomeMid and BiomeMax cases, as can be seen in Appendix H. Tropical and subtropical Asia and the North American desert and desert scrub were the biomes that showed these steps the least. The number of meetings for the family most met, however, ranges from 0, for the tropical and subtropical South America biome in the BiomeMin case, to 279.4 in Australia for the BiomeMax case. This last number is a meeting every 1.3 days, very close to the limit of a meeting every day imposed by the memory function.

In Figures 4.11 and 4.12 we exhibit, as an example, the results for a biome with a high absolute latitude: the Arctic.

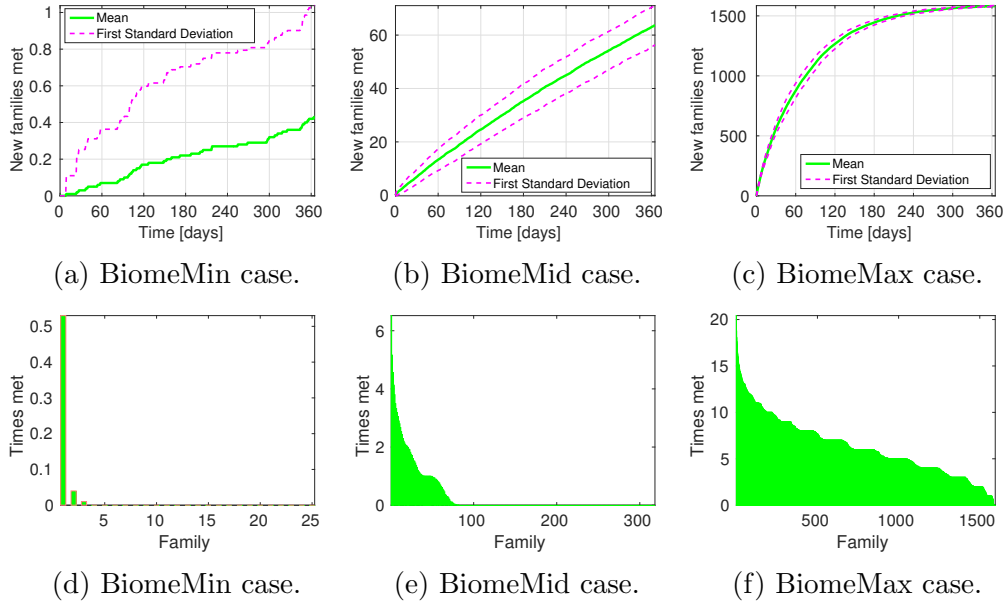


Figure 4.11: Number of new families met over the year and number of times each family was met for the BiomeMin, BiomeMid, and BiomeMax cases for the Arctic biome. This is the case for the two-dimensional model with an average of 4.4 people per family and an encounter distance of 1 km. For the BiomeMin case we had 25 initially unmet families in an area of 370,881 km², and a diffusion coefficient of 9 km²/day, with 100 simulations. For the BiomeMid case we had 318 initially unmet families in an area of 81,225 km², and a diffusion coefficient of 25.8 km²/day, with 100 simulations. For the BiomeMax case we had 1,595 initially unmet families in an area of 18,769 km², and a diffusion coefficient of 132.3 km²/day, with 26 simulations. All parameter values obtained as explained in Appendix G.

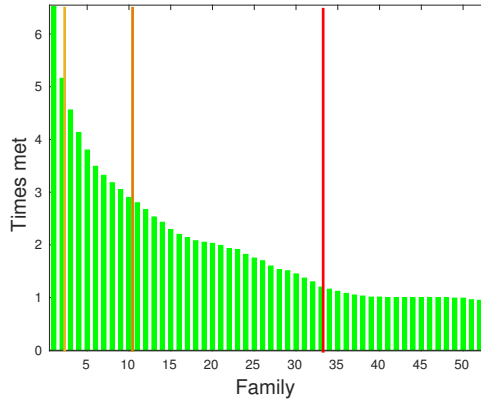


Figure 4.12: Close up of Figure 4.11e. The orange vertical line (furthest to the left) indicates when the 15-individual layer is met, the brown vertical line (middle one) for the 50-individual layer, and the red vertical line (furthest to the right) for the 150-individual layer.

The Arctic biome, having an absolute latitude range between 55.32° and 77.5° , and therefore being amongst the furthest away from the Equator and which could be related to harsh weather conditions, has some of the slowest meetings amongst biomes. We can see in Figure 4.11a its BiomeMin case barely reaches 0.4 family meetings in a year, which corresponds to 2.5 people outside their family. In Figure 4.11c we see the other extreme, where Y reaches 1,581 family meetings in a year, which is 6,958 hunter-gatherers, and the meetings are fast, reaching the 150-individual layer by the third day. But these are both extreme conditions. The BiomeMid case, seen in Figure 4.11b, was more reasonable in the sense that family Y met around 280 other hunter-gatherers (63.69 other families), taking 15 days to meet the foraging party, 52 days to meet the residential band, and 175 days to meet the active network. On the other hand, we can see in Figure 4.11e that the family most met in the BiomeMid case was met a little more than every two months, whilst in the BiomeMax case, in Figure 4.11f, the family most met was met 20 times in a year, almost twice per month. The emerging property, seen as oscillations, can be appreciated in Figures 4.11e and 4.11f, where Figure 4.11e seems to have three clear levels, or layers better shown in Figure 4.12.

Whereas Table 4.4 tells us that it is probable that all biomes meet the 150-people target, it is clear from Tables 4.3 and 4.5 that the lower and upper boundaries are too large, to the point that each biome is indistinguishable from each other. We do find, for example, that harsher climates, like the Arctic or Subarctic biomes, would take longer to meet their societal layers than tropical areas, but all of them have a lower bound of “never meeting 150 people” to an upper bound of “meeting 150

people in 2 or 3 days”. This does not seem precise enough to distinguish between biomes. We can see that the BiomeMin and BiomeMax produce clearly unrealistic results, and while the BiomeMid seems more plausible, it is difficult to know this with certainty. To further investigate this idea, we apply the IMM to each society and show the clear differences between them that could not have been noticed in the biome-guided simulations.

4.4.1.2 Application of IMM to society-specific data

We refer to “society” as the group of people that share a domain and that form part of the same population. This means that in Table 5.01 from Grove et al. [3] (as seen in Appendix D, Table D.1), each row corresponds to a different society, with columns giving location, absolute latitude, general biome, population area (km^2), number of times the society moved base-camp in a year, distance moved in the movement of base-camp, population, and mean size of family. Different societies in similar biomes may not be as similar as we first thought, given the large variation in parameter values seen in the last section. Thus, creating an average for each biome may be misleading for a given society within the biome. To illustrate such disparities we consider societies from the Arctic, and the tropical and subtropical South America biomes, a representation of each is shown separately below.

The parameter values used for the analysis in this section are taken from each society’s row in Grove et al. [3] instead of having to average over several rows as done in the previous section. As before, the family size was taken as 5 when not available in the data. The speed at which each family moves, however, was taken

as the total displacement in a year over the number of times the society moved in a year. When data for the total displacement in a year were not available, the speed was taken as the lower bound used by Pearce [2], 25 km/day. The domain was taken to be a square, so the square root of the total area was taken for the sides' lengths.

Society	Number of distinct people met	Times most frequently met family was met
Baffin Island Inuit	3,115.3	10.11
West Greenland	2,780.5	9.74
Mackenzie Inuit	1,895.9	11.27
Labrador Inuit	1,411.2	14.72
Siberian Eskimo	1,286.4	18.34
Ingulik Inuit	598.3	7.44
Noatak Inuit	546.7	18.53
Kobuk Inuit	496.8	17.52
Great Whale Inuit	445.1	16.84
Caribou Inuit	270.2	6.65
Polar Inuit	264.4	9.77
Nunamiut Inuit	239	26.40
Netsilik Inuit	227.2	6.52
Utkuhikhaling-miut	151.8	9.58
Avivlingmiut Inuit	124.2	11.25

Table 4.6: Number of distinct people met by family Y at the end of the year and maximum number of meetings of a single family in each society in the Arctic biome. Both averages over 100 simulations.

Results on the number of new people met by Y in a year and the number of times the most met family was met by the end of the year for each society in the Arctic biome can be seen in Table 4.6. These, and the results for the tropical and subtropical South America biome, can also be seen in Figure 4.13, compared to the results for the biomes as a whole. We plot the society named “Labrador Inuit” in Figure 4.14 and the “Caribou Inuit” in Figure 4.15. The Labrador Inuit met more

new families than the BiomeMid case of the Arctic biome in the previous section, whilst the Caribou Inuit met only a little less than the BiomeMid case, but definitely more than the BiomeMin case. In general, only 6 out of 15 societies in the Arctic met less than the BiomeMid case of the Arctic biome in Section 4.4.1.1, although not by much. Still, all societies were in between the BiomeMin and BiomeMax cases of the Arctic biome.

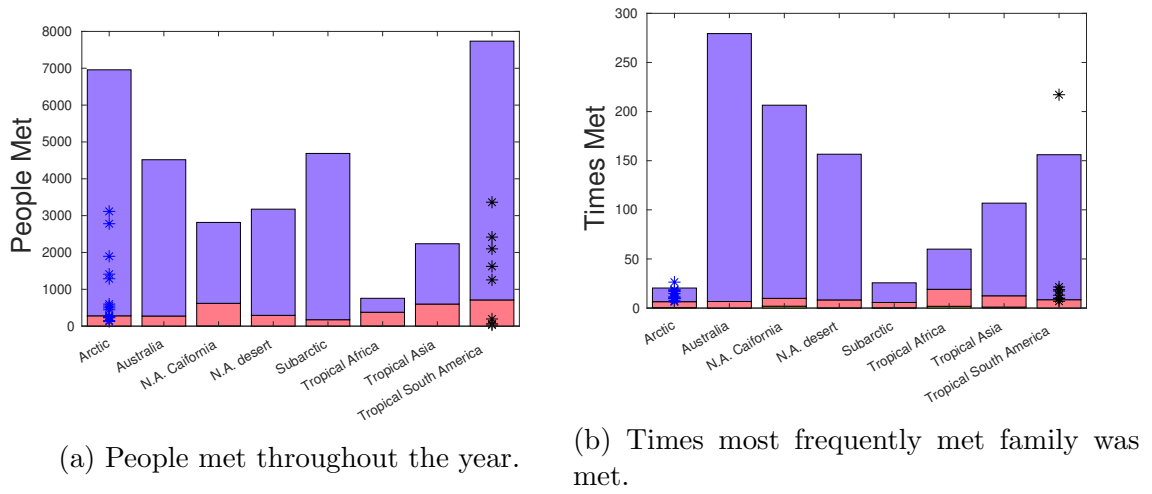


Figure 4.13: Average number of people met (left) and times most frequently met family was met (right) by each biome, for the BiomeMin, BiomeMid, and BiomeMax cases, as explained in Section 4.4.1.1. The BiomeMin case can not be seen because of how small it is compared to the BiomeMax case. BiomeMid is shown in pink (lower bars) and BiomeMax is shown in purple (taller bars). Individual results for each society in the Arctic and tropical and subtropical South America biomes are shown in their respective biomes as asterisk marks so they can be compared, although the names of the societies are not given. That is, all the data points marked as asterisks above the Arctic biome label are results for the simulation of each society in the Arctic biome individually. This makes the comparison between the approach in Section 4.4.1.1 (the coloured columns) and the approach in Section 4.4.1.2 (the asterisks) easier to see.

Regarding the families most met, most societies had between 9 and 19 meetings in a year. This means that almost all of the societies in the Arctic were between the BiomeMid and the BiomeMax cases for the complete Arctic biome in Section

4.4.1.1 (6.5 to 20.4 meetings, respectively), except for an anomalous society with 26 meetings for its most met family.

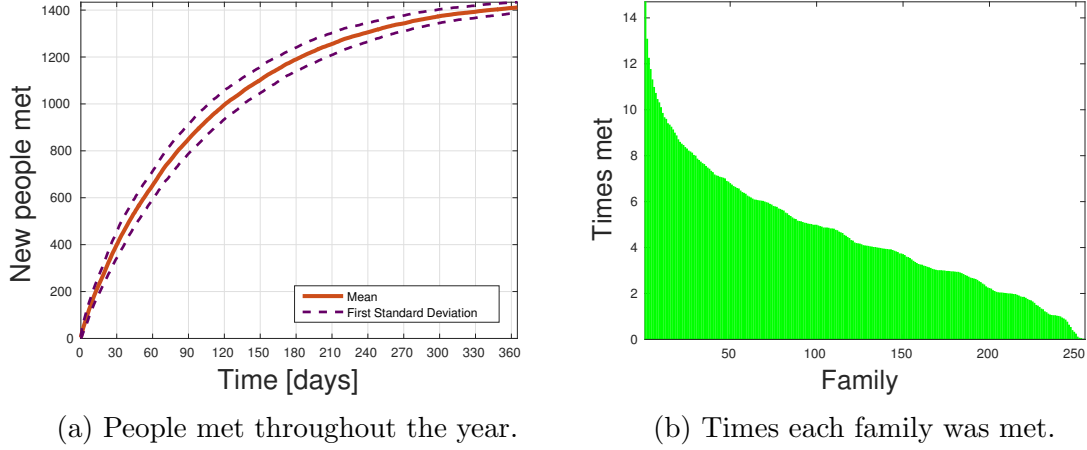


Figure 4.14: Average number of people met (left) and times each family was met (right) by the Labrador Inuit, in the Arctic, over a year, in 100 simulations. This society had 255 unmet families in a 52500 km² domain with an average family of 5.7 individuals and a speed of 33.75 km/day.

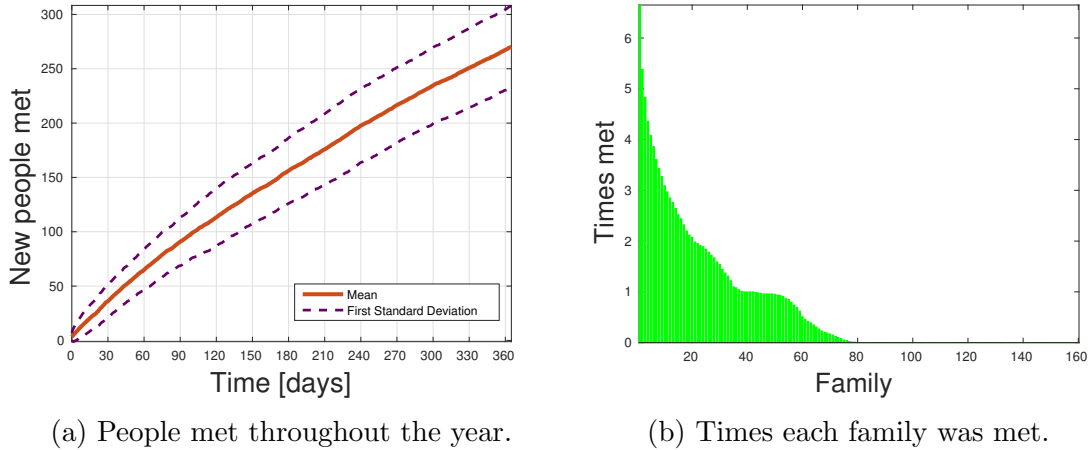


Figure 4.15: Average number of people met (left) and times each family was met (right) by the Caribou Inuit, in the Arctic, over a year, in 100 simulations. This society had 160 unmet families in a 236500 km² domain with an average family of 4.4 individuals and a speed of 27.5 km/day.

Results on the number of new people met by Y in a year and the number of times the most met family was met by the end of the year for each society in the

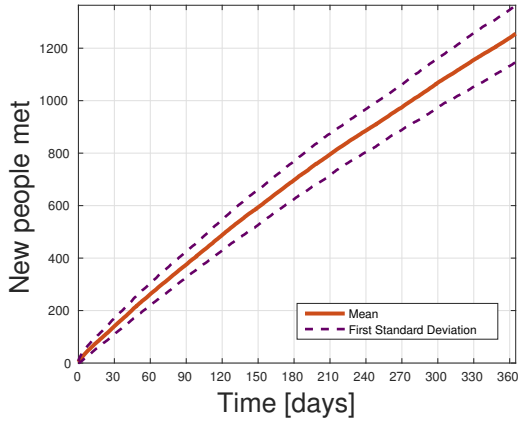
tropical and subtropical South America can be seen in Table 4.7. We choose to showcase the “Tehuelche” and “Ona” societies, as seen in Figure 4.16 and 4.17, respectively. The Tehuelche society was one of the societies that met the least number of people, not including those who did not have more than 100 people in their domain to begin with, yet it managed to meet the 500-individuals layer. The Ona are a good average example for this biome, meeting more than 1600 people. In general, for those societies who could, all met even the 1500-layer, except for the Tehuelche. Also, most societies met everyone in their domain within the year. Regarding the whole biome simulations for tropical and subtropical South America in Section 4.4.1.1, it was clear that no society would fall below the BiomeMin case, since that proved to have 0 new family meetings by the end of the year, and all were within the BiomeMax case bound.

Society	Number of distinct people met	Times most frequently met family was met
Akuriyo	63.6	13.14
Nukak	194.5	10.11
Heta	29	217.17
Tehuelche	1,255.1	6.93
Chono	2,097.1	21.49
Alacaluf	3,364	18.75
Ona	1,622.6	9.54
Yahgan	2,417.2	17.39

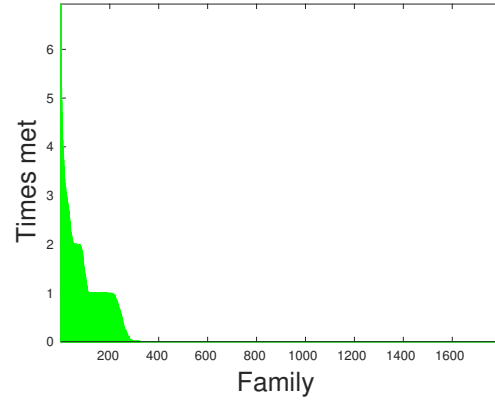
Table 4.7: Number of total new people met by family Y at the end of the year and number of times the family most met was met in each society in the tropical and subtropical South America biome.

Even though this biome was smaller, with only 8 societies, the range of number of times the most met family was met was very large, ranging between 7 and 21.5.

We have, though an anomalous case, the Heta society in south Brazil where the 5 families in the domain were met more than 200 times. What remains constant are the emergent steps, except where there were too few families for these to be clear. Once again, it was obvious that no family would be met fewer times than the number of meetings for any family in the BiomeMin case for the whole tropical and subtropical South America biome in Section 4.4.1.1 because those had 0 meetings. Still only one society had its most met family fall below the BiomeMid case in that section and only one other society had its most met family being met more times than the BiomeMax case, with 217.17 against only 156.1. The society that had its most met family having met several more times than the BiomeMax case for the same tropical and subtropical South America biome in Section 4.4.1.1 had the same domain size as the BiomeMax case, and, because theoretically the path of each family is independent of every other family, so the number of families in a domain should not influence the number of encounters and meetings each family had in this model, and both speeds were similar, both simulations should have approximately the same number of meetings for the most met family. The difference could be attributed to the low number of repetitions of the BiomeMax case, only having 19, against the 100 repetitions for the particular society.

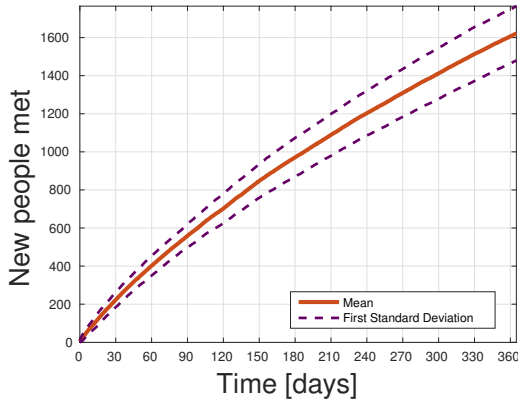


(a) People met throughout the year.

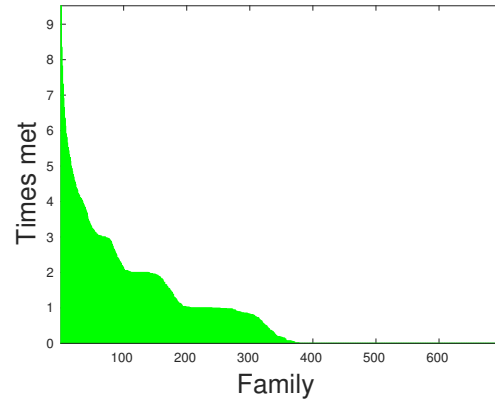


(b) Times each family was met.

Figure 4.16: Average number of people met (left) and times each family was met (right) by the Tehuelche, in Argentina, tropical and subtropical South America, over a year, in 100 simulations. This society had 1799 unmet families in a 475300 km² domain with an average family of 5 individuals and a speed of 21.18 km/day.



(a) People met throughout the year.



(b) Times each family was met.

Figure 4.17: Average number of people met (left) and times each family was met (right) by the Ona, in Argentina, tropical and subtropical South America, over a year, in 100 simulations. This society had 698 unmet families in a 48100 km² domain with an average family of 5 individuals and a speed of 13.33 km/day.

In general it seems that the BiomeMin and BiomeMax cases of Section 4.4.1.1 worked as lower and upper bounds for the results of each society in this section, as seen partly in Figure 4.13. For the cases where a society did go above the BiomeMax case, this BiomeMax as a top limit could be improved by running many

more iterations. Yet, the range given by the BiomeMin and BiomeMax cases in Section 4.4.1.1 tends to be so large that differentiation between biomes is mainly lost. Also, by dividing societies into whole biomes, information on whether societies could have met all layers of networking, and when, is lost, especially for societies that did not have a sufficient population in their domain to cover all layers. It would seem like it was useful to break the biomes into individual societies, as there is large variance within the same biomes. However, all societies seem to have met their 150-individuals layer.

The fact that in several cases everyone, or almost everyone, was met, even with such a large range of areas and variation in the number of initial families, leads to the idea that the speed chosen for these examples might have been too high. This could have been because we took the total displacement in a year over the number of times they moved base-camp, because this assumes that every time base-camp was moved, its re-establishment was completed in a day, whereas in real life this could have taken several days. This is also assuming their speed when moving base-camp is the same as when hunting or gathering. The speed parameter might have to be chosen differently, for example using velocities from different sources as we did in Section 4.4.1.1. It will also be more realistic to represent hunting and gathering groups with different speeds, which is more precisely given in Layton et al. [111], as we suggest in our further work section, Section 4.6.

In Section 4.4.2 we address doubts regarding the influence of different speeds and area parameter values by testing all possible parameter combinations and analysing which make the meetings possible and which do not.

4.4.2 Parameter Sweep of MFRM

While it is useful to examine the possibilities of societal cohesion of societies of which we know, we would like to provide information such that societies we do not yet know of can be assessed in the future. This, and the desire to try and make a rule for when are the layers met, are the stimulus for making a sweep of all realistic parameter values. For the MFRM, the parameter values needed for each simulation are: number of initial families, speed at which they diffuse, and area size. Using information provided by Prof. Robin Dunbar and Dr. Eiluned Pearce, a viable range for each of the three input parameter values was proposed. These ranges were split into what can be seen in Table 4.8 and a parameter sweep was made. This will aid in producing a quick guide on which combinations of parameter values would produce the number of desired meetings in a year. One hundred simulations were run for each combination of parameter values and the average was taken.

N_0 [families]	Corresponds to Total number of people in area	Length of one side [km]	Corresponds to Total area [km ²]	Speed [km/day]
2	15	9	81	5
9	50	99	9,801	10
29	150	199	39,601	15
99	500	299	89,401	20
150	755	399	159,201	25
299	1500	499	249,001	30
		599	358,801	35
		699	488,601	40
		799	638,401	45

Table 4.8: Ranges of parameter values used in the computer simulations for the parameter sweep.

As before, we used the assumption that humans would not live or hunt alone, and

would therefore always travel in cliques because simulating this reduces the number of individuals, whilst maintaining realism. In previous sections we have used the number of individuals per family as available in data such as Table 5.01 from Grove et al. [3] (as seen in Appendix D). From here on, to standardise, we will use exactly 5 people to a family. This reduced the need to simulate up to 1500 people in a domain to simulating only 300 families. Although it would have been interesting to test higher number of families interacting, some set of the larger parameter values used took three days or longer to run. Much higher numbers were not feasible.

With this in mind, the range for number of families in Table 4.8 was chosen so that at least enough people were available in the domain to meet the 15-people layer and all the way up to having a whole 1500-people population available, making all societal layers available through the range. The range for total area of domain was chosen such that the smallest and largest domains recorded in Binford’s data were included. Binford’s data [117] suggest that the hunter-gatherer walking speeds vary from 2 to 4 km/hour. We therefore produce a speed range, in 5 km/day intervals, from minimal walking throughout the day, to walking for 9 hours, which corresponds with some of the distances base-camps were moved if moved in one day.

We do not have information on the meeting rate of our families, μ , so we needed to find a rate that would be high enough, such that we would ensure two families falling into the same box would meet before one of them jumped out of the box again. The rate can not be too high, though, as larger values significantly slow down the simulations. We estimated that a rate bigger than that of the second highest rate would suffice. The second highest rate is the diffusion rate. From Okubo et al. [148],

we know that the two-dimensional diffusion of a species can be obtained from their walking speed by:

$$\text{diffusion} = (\text{distance walked})^2 / (4 \times \text{time taken}).$$

We know, from the mean-field equations of diffusion, that this diffusion coefficient, D , is linked to the diffusion rate, d , by:

$$D = dh^2,$$

with h the length of a box in our domain, which in this case is 1 km. We can see in Table 4.8 that the fastest walking speed taken is of 45 km/day, making our second largest rate $d_{max} = 506.25 \text{ day}^{-1}$.

We consider a family would need at least 5 minutes to recognise and appreciate the members of another family, but would need to sleep 8 hours a day, leaving 960 useful minutes in each day. We tentatively set $\mu = 960$ per day. To check if $\mu = 960$ per day is large enough we check the following: We want the probability of two families meeting within a few minutes, say 5 minutes, to be very close to 1. We assume that the meeting event is a Poisson random variable, as its occurrence is independent of each event, uniform within a time interval, and we consider two events can not happen at the same time. Hence, the probability of observing an event in a time interval t is given by:

$$P(k \text{ events in time interval } t) = (rt)^k \exp^{-rt} / k! \quad (4.51)$$

with r being the event rate. In our case the event rate is μ . If we set $\mu = 960$ meetings per day, convert μ into meetings per minute to make it consistent with the

time interval we are interested in, and make the time interval $t = 5$ minutes, then the probability of having one meeting within 5 minutes is 0.1189, and the probability of having 2 or more events in this time is of 0.8454. The probability of one or more meetings within 5 minutes will then be 0.9643, which we consider is close enough to 1. The forgetting rate will be left as $f = 1$ per day.

A quick test was made to ensure that the meeting and forgetting rates given for the computational simulations would be consistent with what would be expected in real life. These two parameter values were tested in a sole compartment (no space and no diffusion involved) of 1 km^2 with one Y family and 1000 other families. This is not a realistic density but was only used to test the functionality of the chosen rates. In this case, family Y met all other families within the first day, but only got to around 366,000 meetings by the end of the year, which means each family was forgotten more or less every day, in time to be met approximately once a day. What is more, all families had been met by the time of our first data point, at day 0.05, so within the first 1.2 hours, agreeing with our requirement: families having met after a few minutes in contact with each other.

The amount of time taken for each set of parameter values to get to 150 people met (the active network) is shown in Figure 4.18. As is logical, entries with fewer than 150 people in the domain to start with, $N_0 = 2$ and 9 families, did not reach the 150-individuals level of society, the active network. It can also be seen that as the domain gets larger, as indicated in the x -axis, it takes longer for people to meet 149 other people. It is also clear that the faster the people move, the quicker they will meet each other. Qualitatively similar graphs were obtained for meeting all

different hierarchical levels of society.

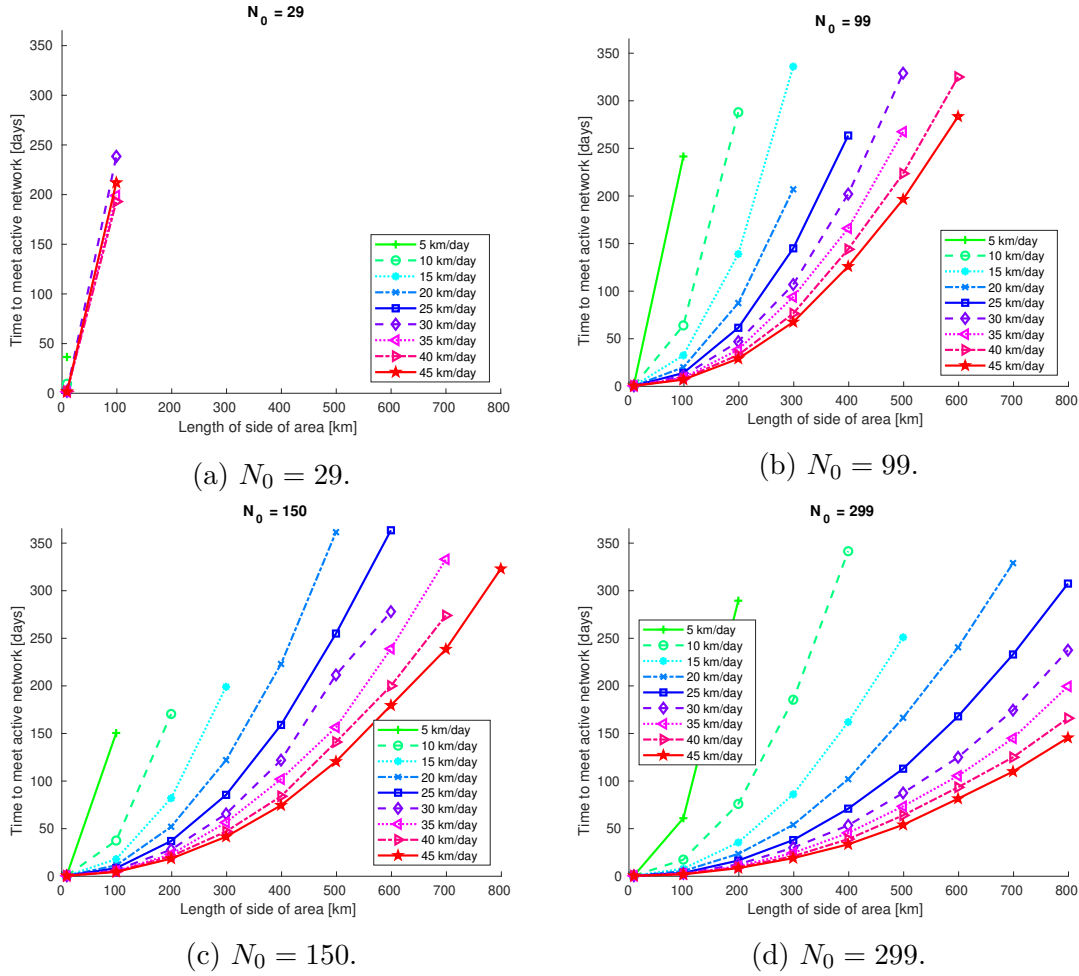


Figure 4.18: Time taken to meet 150 people, given different: areas in which families move, speeds at which they move, and the number of families in total in the area, N_0 . Because the domains were taken as squares, only the length of one side of the domain is given. It can be seen that larger length increases the time to meet 150 people, whereas the increase in people in the area (across subfigures), or the increase in speed (lower lines in each graph), both decrease the time taken to meet the hierarchical level.

By plotting these graphs in a log-log plot we realise the relation between the side length and the time taken to meet 150 people behaves like a power law. The graphs are mostly linear, except for the first point in the graph. As we output every 0.5 days, the first data point might not be an accurate representation of the actual time

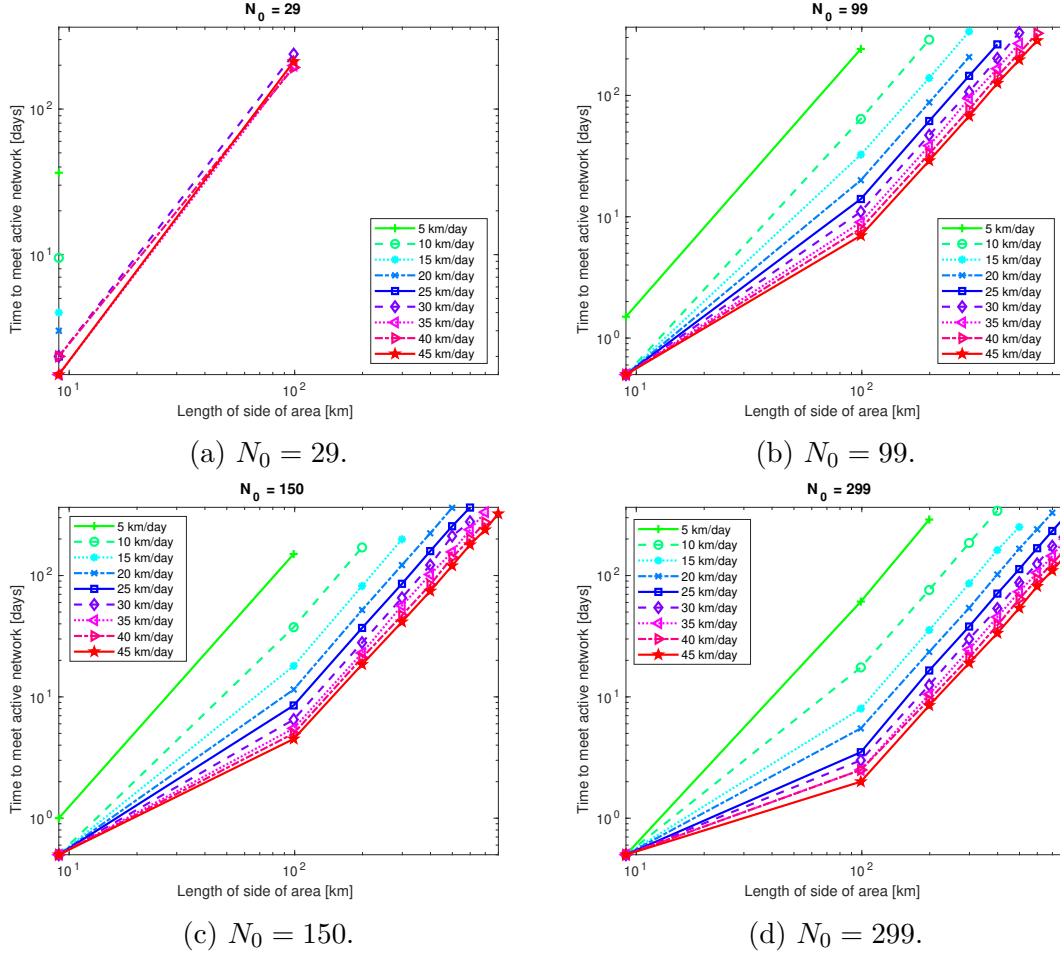


Figure 4.19: Log-log plot of Figure 4.18. Time taken to meet 150 people, given different: areas in which families move, speeds at which they move, and the number of families in total in the area, N_0 . Because the domains were taken as squares, only the length of one side of the domain is given. It can be seen that larger length increases the time to meet 150 people, whereas the increase in people in the area (across subfigures), or the increase in speed (lower lines in each graph), both decrease the time taken to meet the hierarchical level. These graphs show a power relationship.

for meeting 150 people as the meeting may have happened much earlier than day 0.5. Therefore we ignore this first point when doing linear regression. We also ignore graphs with just one data point. The gradients from the log-log plots are shown in Figure 4.20, which would correlate with the index of the power law, whereas the intercept points are shown in Figure 4.21, which would indicate the constant

term multiplying the independent variable. It can be seen that the gradients have an ascending relationship with velocity within the same parameter values, and are larger with ascending numbers of people in the domain.

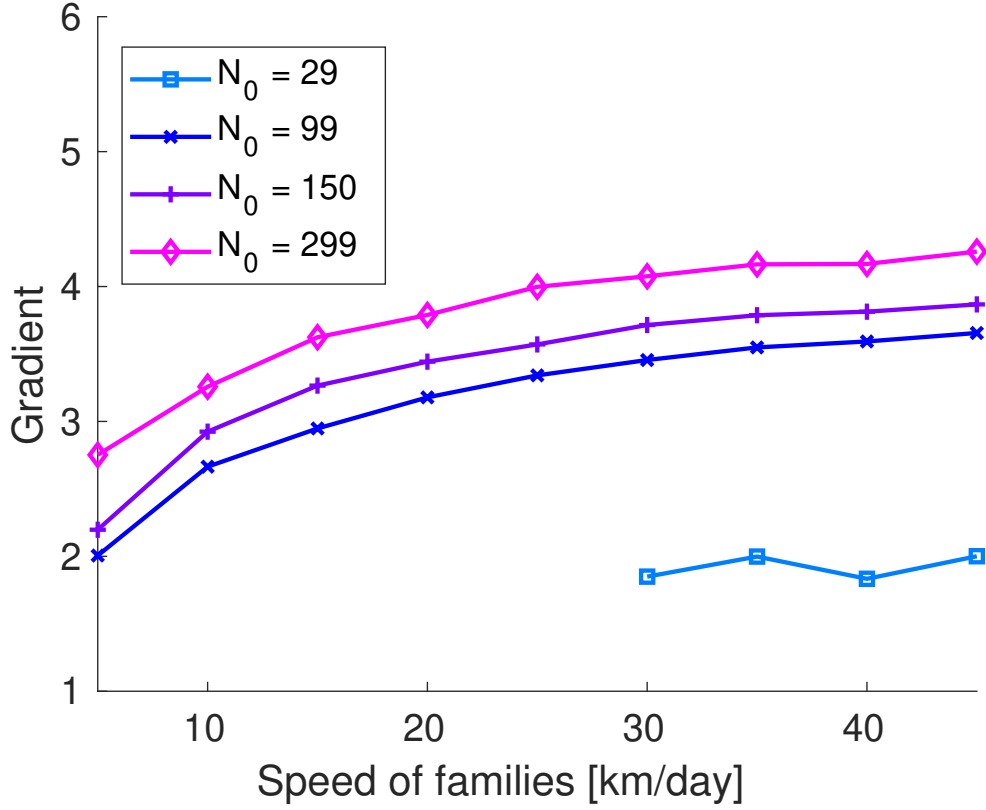


Figure 4.20: Gradients from linear regression results corresponding to Figure 4.19a (lower line in light blue with squares), Figure 4.19b (dark blue line with crosses), Figure 4.19c (purple line with plus signs), and Figure 4.19d (top line in magenta with diamond markers). The gradient is the power term of the power law found in Figure 4.19.

The same procedure was carried out for the time taken to meet other network layers (sympathy group at 15 people, affinity group at 50 people, mating pool at 500 people, and the tribe at 1500 people) and the gradients and intersection points are shown in Figures 4.22 and 4.23, respectively.

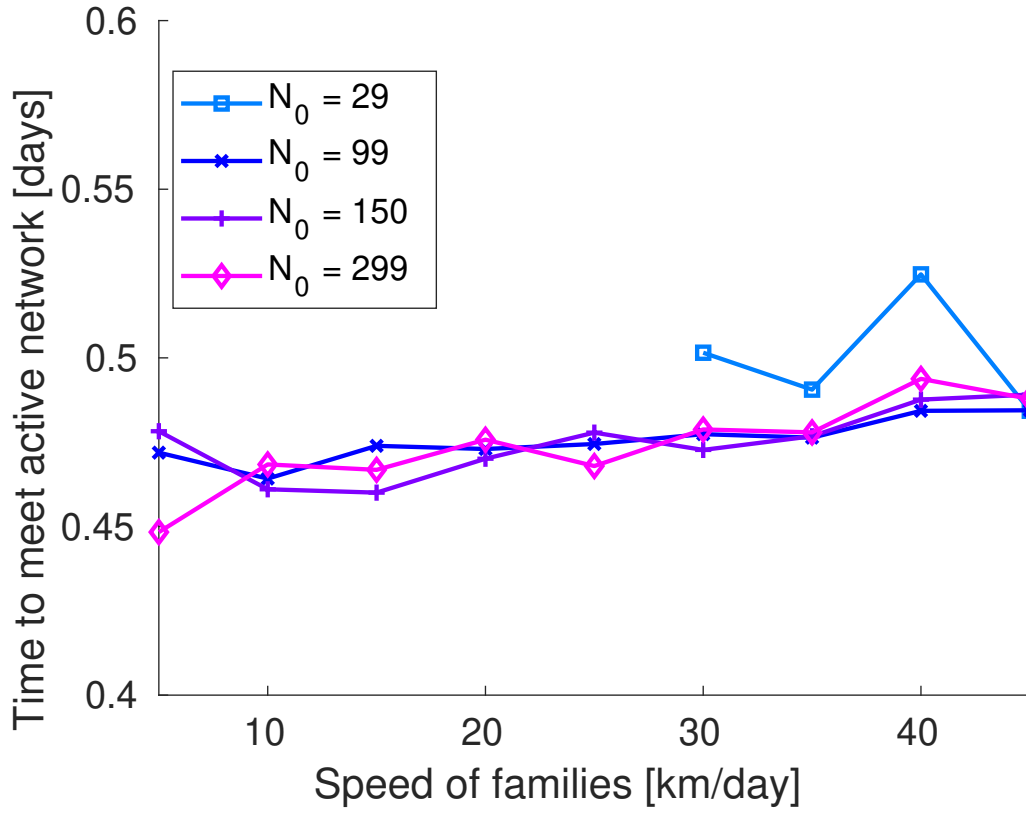
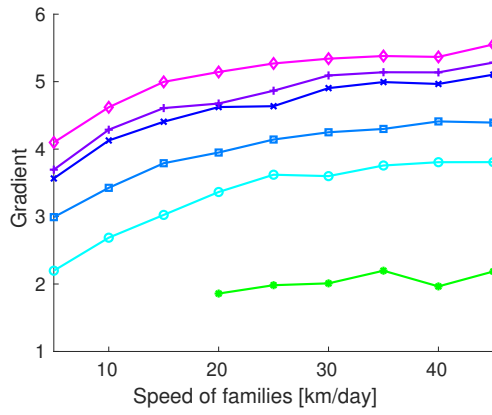
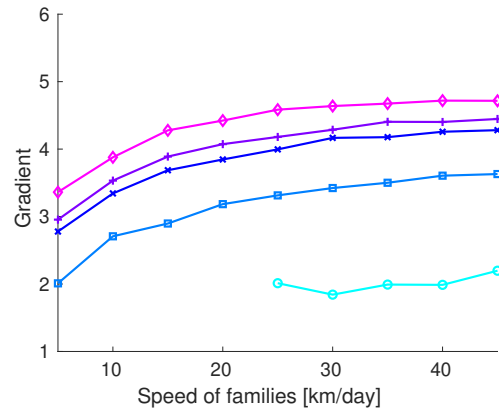


Figure 4.21: Intercept with y-axis when doing linear regression to Figure 4.19a (top line in light blue with squares), Figure 4.19b (dark blue line with crosses), Figure 4.19c (purple line with plus signs), and Figure 4.19d (magenta line with diamond markers). The intercept is the constant term multiplying the powered independent variable of the power law.

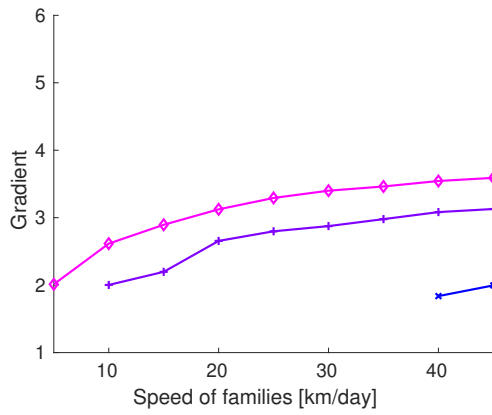
We can deduce that with an increase in density and diffusive speed, the time at which the desired society level is met decreases. But we have yet to deduce an equation that can take the parameter values and produce similar results.



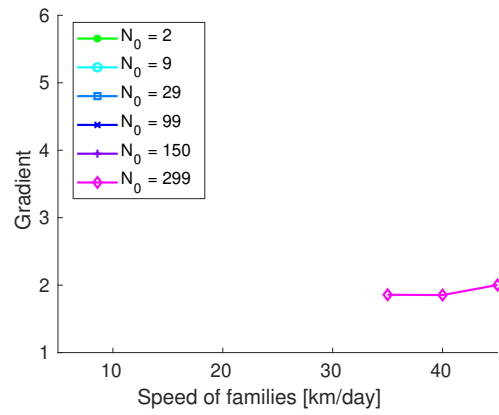
(a) Time to meet 15 people.



(b) Time to meet 50 people.



(c) Time to meet 500 people.



(d) Time to meet 1500 people.

Figure 4.22: Gradients from linear regression results similar to Figure 4.20 but corresponding to the time taken to meet the 15, 50, 500, and the 1500 social layers.

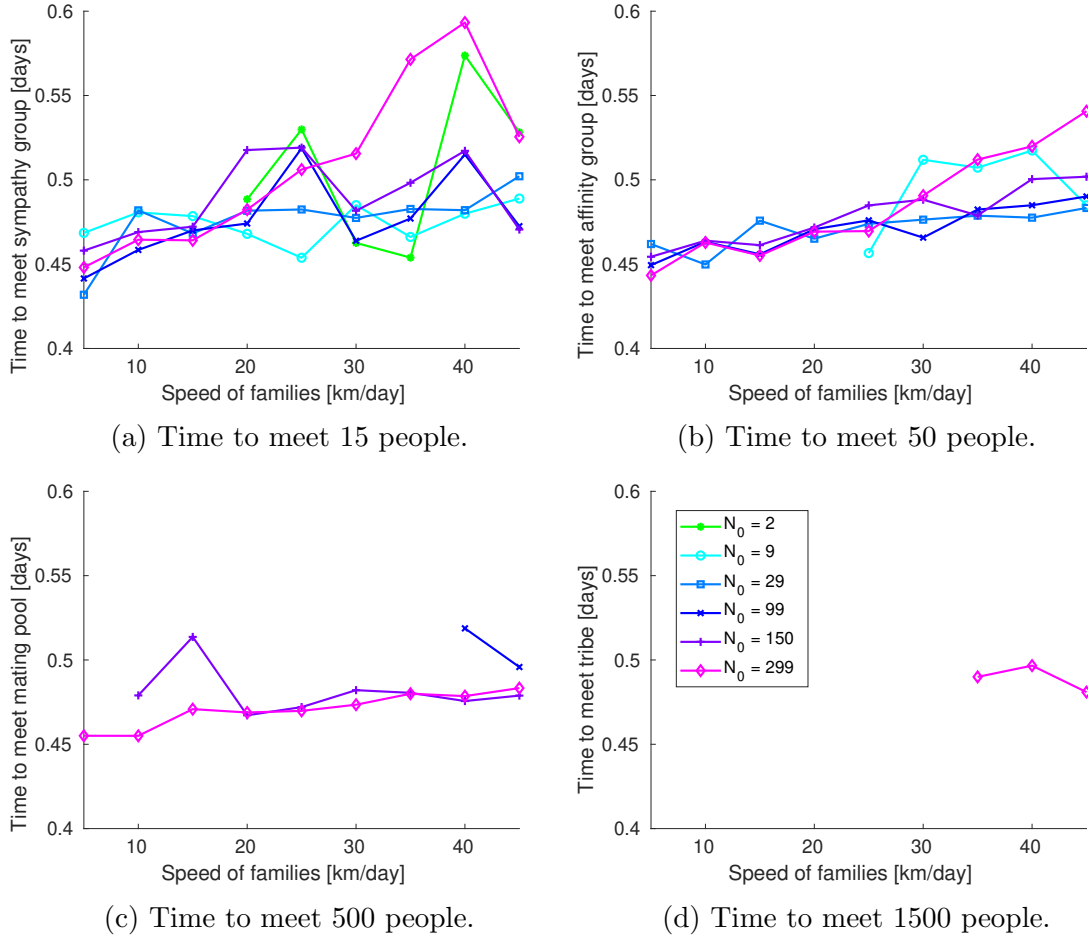


Figure 4.23: Intersect with Y-axis when carrying out linear regression similar to Figure 4.20 but corresponding to the time taken to meet the 15, 50, 500, and the 1500 social layers.

Finally, our parameter sweep produces Figure 4.24, the range of parameter values checked for the length of one side of the square domain, and the speed at which the families move per day. It also marks the cases of initial number of families for which the 150-people layers were met. Figure 4.25 indicates the same for the other societal tiers. These figures could be used as a rough look-up table for when are people meeting their network layers and when they might not. We can see that many parameter combinations do reach the 150-people layer, especially with speeds of 25

km/day or more, which seems well within human capacity. When there are more than enough people in the domain, the domain has to be very large for the people to not meet their active network. Yet it seems very difficult for them to meet their whole ethnolinguistic tribe when there are barely enough people for the 1500 layer to exist. In all, meeting the 150 people within a year by random movement seems plausible in most cases.

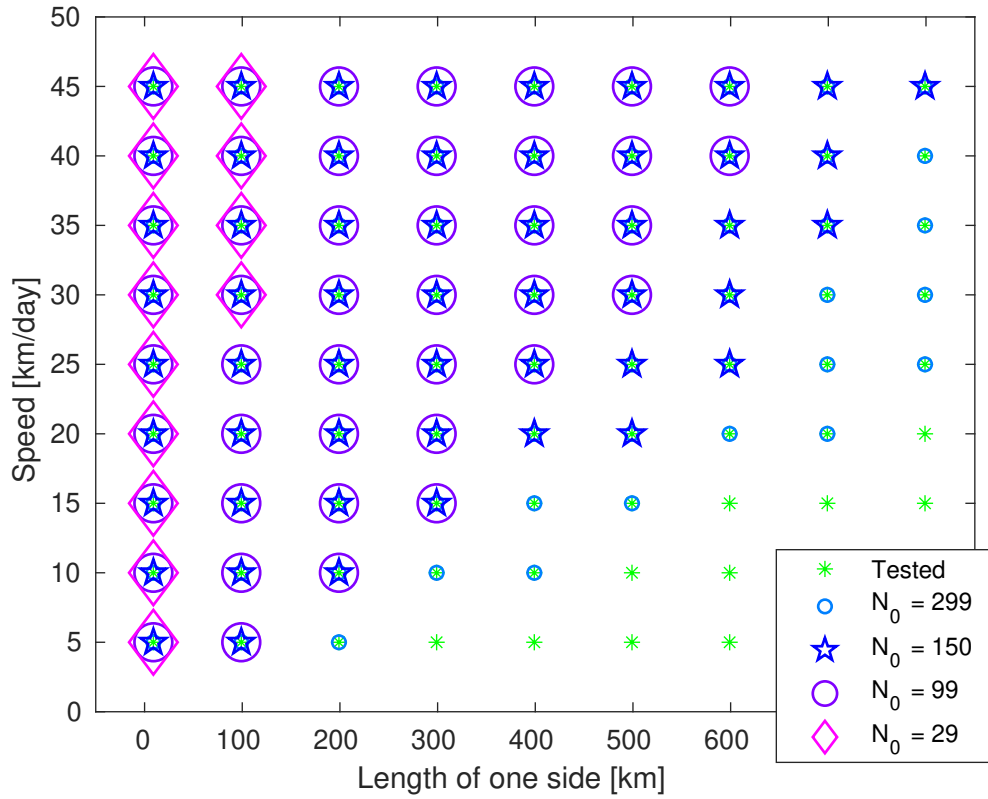
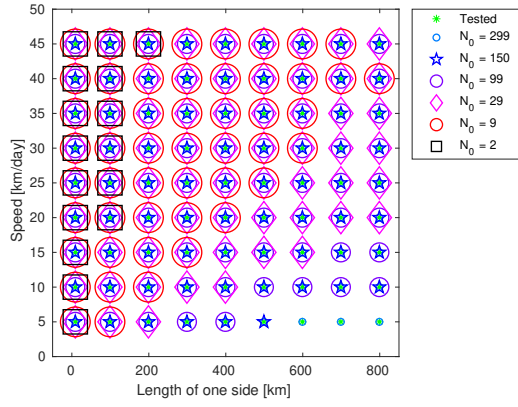
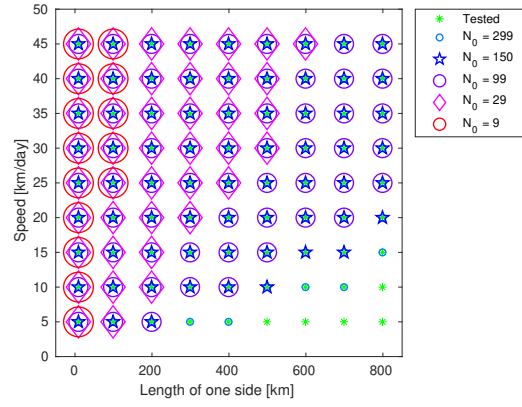


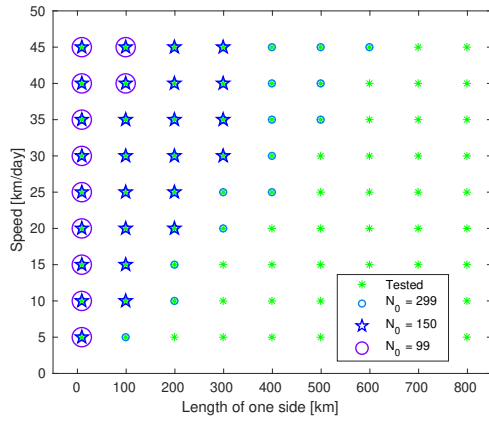
Figure 4.24: Out of the parameter values tested, those which hit the 150 level of number of people met throughout a year are marked as indicated. N_0 refers to the number of ‘other’ families in the domain. Meeting one family would mean meeting 5 people.



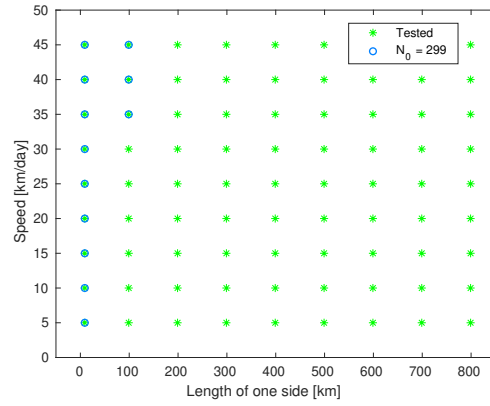
(a) Parameter values that met 15 people.



(b) Parameter values that met 50 people.



(c) Parameter values that met 500 people.

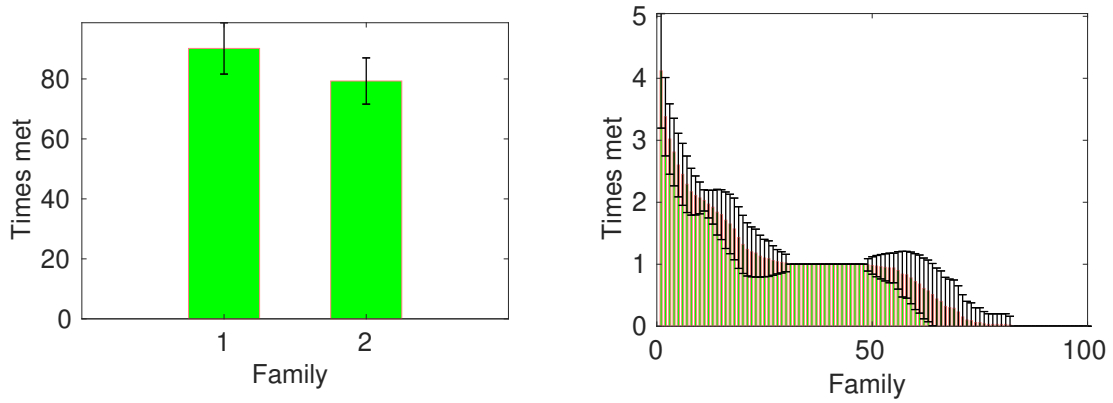


(d) Parameter values that met 1500 people.

Figure 4.25: Out of the parameter values tested, those which, within the year, hit the number of people indicated below the subfigure are marked with a symbol accordingly.

Following up on the “waves” seen before, we plot in Figure 4.26 the graphs for the ranking of families obtained from using all the minimum parameter values (Figure 4.26a) to using all the maximum parameter values (Figure 4.26b). As we analyse all the 486 results we realise that all set of parameter values for which a pattern could be seen, the waves seen in Figure 4.26b, would form.

We noticed that in Figure 4.26b, around 10 families seemed to be met exactly once during the year, without deviation, so we explored this further. We ran a par-



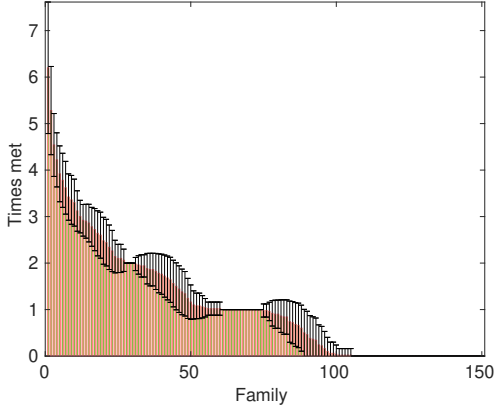
(a) Minimum value of all parameter values: $N_0 = 2$ other families, $L = 9$ km per side, and moving at 5 km per day. (b) Maximum value of all parameter values: $N_0 = 299$ other families, $L = 799$ km per side, and moving at 45 km per day.

Figure 4.26: Ranked families from most met to least. We obtain the number of times each family was met by family Y within a year and then sort them from most met to least, then average over all 100 simulations. (a) Minimum in all parameter values: 2 other families, apart from the main one, and movement at a speed of 5 km/day in a 9×9 km² domain. (b) Maximum in all parameter values: 299 other families, apart from the main one, each moving at the speed of 45 km/day in a 799×7999 km² domain. In each simulation Y is initialised in the middle of the domain and all other families are placed uniformly randomly throughout the domain.

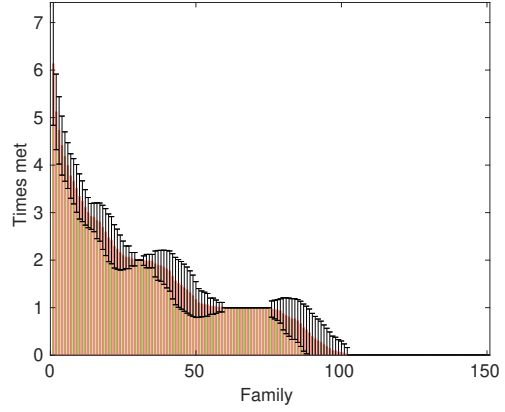
ticular set of parameter values, where the wave-like formations, or “steps”, could be seen, for varying numbers of simulations. The results can be seen in Figure 4.27. It is difficult to see in Figures 4.27a to 4.27c, but it is clear by Figure 4.27d, with one thousand simulations, that the lack of deviation is a product of an insufficient number of simulations. One thousand simulations were needed in order to see variation in the number of times all families were met in this particular case.

Nonetheless, the waves are still present, despite the hundred-fold repeated simulations. The deviation in Figure 4.26b still gets smaller periodically. We notice that the “plateau” of the waves seems to coincide with whole numbers and probably stems from the integer characteristic of this kind of measurement. This phenomenon

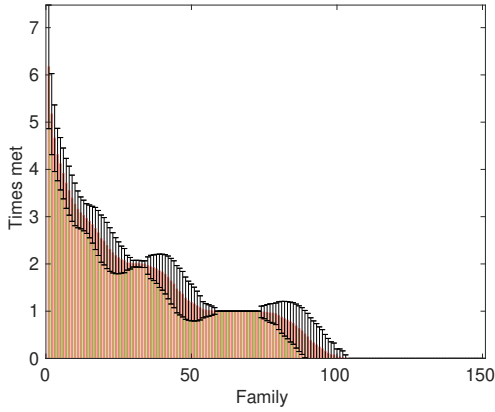
may be explained by an area of mathematics called “Order Statistics” [149, 150], but this is beyond the scope of this thesis.



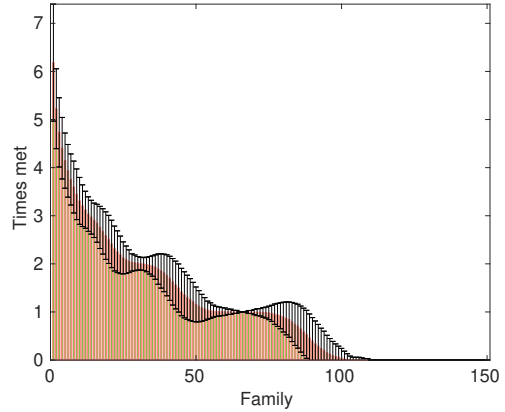
(a) 50 repetitions.



(b) 100 repetitions.



(c) 200 repetitions.



(d) 1000 repetitions.

Figure 4.27: Examining the plateau observed in the ranking of families, such as in Figure 4.26b. Varying the number of simulations for the mean of ranked families. 150 other families were used, with 399 km per side of the domain, and a speed of 40 km/day.

4.5 Conclusions

In this chapter, we set out to answer whether it was possible for hunter-gatherers to meet their network layers as a by-product of their roaming to fulfil other necessities, or if other means of bonding would be necessary. We studied simple diffusive motion,

and we developed several, progressively more complicated, models to gain insight into this problem. We then applied these models to several real-life cases based on data-driven parameter values.

At the beginning we thought that when applying the model to real-life cases we could divide the societies into the biomes to which they belonged, obtaining upper and lower bounds of number of people met for those biomes. We hoped this could be applied to societies for which we had no information but were in similar biomes. However, we realised that dividing the societies into biomes was unnecessary and gave bounds for the possible number of meetings for families in each biome that were too large to distinguish between societies. We also realised that the absolute number of meetings one family had throughout the year can be approximated by the ideal gas model, as it is linear, but not so the distinct number of people met.

In Section 4.3 we showed that a mean field approach is not valid for the meeting process in which we are interested because of the importance of stochasticity. Critically, we are dealing with only one family and its interactions with the other families in the tribe.

Finally, the parameter sweep helped us identify parameter spaces in which meeting 150 people would be plausible, allowing us to generate look-up tables (Figures 4.24 and 4.25). These tables mean that someone wanting to know whether a society could be meeting its different network layers in a year would only need to know the size of the area, the number of people in that society, and the speed at which they roam, to answer their question. In the analysis of these tables we showed that there is a power-law relation between the number of meetings in a year and the length of

the side of the domain.

Section 4.2.1 provides us with an analytical solution to the probability of a particle being at a given place, given our particular case of discrete space and continuous time diffusion with reflective boundary conditions, Equation (4.24), which is an important result of our work. This could be used in the future to search for an analytical solution for the number of encounters a family has in a year.

4.6 Further work

The work presented in this chapter could take several different paths going forward. Some of the things that can be done to carry the work forward are fairly easy and would just take time, like extending the parameter sweep in Section 4.4.2 to account for more entities moving. Some other things require a lot more work and depend deeply on the specific questions asked, like exploring groups with different tasks, or the source of the wave-like phenomena seen in Figure 4.27. Here we present some ideas we have explored in a preliminary fashion.

As mentioned before, by grouping people into families of 5 we found a region of parameter space that allowed people to meet 29 other families, namely 150 people in total. To fully address the problem of each person moving individually and covering all the possible combinations of parameter values would require a significant amount of computational time. The coding needed for this is already available, but the simulations would take several months.

Another step that would be simple to take would be to find an *ad hoc* equation that could predict whether someone in a particular society could meet its 150-people active network. The work done to obtain Figures 4.20 and 4.21 already lets us see the relation that exists between the three parameter values used in our model. An equation could be sought for the meeting of each network layer, which could be quicker and easier to use than the look-up tables in Figures 4.24 and 4.25. As it would be *ad hoc*, this endeavour may prove to be quick or may encounter some other problems on the way.

One of the main reasons for changing to the MFRM was the possibility to translate the computational model into equations that may be quicker to solve and might be more insightful. Section 4.3 showed that the Mean Field approach would not suffice, but we could use the tools from Chapters 2 and 3 and derive the equations from first principles. The multi-centre particle density formalism has the ability to take into account the relative distance between particles, and so can incorporate the stochasticity that proves to be important in the encounter between hunter-gatherers.

Since animals have memory and take sensory cues from their environment, a purely random movement will not be realistic when looking for specific details about their displacement and decision-making. In general, we would have to incorporate movement and interaction with the environment, to produce a more realistic hunter-gatherer movement model. To do this we would implement the idea of resource use. The following section, Section 4.6.1, shows our preliminary advances on studying how the diffusive movement of humans and their meetings are affected by spatially-distributed resources.

4.6.1 Mathematical model of resource use and meeting

An animal's pattern of movement and home range size greatly influences its survival. Although movement and environment are individually actively studied, the coupling of environmental factors with population dynamics has received little attention [151] and its better understanding is needed in order to bring our models closer to empirical observations. For example, animals may be restricted by patches of land where food does not grow, or attracted to bodies of fresh water. A model

by Possingham and Houston [152] shows that, for territorial foragers, food renewal is an important element of foraging strategy.

Building on the purely diffusive model from Section 4.2.1, and with the assumption that the hunter-gatherers move mainly following resources, we will add a resource density, R . In this section we first get an idea of how the motion of one type of family is affected by the presence of resource being followed in the domain. Alongside this, we also obtain a set of PDEs for the time evolution of the one type of family, Y , and the resource field, R . We then combine the meeting action that we developed in Section 4.2 with this new resource use and start studying how is the distribution of families Y , U and M affected by this combination. As a first example we consider a case with high eating rate which shows how the number of new meetings in a year can be highly impaired when the resource field is quickly depleted.

In general, this section suggests that if the rate between resource regrowth and resource usage is not large enough, it is very easy for the families to quickly consume available resources, get trapped, and not meet enough people. We can predict that if the amount of available food is infinite, this case should go back to the case where resource had no inference, like in Section 4.2, so these two extremes should mark the boundaries for the amount of people that can be met in a year by moving to collect resources.

4.6.1.1 Preliminary investigation into a model including resource use

Firstly, we explore the effects of resource use on the movement and distribution of one type of family. We may see this resource as food, which will be eaten by the hunter-gatherers with an eating rate E and will grow back with a growing rate G , with the maximum level of available resources being R_{max} . Initially, we start with a homogeneous source of food at R_{max} everywhere in a one-dimensional domain.

In this case, human movement will not just depend on diffusion, but there will be a transition function, $T^\pm \Delta t$ which, in turn, depends on the resource density. This transition function will dictate the probability that a family will move left (T^-) or right (T^+) in the time interval Δt . We use a discretized space and time, where time is divided into intervals of size Δt , and t_j refers to an interval $[(j-1)\Delta t, j\Delta t)$, and a space domain $[0, L]$ which is divided in compartments of length Δx and x_k refers to the compartment $[(k-1)\Delta x, k\Delta x)$. For the probability that a particle will be in compartment x_k at time t_j , $Y(x_k, t_j)$, we have:

$$Y(x_k, t_j + \Delta t) = T_k^+ \Delta t Y(x_k - \Delta x, t_j) + T_k^- \Delta t Y(x_k + \Delta x, t_j) + (1 - T_{k+1}^+ \Delta t - T_{k-1}^- \Delta t) Y(x_k, t_j),$$

where T_k^- is the transition left into compartment x_k , and T_k^+ is the transition right into compartment x_k . We define

$$T_k^\pm = dR_k, \tag{4.52}$$

where R_k will be the amount of resource available at x_k , and d is the diffusion rate constant.

Using the Taylor expansion for $Y(x_k, t_j)$ around x :

$$\begin{aligned}
Y(x_k, t_j + \Delta t) &= dR_k \Delta t Y(x_k - \Delta x, t_j) + dR_k \Delta t Y(x_k + \Delta x, t_j) \\
&\quad + (1 - dR_{k+1} \Delta t - dR_{k-1} \Delta t) Y(x_k, t_j), \\
\frac{Y(x_k, t_j + \Delta t) - Y(x_k, t_j)}{\Delta t} &= d[R_k Y(x_k - \Delta x, t_j) + R_k Y(x_k + \Delta x, t_j) \\
&\quad - (R_{k+1} + R_{k-1}) Y(x_k, t_j)], \\
&= d \left[2R_k Y(x_k, t_j) + R_k \frac{\partial^2 Y(x_k, t_j)}{\partial x^2} (\Delta x)^2 + \mathcal{O}(\Delta x^3) \right. \\
&\quad \left. - (R_{k+1} + R_{k-1}) Y(x_k, t_j) \right].
\end{aligned}$$

Assuming $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$ with d such that $d\Delta x^2 \rightarrow D$, D a positive constant, then $\mathcal{O}(\Delta x^3)$ will be negligible compared to the remaining terms and, letting $d = D/\Delta x^2$:

$$\frac{\partial Y(x, t)}{\partial t} = D \left[R(x, t) \frac{\partial^2 Y(x, t)}{\partial x^2} - \frac{\partial^2 R(x, t)}{\partial x^2} Y(x, t) \right], \quad (4.53)$$

where x and t are now continuous, as is $R(x, t)$, and $Y(x, t)$ is the density of a particle in x at any time point t . This simplifies to:

$$\frac{\partial Y(x, t)}{\partial t} = D \frac{\partial}{\partial x} \left[R(x, t) \frac{\partial Y(x, t)}{\partial x} - \frac{\partial R(x, t)}{\partial x} Y(x, t) \right]. \quad (4.54)$$

We now develop the form of the resource equation, $R(x, t)$, which is a function that gives us the amount of resource at time t in a compartment around x of width Δx . We assume there is a maximum level of food, therefore $R(x, t) \in [0, R_{max}]$. Intuitively, the concentration of $R(x, t)$ will decrease with the eating rate, E , times the amount of people in x , and grow with rate G , such that:

$$\frac{\partial R(x, t)}{\partial t} = -EY(x, t)R(x, t) + G(R_{max} - R(x, t)). \quad (4.55)$$

We can non-dimensionalize Equations (4.54) and (4.55) with:

$$R = R_{max}r, \quad Y = \frac{G}{E}y, \quad t = \frac{1}{G}t', \quad x = \sqrt{\frac{DR_{max}}{G}}x',$$

where r , y , t' , and x' are now dimensionless. So that, upon dropping primes for simplicity:

$$\frac{\partial y(x, t)}{\partial t} = \frac{\partial}{\partial x} \left[r(x, t) \frac{\partial y(x, t)}{\partial x} - y(x, t) \frac{\partial r(x, t)}{\partial x} \right], \quad (4.56)$$

$$\frac{\partial r(x, t)}{\partial t} = -y(x, t)r(x, t) + 1 - r(x, t). \quad (4.57)$$

As initial conditions we have:

$$r(x, t = 0) = 1, \\ y(x, t = 0) = \begin{cases} \frac{E}{G2\delta L} & \text{if } x \in [L/2 - \delta L, L/2 + \delta L] \\ 0 & \text{elsewhere} \end{cases},$$

where L is the length of the domain, and δL is the size of the partitions of L . This means the resource starts at its maximum capacity, and the particles are initiated at the middle of the domain. Also, at the boundary, we have reflective boundary conditions:

$$\begin{aligned} \left. \frac{\partial y(x, t)}{\partial x} \right|_{x=0} &= 0, \\ \left. \frac{\partial y(x, t)}{\partial x} \right|_{x=L} &= 0, \\ \left. \frac{\partial r(x, t)}{\partial x} \right|_{x=0} &= 0, \\ \left. \frac{\partial r(x, t)}{\partial x} \right|_{x=L} &= 0. \end{aligned}$$

A numerical solution can be seen in Figure 4.28. The difference with the simple diffusion case (an example of which can be seen in Figure 4.7, from Equation (4.40))

can be readily observed in Figure 4.29, where the densities of y and r are shown for time $t = 0.01$ days.

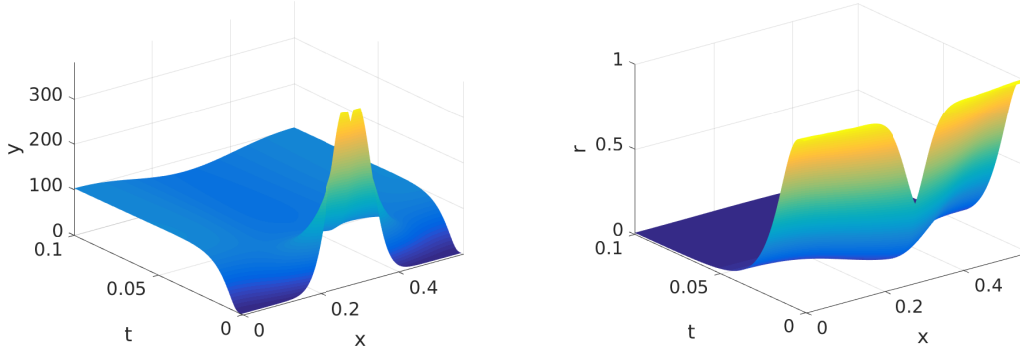


Figure 4.28: Numerical solution to Equations (4.56) (left) and (4.57) (right). Hunter-gatherers diffusing and eating in a one-dimensional space of length 200 km, diffusing at $25 \text{ km}^2/\text{day}$, with an eating rate of 50 resources per day, a growth rate of 1 per day, an initial maximum food level of $R_{max} = 5000$ units in every compartment and y initially as a step function of area 1 in the centre of the domain.

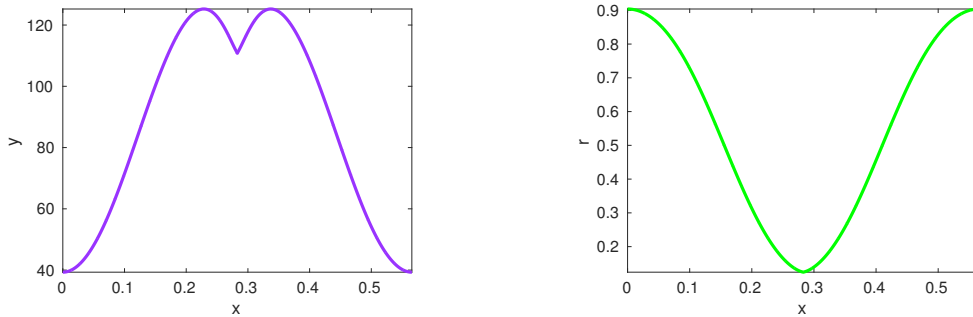


Figure 4.29: Numerical solution to Equations (4.56) (left) and (4.57) (right) at time $t = 0.01$ days. Same parameter values as in Figure 4.28.

We can see in Figure 4.29 that, because of the initial depletion of food in the centre of the domain due to the initial conditions on y , y falls in the middle of the domain. As time goes by, these “indentations” move sideways throughout the domain, but the shape of the y curve, in general, quickly tends to a uniform distribution, so that these indentations become smaller and are unnoticeable by the end

of the first unit of time. Something similar happens to r as it mirrors y 's movement, depleting faster wherever y is more concentrated and experiencing some lag time before “regrowing”.

Since a numerical solution is much faster than a stochastic simulation, although not as exact, solving this set of equations should give us a good estimate of the parameter range to be used in the stochastic simulations to obtain the types of behaviour in which we are interested.

4.6.1.2 Combining meeting and resource use

In this section we combine the MRM with the model of resources to gain some knowledge of how the search for resources affects the number of meetings. An example with people eating too fast, or too little food available, can then be seen.

The MRM without forgetting is similar to Equations (4.40) for Y , (4.41) for U , and (4.42) for M , but the incorporation of the resource usage is analogous to Equation (4.54) for Y and Equation (4.55) for R , such that the PDEs for the deterministic mean-field approximation of this model are:

$$\frac{\partial Y(x, t)}{\partial t} = D \frac{\partial}{\partial x} \left[R(x, t) \frac{\partial Y(x, t)}{\partial x} - \frac{\partial R(x, t)}{\partial x} Y(x, t) \right], \quad (4.58)$$

$$\frac{\partial U(x, t)}{\partial t} = D \frac{\partial}{\partial x} \left[R(x, t) \frac{\partial U(x, t)}{\partial x} - \frac{\partial R(x, t)}{\partial x} U(x, t) \right] - \mu Y(x, t) U(x, t), \quad (4.59)$$

$$\frac{\partial M(x, t)}{\partial t} = D \frac{\partial}{\partial x} \left[R(x, t) \frac{\partial M(x, t)}{\partial x} - \frac{\partial R(x, t)}{\partial x} M(x, t) \right] + \mu Y(x, t) U(x, t), \quad (4.60)$$

$$\frac{\partial R(x, t)}{\partial t} = -E[Y(x, t) + U(x, t) + M(x, t)]R(x, t) + G(R_{max} - R(x, t)), \quad (4.61)$$

where $Y(x, t)$, $U(x, t)$, $M(x, t)$, and $R(x, t)$ are the density of the main family, the unmet families, the met families, and food units in x at any time t , respectively. As previously, μ is the meeting rate, E is the eating rate, and G is the growth rate.

As before, we can non-dimensionalize the set of Equation (4.58) to (4.61) by setting new variables:

$$r = \frac{R}{R_{max}}, \quad y = \frac{E}{G}Y, \quad u = \frac{E}{G}U, \quad m = \frac{E}{G}M, \quad t' = Gt, \quad x' = \sqrt{\frac{G}{DR_{max}}}x,$$

where the variables on the left are now dimensionless. So that upon dropping primes for simplicity:

$$\frac{\partial y(x, t)}{\partial t} = \frac{\partial}{\partial x} \left[r(x, t) \frac{\partial y(x, t)}{\partial x} - \frac{\partial r(x, t)}{\partial x} y(x, t) \right], \quad (4.62)$$

$$\frac{\partial u(x, t)}{\partial t} = \frac{\partial}{\partial x} \left[r(x, t) \frac{\partial u(x, t)}{\partial x} - \frac{\partial r(x, t)}{\partial x} u(x, t) \right] - \frac{\mu}{E} y(x, t) u(x, t), \quad (4.63)$$

$$\frac{\partial m(x, t)}{\partial t} = \frac{\partial}{\partial x} \left[r(x, t) \frac{\partial m(x, t)}{\partial x} - \frac{\partial r(x, t)}{\partial x} m(x, t) \right] + \frac{\mu}{E} y(x, t) u(x, t), \quad (4.64)$$

$$\frac{\partial r(x, t)}{\partial t} = -[y(x, t) + u(x, t) + m(x, t) + 1]r(x, t) + 1. \quad (4.65)$$

As initial conditions we have:

$$y(x, t = 0) = \begin{cases} \frac{E}{G^{2\delta L}} & \text{if } x \in [L/2 - \delta L, L/2 + \delta L] \\ 0 & \text{elsewhere} \end{cases},$$

$$u(x, t = 0) = \frac{EN}{GL},$$

$$m(x, t = 0) = 0,$$

$$r(x, t = 0) = 1,$$

where L is the length of the domain, δL is the size of the partitions of L , and N is the number of other families, initialised all as unmet and spread evenly in the domain. At the boundary, we again have reflective boundary conditions:

$$\begin{aligned} \left. \frac{\partial y(x, t)}{\partial x} \right|_{x=0} &= 0, & \left. \frac{\partial u(x, t)}{\partial x} \right|_{x=0} &= 0, \\ \left. \frac{\partial y(x, t)}{\partial x} \right|_{x=L} &= 0, & \left. \frac{\partial u(x, t)}{\partial x} \right|_{x=L} &= 0, \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial m(x,t)}{\partial x} \right|_{x=0} &= 0, & \left. \frac{\partial r(x,t)}{\partial x} \right|_{x=0} &= 0, \\ \left. \frac{\partial m(x,t)}{\partial x} \right|_{x=L} &= 0, & \left. \frac{\partial r(x,t)}{\partial x} \right|_{x=L} &= 0. \end{aligned}$$

For example, the numerical solution to these equations with the same parameter values and at the same time as Figure 4.29 is shown in Figure 4.30. These figures do not show the indentation for y , but here r shows some undulations that were not present in Figure 4.29 and u has a couple of peaks before the dent in the middle. This shows that the presence of both types of families, Y and U , the meeting events, and resource availability has altered the distribution of Y , U , and R .

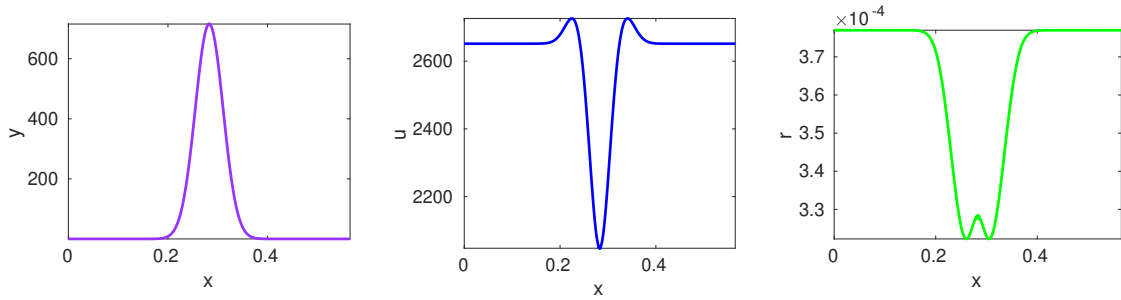


Figure 4.30: Numerical solution to Equations (4.62) to (4.65) for y (left), u (centre), and r (right) at time $t = 0.01$ days. One Y and 30 unmet families diffusing, eating, and meeting in one-dimensional space of length 200 km, diffusing at $25 \text{ km}^2/\text{day}$, with an eating rate of 50 per day, a growth rate of 1 per day, a meeting rate of $\mu = 1$ per day, and an initial maximum food level of $R_{max} = 5000$.

4.6.1.3 An example of high eating rate

Up to now we had been considering Y as a density in the mean-field equations, but to better see the interaction of a lone Y family with several U families and their surrounding resources we switch to a Gillespie stochastic simulation. The model and simulations will work as for the IMM from Section 4.2.2, but we will include the

resource as a non-diffusing entity that gets consumed by families in the space. As with the dynamics described in Section 4.6.1.2, the families have reflective boundary conditions and the resource starts at its maximum capacity.

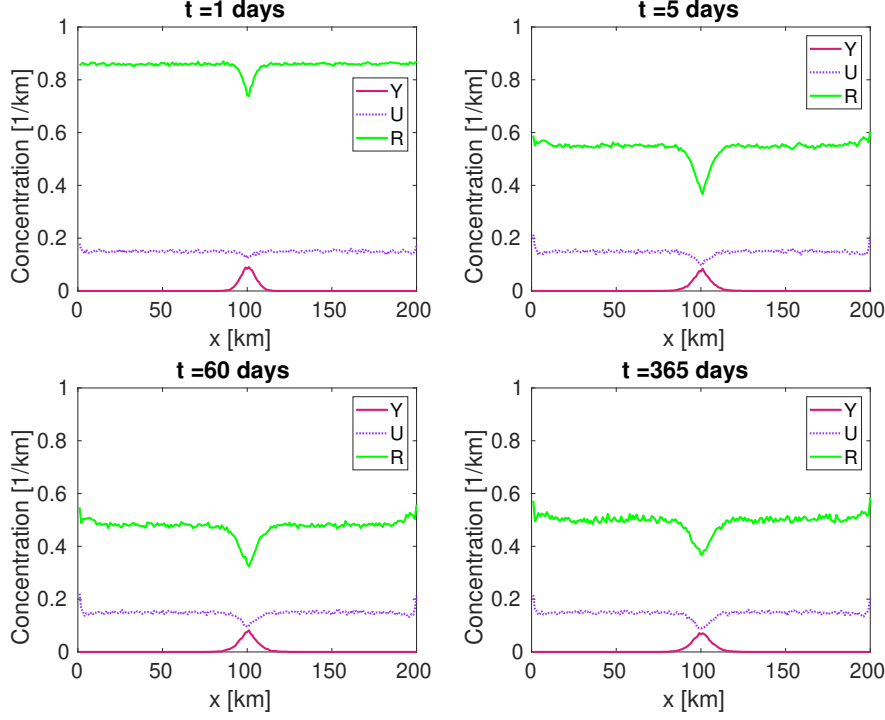


Figure 4.31: Average from 10,000 stochastic simulations. A family Y and 30 other initially unmet families diffusing in a one-dimensional domain of length $L = 200$ km, meeting and eating with rates $\mu = E = 1$ per day, forgetting each other every 24 hours, a detection distance of 1 km, a maximum resource of $R_{max} = 1$ unit, and a regrowth rate of $G = 1/365$ units per day. Each graph has the day stated at the top of the graph. All events treated as reactions except the forgetting event, which happens a day after a meeting occurs.

The first simulation we ran had a very high eating rate because the initial density of food was one unit throughout, whilst the eating rate was also 1 unit per day, meaning people would very quickly deplete the only available resource as soon as they reached a compartment. The result can be seen in Figure 4.31. Here, the population barely moves from its starting point because of the resources quickly disappearing and not growing back fast enough. This means that family Y will

only get to meet the closest people as soon as the year starts and then will have difficulties meeting other families, as shown in Figure 4.32.

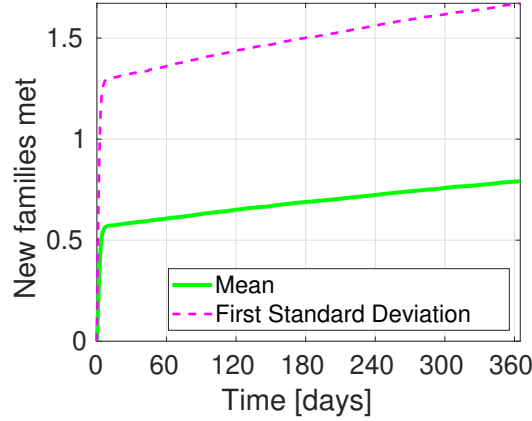


Figure 4.32: Number of new families Y meets when all families have a very high eating rate. Average from 10,000 stochastic simulations. All parameter values as in Figure 4.31.

It is clear from the work done in this section that the presence of a resource to be consumed and followed changes the behaviour of the families present, whereas, on its own, family Y normally just diffuses into a uniform distribution. During preliminary testing, parameter values that strongly showed the oscillatory behaviour of the distributions were the ones where a lot of initial resources were available and the eating rate was much higher than the growing rate, but orders of magnitude smaller than the initial resource level. This type of behaviour could directly affect the number families met by Y by keeping families restricted to specific areas of the domain, as demonstrated in the last case.

The results obtained when adding an interactive resource field to the models, even in the one-dimensional case, would point to the fact that a threshold ratio between resources available and resources eaten exists, all measured in terms of

resource growth, below which families will easily get trapped and stop meeting more families. If this happens, families will heavily depend on the number of families that closely surround them for support and sharing of knowledge, and will possibly not meet the 150-individual layer, or even the 15-individual layer. Whereas in the simple cases without food the families had easily met everyone in their 150-individual layer and we had considered the speed of the families to be too large for our interest, when resources are added, the variations in speed we were debating about become minor and play a less significant role on the number of meetings, and other parameter values, such as eating rate and initial amount of resources, become more important. As always, these models would have to be translated into two dimensions. We suggest that, for Dunbar's number to be reached in more realistic conditions, additional means of network cohesion would be needed.

Chapter 5

Conclusions

We have shown, throughout this Thesis, the importance of taking into account phenomena such as stochasticity and location when modelling certain biological problems. In Chapters 2 and 3 we developed a set of analytical and computational models to improve our understanding of the interactions that take place in the T cell membrane. In Chapter 4 we explored the possibility of societal maintenance by gathering and hunting movements of non-sedentary humans. The difference between the mean-field approach and ours is that we consider a specific element interacting with a small number of the surrounding elements, independently of the size scale of the problem.

Chapters 2 and 3 modelled the dephosphorylation and phosphorylation, respectively, of tethered receptors that are present in T cells. By comparing the analytically posed, but numerically solved, models to stochastic computational simulations we proved that our models fit the *in silico* data generated by the computational simulations (which we take to be the ground truth) better than the traditional mean-field ODEs in both cases. The numerical models in Chapter 2 have now been tested against laboratory experiments and have proven to be useful [1, 91] by providing

values for four biochemical parameters (those included in the model) and a good estimate of the length of the tethered tail, which is otherwise difficult to measure, from a single experiment. The model helped estimate that the tethered binding makes SHP-1 more active and helps cluster the dephosphorylating tails together, making the dephosphorylation more effective than if the proteins were floating in the cytoplasm. The MPDPDE model from Chapter 2 is also used to examine which sections of the tail are more relevant for dephosphorylation of surrounding peptides to occur [1]. Further on, the model was used to determine the length of SHP-1 when bound, showing that it had a longer reach than had been previously estimated from its structure [91]. We wanted to verify that the main interaction between SHP-1 and the peptide tail is as previously thought and the success of these experiments confirms our hypothesis.

Once experiments are available, the numerical, deterministic models in Chapter 3 can also be checked against experiments and, if verified, all of these numerical models could be useful to inform future experiments. As with the model in Chapter 2, the model in Chapter 3 can provide estimates for four different reaction rates, as well as the length of the protein being tethered, by fitting to data from just one experiment. Thus, these models provide a good insight into the mechanistic workings of processes we can not see. As such, the models would be very useful to help us understand further experiments where we know a change in the set-up produces a change in the outcome, but the reasons for this change are not yet understood. Further, the way these models are written and solved means they can be plugged into larger models, where larger sections of the immune system are incorporated. Despite the

existence of the stochastic simulations, the analytical models are preferred because they are quicker to solve, use less computational memory, and are as easy to use. The stochastic simulations are used, though, to ensure that our deterministic models correctly capture the chemistry we are describing, to determine the accuracy of our parameter predictions, and can be used to conclude if differences between our deterministic model and the experiments are due to numerical approximations or because we are not fully capturing the behaviour in the experiments.

Chapter 2 saw the reduction of the three-dimensional model to a two-dimensional model, and its analysis. This confirmed the suitability of each model for experiments where the proteins are tethered to matrices taller or shorter than the protein's length, respectively. A similar extension could be made for the model in Chapter 3, and we would expect a similar behaviour because of the similarity of protein movement, despite the difference of interaction.

Despite the encouraging results, the analytical models are not perfect. One potential source of error that they all share is the use of the Kirkwood approximation, where the interaction between three elements is approximated with three two-element interactions. Testing this hypothesis requires changing this approximation to one where four-element interactions are approximated with the four three-element possible interactions, which is a higher order moment closure. This would significantly increase the number of equations to be solved and would also increase the time to generate the numerical solution. Still, this does not seem to be necessary so far in light of the small errors and the proven usefulness of the models from Chapter 2 in laboratory experiments.

Chapters 2 and 3 can be seen as indicators of the importance of stochastic interactions, specifically within immune system regulation. Whereas the dephosphorylating SHP-1 reduces the amount of activated receptor tails and so helps down-regulate the signalling within the T cell, the phosphorylating ZAP70 does the opposite. Adding stochasticity to the system does not change the final number of phosphorylated tails at the steady state, but it does change the reaction rates, the speed with which we get to the steady state. Furthermore, the space-dependent nature of these systems changes the final conformation of the phosphorylated tails, meaning they will end up spread apart or grouped together in clusters, instead of uniformly randomly located. These properties probably have an effect on the immune system's efficiency, although we do not yet fully understand how. For example, having the activated tails in a cluster may help other proteins find them more quickly or interact with other necessary sections of the chain of reaction that are now nearby. Also, a steady state may never be reached, and the speed with which reactions take place may be crucial.

Study and awareness of these stochastic and location-dependent elements seem important to understand the immune system and its possible failures, whether by an over active response, or an insufficient response. We wanted to better understand how dephosphorylation and phosphorylation affect the signal transduction pathway of the immune system, helping turn it on and off, and with this work we now do, understanding how the tethered state influences the speed of reaction of dephosphorylation or phosphorylation, respectively. We better appreciate the regulatory mechanism in T cells and have successfully introduced space into the modelling of

these problems.

The work in Chapters 2 and 3 is of a basic nature, as it looks into the fundamental activation of T cells. As a result, it would apply to any disease or therapy involved with T cells, be it development of vaccines, understanding of autoimmune diseases, allergies, cancer, developing new diagnosis methods, or many more. Improved understanding of the process may allow us to exploit this knowledge for the benefit of human health.

In Chapter 4 we moved to a problem on a much larger spatial scale and considered the importance of face-to-face meeting for primitive humans. The interest sparked from the unusually large network that human individuals have, compared to other primates, where we all keep a conscious log of around 150 people (Dunbar’s number), but with restrictions on time to spend with “friends” solely socially, and the difficulty to find them when spread across large areas. Face-to-face meetings kept the social bonds alive in a time when no other means of communication existed. These social bonds were important because humans rely on others for aid against threats, physical and mental support, assistance in daily tasks like child-rearing, and for safe-keeping and dispersal of communal intelligence, amongst many other things, as explained in Section 4.1.1. So we decided to investigate whether a certain number of face-to-face meetings was possible through essential movement, or if other means for meeting and remembering each other would be necessary. The hierarchical arrangement of these societies was also noted when analysing our results. Initial attempts at modelling the interactions taught us that modelling hunter-gatherer societies in similar environments (biomes) did not suffice, as the range of number of meetings

for each individual was too large, biomes became mainly indistinguishable from each other, and it was not possible to say whether people would meet their different network layers (as seen in Figure 4.1). It was also clear that the absolute number of encounters each individual had could be approximated using the ideal-gas model, but this model was not able to provide information on the question that we are interested in, namely, how many different people our subject met. Here is where the stochasticity of the problem came into play. We further proved the need to include stochasticity into our models by showing that the traditional mean-field approximation did not accurately fit the computational simulations either.

In the end, we wrote a computational model with which we carried out a systematic testing of all the plausible parameter ranges, the parameters being the number of people in an area, size of that area, and the walking speed. With this we produced a set of look-up figures (Figures 4.24 and 4.25) that will in future allow us to see whether someone in their society of interest would meet their “active layer” (150 people) in a year, or any of the other network layers. Of course, this is for a randomly moving individual who moves all day at constant speed with their family of 5 and with a fixed memory function. All, and any, of these properties, along with many more, could be changed to generate a more realistic result, but this would depend on the specific research question as each change would require several hours of simulation, if not days or months, with the present method.

We set out to investigate whether Dunbar’s number (150 people) could be met by an individual roaming around an area in search of resources and we found that this is very dependent on the size of the area, the number of people in it, and the speed

at which the individual walks. There are several occasions when 150 people were met, especially when the area had more than enough people for the 150 number to be met (with more than 500 people in the domain, Dunbar's number was met more than half the time). In fact, this random movement could possibly work extremely well for fast-paced humans, or societies with a lot of people in a small space, where the whole tribe (1500 people) could be met in a year. But there are certainly some parameters where Dunbar's number was not met, those where the space was large and people were spread apart because of low concentrations, for example. This feeds into the idea that in harsher places, like the Arctic or the desert, with vast areas and a small population, mechanisms for meeting the 150 people would be necessary. The hypothesis is that this is how gift-giving and other kinds of material exchanges developed, as explained in Pearce et al. [153], making the memory of other members of the society last longer and so not needing to meet as often. The model in Chapter 4 could be enhanced to explore this by changing the memory function to a longer one in the case of a gift being exchanged, and running the simulations for several years. Another way of fixing the problem of insufficient meetings could have been the arrangement of periodical meetings, from every few days to yearly. In modern times we have restaurants and means of agreeing meetings, but some early humans had annual events, gathering for the solstice, or developed religion that brought people together. Later on, humans developed social events, such as sports. Presently, if we can not get interaction from movement, we have to put things in place to substitute, thus the importance of Skype, Zoom, Facebook, other forms of communication, and all kind of media. With the onset of the 2020 pandemic

and restrictions on socialisation, online gaming has seen a boost, especially multi-player games, as opposed to single-player games [154]. Yet, because for most of its existence the human evolved by adapting to face-to-face interaction, it is plausible that technological communication is not a good enough substitute. The effect that this change of type of interaction has on mental health remains as an open question.

Chapter 4 also saw a natural rise of different numbers of meetings for different people in the society of an individual, as seen in Figure 4.12. Where anthropologists talk about a hierarchical organisation of an individual's society (Figure 4.1), it is possible that this organisation emerged naturally as a mathematical and geographical construct, rather than a human or biological imposition. The causality of this hierarchical arrangement would be difficult to verify, but further research could be carried out into the source of the 'steps' seen in the plots for numbers of meetings per family (Figure 4.12).

Along the way, an analytical, exact solution was found for the probability of a particle's position given a discrete space and continuous time diffusion, in a space with reflective boundaries. Diffusion in a discrete space over continuous time is a fundamental problem that appears in numerous contexts. Different versions of this problem have been solved with different methods throughout time [145, 146, 147], but the exact problem we deal with has not been solved analytically and without approximations before, as explained at the end of Section 4.2.1.2. Thus, our result could be applicable in a wide variety of problems such as in random walks [155], stochastic processes in physics and astronomy [145], heat dispersion in nano-films [156], signal and image processing [146], and mutations in biology [147].

The models presented in this thesis can be improved in several different ways: accuracy, speed, compatibility with other models in the area, to mention but a few. But we have already improved on our mean-field, homogeneous counterparts (the mean-field models in Sections 2.2, 3.2, and 4.2.3.1). We have taken methods developed in other areas of mathematics and computation and used them to gain insight into biological questions. In doing so we have made some extensions to these techniques. We have worked in an interdisciplinary manner to extend the reach of mathematics to address problems of interest to non-mathematicians. And, more importantly, we have once again pointed out the importance of not taking small numbers or relative position lightly.

Appendix A

Integral in Mean Field model in Chapter 2

We want to compute the integral term for Equation (2.14) in Section 2.2.2 as we try to solve the traditional mean field approximation of our system in Chapter 2.

Because $\sigma(r)$ only depends on the absolute distance between particles we can write it as $\sigma(|\vec{x} - \vec{x}'|)$. In spherical coordinates this is:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} r'^2 \sin \theta \sigma(|\vec{x} - \vec{x}'|) dr' d\theta d\phi. \quad (\text{A.1})$$

If we make a change of variable and use $q = |\vec{x} - \vec{x}'| = \sqrt{r'^2 + r^2 - 2rr' \cos \theta}$, then the expression (A.1) becomes:

$$\frac{2\pi}{r} \left(\frac{3}{2\pi L^2} \right)^{3/2} \int_{r'=0}^{\infty} r' \int_{q=|r-r'|}^{|r+r'|} q \exp \left(-\frac{3q^2}{2L^2} \right) dq dr'. \quad (\text{A.2})$$

Calculating the integral with respect to q :

$$\int_{q=|r-r'|}^{|r+r'|} q \exp \left(-\frac{3q^2}{2L^2} \right) dq = \frac{L^2}{3} \left[e^{-\frac{3}{2L^2}(r'^2+r^2-2rr')} - e^{-\frac{3}{2L^2}(r'^2+r^2+2rr')} \right]. \quad (\text{A.3})$$

Substituting Equation (A.3) into the integral with respect to r' in the expression

(A.2) we obtain:

$$\begin{aligned}
& \int_{r'=0}^{\infty} r' \int_{q=|r-r'|}^{|r+r'|} q \exp\left(-\frac{3q^2}{2L^2}\right) dq dr' \\
&= \frac{L^2}{3} \left\{ \frac{1}{6} L \left[\sqrt{6\pi} r \left(\text{Erf}\left(\frac{\sqrt{\frac{3}{2}}(r+r')}{L}\right) - \text{Erf}\left(\frac{\sqrt{\frac{3}{2}}(r-r')}{L}\right) \right) \right. \right. \\
&\quad \left. \left. + 2L \left(e^{-\frac{3(r+r')^2}{2L^2}} - e^{-\frac{3(r-r')^2}{2L^2}} \right) \right] \right\} \Bigg|_0^{\infty},
\end{aligned}$$

where $\text{Erf}(z)$ is the “error function”, defined by:

$$\text{Erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp^{-t^2} dt.$$

In the limit as z tends to infinity, $\text{Erf}(z)$ is 1 for $z \neq 0$, zero when $z = 0$, and $\text{Erf}(-z) = -\text{Erf}(z)$. This leads us to the evaluated integral:

$$\int_{r'=0}^{\infty} r' \int_{q=|r-r'|}^{|r+r'|} q \exp\left(-\frac{3q^2}{2L^2}\right) dq dr' = \frac{L^3}{3} \sqrt{\frac{2\pi}{3}} r. \quad (\text{A.4})$$

Substituting Equation (A.4) back into the expression (A.2) we obtain:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r'=0}^{\infty} r'^2 \sin \theta \sigma(|\vec{x} - \vec{x}'|) dr' d\theta d\phi = 1. \quad (\text{A.5})$$

Appendix B

Ode23 and ode15s MATLAB solvers for the MPDPDE of Chapter 3

Whilst solving the MPDPDE model in three dimensions in Chapter 3, the numerical solver used was taking too long to solve. We study here the cause of this.

The numerical package that we used at first to solve Equations (3.55) to (3.63) was MATLAB's "ode23", using the trapezoidal rule for the integrals, which states:

$$\int_a^b f(x) \, dx \approx (b - a) \left[\frac{f(a) + f(b)}{2} \right]. \quad (\text{B.1})$$

It was seen that the behaviour of the solution represented closely that of the solution of a "stiff" set of differential equations. This can be inferred in Figure B.1, where the functions X_A and X_B change very rapidly in the beginning, whereas X_C takes a while to change. This is, they seem to be in different time scales.

On the other hand, it can be seen in Figure B.2 that the time taken to solve the set of equations seems to be taking exponentially longer, despite the system seemingly having reached the steady state since the Experimental Time $\times k_{off} = 30$. Although the laboratory experiments typically run for 300 seconds, if we

extrapolate the exponential growth in Figure B.2 the expected time to reach this ‘Experimental Time’ is of 1.57×10^9 days (taking into account a reasonable value of k_{off} of 1 per second), which is prohibitively long. We hypothesise that this long solving time toward the end may be due to n_C approaching 0, where it is no longer adequate to talk about densities, whereas our MPDPDE model deals with densities.

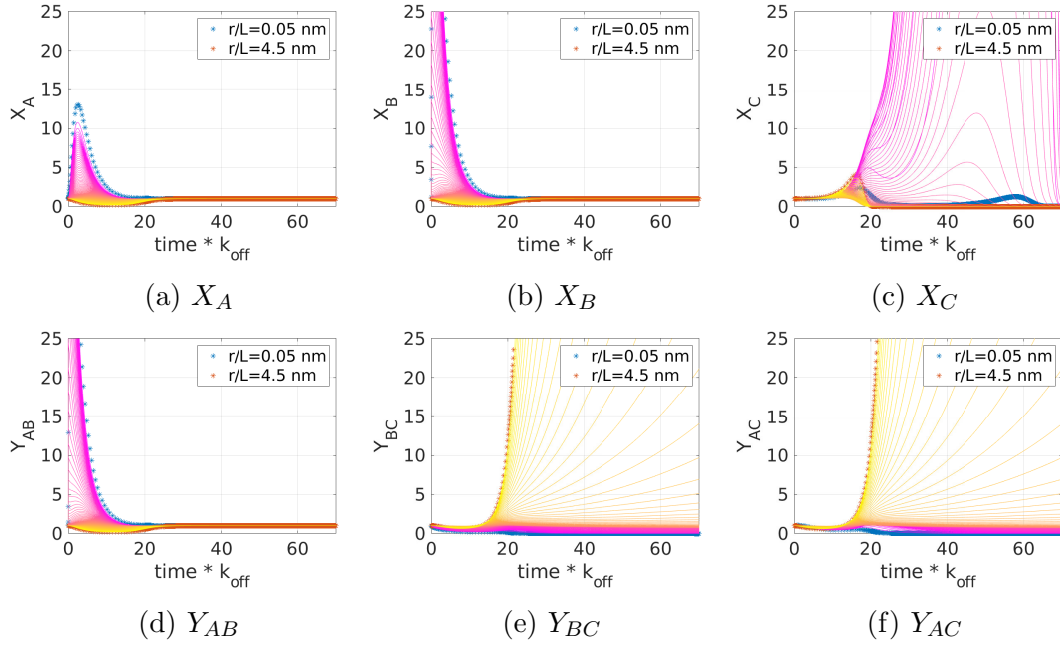


Figure B.1: Pair correlation functions of three-dimensional MPDPDE model of Chapter 3, Equations (3.55) to (3.63). Shown are the solutions to each pair correlation function over time at each discretisation of r , from the closest $r/L = 0.05$ in blue, to the furthest $r/L = 4.5$ in red. Hence 90 solutions to each function. Numerically solved by MATLAB’s ode23. Parameters used are a peptide concentration of 3×10^{-4} peptides/nm³ (equivalent to a height and width of 1,000 nm with 300,000 peptides), $k_{on}^* = 0.1 \text{ s}^{-1}$, $k_{off} = 1 \text{ s}^{-1}$, $k_{cat} = 0.02 \mu\text{M}^{-1}\text{s}^{-1}$, $L = 20 \text{ nm}$, and $\lambda = 0.005 \text{ s}^{-1}$.

For the first problem, a stiff ODE solver was then used in MATLAB, “ode15s”.

This solver took longer to solve than the ode23 solver for Experimental Time 30 (about 3 minutes), but then took about 4.8 minutes to solve the equations up to 70. Yet, ode15s stops at Experimental Time 211 because it was “Unable to meet

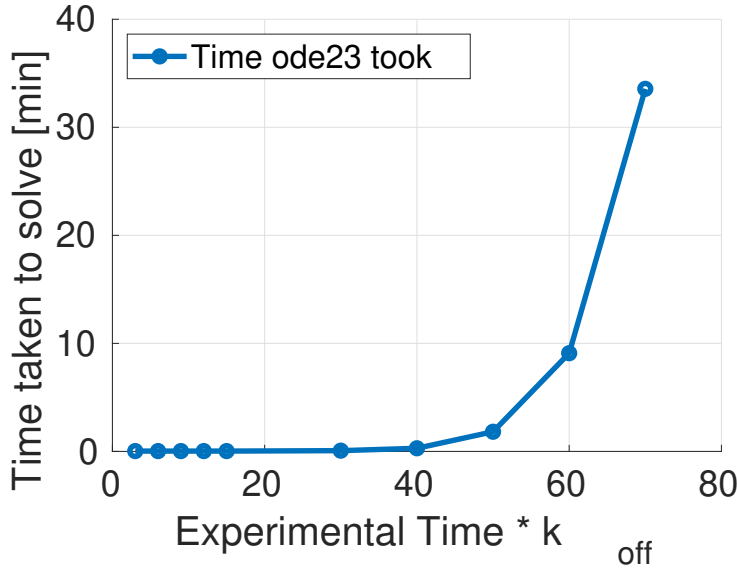


Figure B.2: Time taken for numerical solver, MATLAB’s “ode23”, to solve our MPDPDE system of equations up to a given ‘Experimental Time’. $p_1 = 0.100000$, $p_2 = 9.961813$, $p_3 = 4.150755$, $p_4 = 0.005000$, $C_T = 0.0003$ peptides per nm^3 , and $k_{off} = 1$ per s.

integration tolerances without reducing the step size below the smallest value allowed (4.547474×10^{-13}).” Still, ode15s reaches this point in less than 15 minutes of running time, which is still better than ode23.

When compared, it was seen that both solvers had very similar solutions to the MPDPDE system of equation, and their difference with the stochastic simulation is small. The comparison between both ODE solvers and how close they get to the stochastically produced data can be seen in Figure B.3. We note that the close agreement between the stochastic simulations and the MPDPDE model suggests that the MPDPDE model does reproduce the stochastic simulations, even though the MPDPDE model includes simplifications (like Kirkwood approximation and integral truncation) that the stochastic simulations do not have. This suggests that the MPDPDE model can be used to fit data.

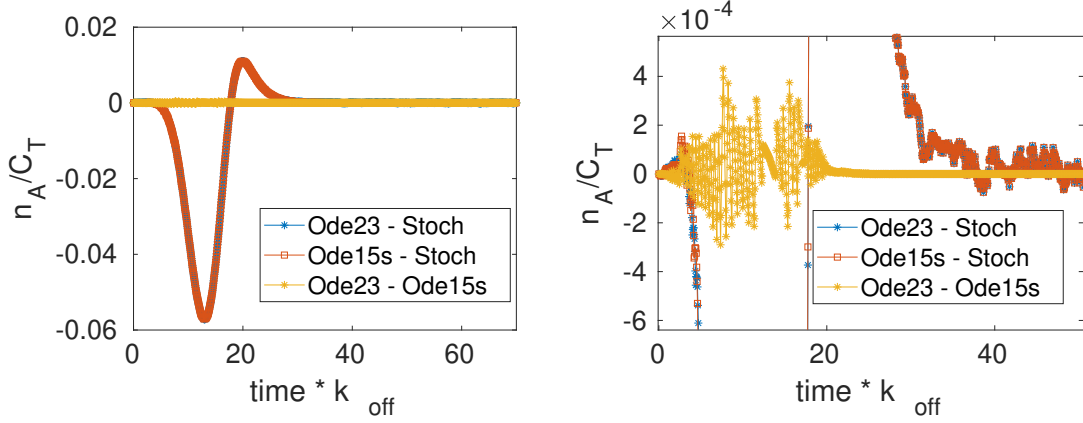


Figure B.3: Comparing the stiff (ode15s) and non-stiff (ode23) ODE solvers in MATLAB with the data produced by the stochastic simulations from Chapter 3. The blue line with asterisks is the difference in A concentration between non-stiff solver and data from stochastic simulations, whilst the red line with squares is the difference between the stiff solver and the same data from stochastic simulations. We note these are very similar. This is confirmed by plotting the difference between the stiff and non-stiff solutions, shown by the yellow line with asterisks, which seems flat (at zero) on the left. The difference can only be appreciated when we zoom in (on the right). Clearly, the solutions generated by the ODE solvers are not too different from each other, but ode15s managed to solve the equations much faster.

Given that none of the ODE solvers reach the 300 seconds of Experimental Time in a satisfactory amount of time, and taking into account that the MPDPDE model should not be used when the number of particles is too close to 0, none of the numerical solutions could be used alone for data fitting. Yet, given that we know that the MPDPDE model reaches the steady state quickly, this knowledge could be used to shorten drastically the time needed to run the model. That is, once the solutions hit the steady state values, given in Equations (3.14), the MATLAB solver could be stopped.

Appendix C

Gillespie and Smoluchowski comparison

Following work done in Section 4.1.3, in this section we show an example of the program used to compare our spatially extended Gillespie Stochastic Simulation Algorithm (SSA) with the Smoluchowski SSA. An example of the visualization produced by *Smoldyn*, which is based on the Smoluchowski SSA, at two different points in time can be seen in Figure C.1. This figure is also an example of how the met families distribute after having been met by family Y .

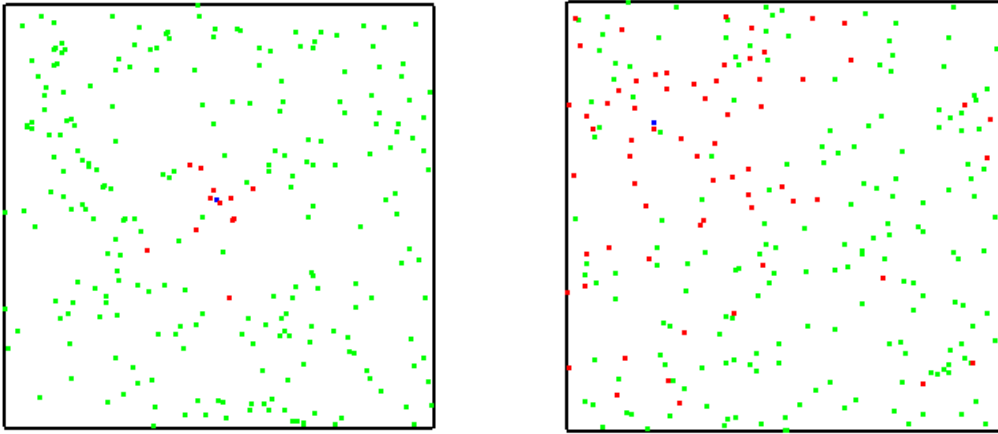


Figure C.1: An example of the visualization of the *Smoldyn* dynamics at two points in time. Family Y in blue, unmet families in green and met families in red. Done with 245 initial unmet families, for 365 days, with a forgetting time of 1 day, a detection distance of 1 km, a diffusion constant of $9 \text{ km}^2/\text{day}$, and a domain of $78 \times 77 \text{ km}^2$.

The comparison between two Smoluchowski SSA realizations and the average of 2000 realizations with the Gillespie SSA that we used all throughout the dissertation in any stochastic simulations is shown in Figure C.2. Here we can see that, despite the Smoluchowski not fitting exactly between the first standard deviations of the mean of the Gillespie SSA, both Smoluchowski simulations follow the Gillespie's mean shape and the difference could be attributed to the stochastic nature of the simulations. This points towards the fact that both algorithms can be interchangeably used for our purposes and choosing the Gillespie SSA because of speed and eventual specific use of the domain for, for example, resource allocation, can be admitted.

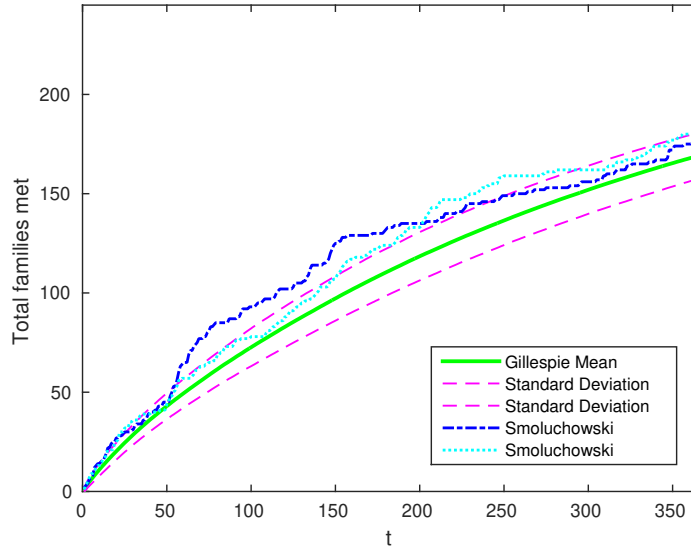


Figure C.2: Number of families that family Y has met over time, only counting the first time they were met. Average of 2000 runs with the Gillespie SSA compared to two particular runs done with the Smoluchowski method using *Smoldyn*. The mean of the Gillespie simulations is shown in green, whilst its first standard deviations have been added above and below in a dashed magenta line. Done in two dimensions with 245 initial unmet families, for 365 days, with a forgetting time of 1 day, a detection distance of 1 km, a diffusion constant of $9 \text{ km}^2/\text{day}$, and a domain of $78 \times 77 \text{ km}^2$.

Appendix D

Table of data used for parameter values

Here we present an extract of the table from where real-world data were obtained to calculate the parameter values used in the two-dimensional simulations for biomes, in Section 4.4.1.1, and societies, in Section 4.4.1.2. Put together by Binford [117], found in Grove et al. [3, Table 5.01], and kindly provided to us by Dr. Eiluned Pearce.

Table D.1: An extract of the data used from Grove et al. [3, Table 5.01]

Society Name	Location	Absolute Latitude	General Biome	Area [km ²]	Estimated Total Annual Distance Moved [k]	Population	Mean Family
West Greenland	Greenland	69.31	Arctic	108000	70	5112	4.84
Great Whale Inuit	Quebec	55.32	Arctic	24200	85	450	4.72
Siberian Eskimo	Russia	65.96	Arctic	27490	90	1292	4.72
Labrador Inuit	Newfoundland	57.97	Arctic	52500	135	1460	5.7
Wackende Inuit	Northwest Territories	69.86	Arctic	65000	145	2500	
Baffin Island Inuit	Northwest Territories	65	Arctic	370000	175	6000	3.76
Kobuk Inuit	Alaska	66.95	Arctic	18700	225	500	4.5
Netsilik Inuit	Northwest Territories	71.46	Arctic	197000	307	500	3.74
Polar Inuit	Greenland	77.49	Arctic	73100	350	300	4.29
Noatak Inuit	Alaska	68.09	Arctic	25000	360	550	3
Ukukihaling-miut	Northwest Territories	66.99	Arctic	43100	380	164	4.37
Inguik Inuit	Northwest Territories	69.44	Arctic	221000	385	1193	
Avillingmiut Inuit	Northwest Territories	65.16	Arctic	40000	400	130	
Caribou Inuit	Northwest Territories	64.33	Arctic	236500	440	700	4.36
Nunamut Inuit	Alaska	68.18	Arctic	24900	501	240	
Lungga	Australia - Western	16.91	Australia: subtropical desert through temperate settings	12500	0	562	
Karaderi	Australia - Western	18.92	Australia: subtropical desert through temperate settings	14300	0	536	
Aranda-southern	Australia - Northern Territory	26.09	Australia: subtropical desert through temperate settings	45500	0	500	
Jalaria	Australia - New South Wales	29.3	Australia: subtropical desert through temperate settings	12000	0	1080	
Barkindji	Australia - New South Wales	32.4	Australia: subtropical desert through temperate settings	19500	0	3008	
Wongalbon	Australia - New South Wales	32.14	Australia: subtropical desert through temperate settings	70110	0	3589	
Undanbi	Australia - Queensland	27.28	Australia: subtropical desert through temperate settings	2300	0	500	
Warunggu	Australia - Queensland	18.41	Australia: subtropical desert through temperate settings	9600	0	1562	
Kurnai	Australia - Victoria	37.59	Australia: subtropical desert through temperate settings	1900	0	336	
Minng	Australia - Western	34.95	Australia: subtropical desert through temperate settings	12700	0	889	6.66
Dieri	Australia - South	28.54	Australia: subtropical desert through temperate settings	21800	0	420	
Anbara	Australia - Northern Territory	12.2	Australia: subtropical desert through temperate settings	80	17	35	
Kuku-Yalanji	Australia - Queensland	15.92	Australia: subtropical desert through temperate settings	1300	18	650	6
Yinjiingga	Australia - Queensland	14	Australia: subtropical desert through temperate settings	700	25	217	
Imbarga	Australia - Queensland	27.25	Australia: subtropical desert through temperate settings	1370	32	219	
Lardi	Australia - Queensland	16.3	Australia: subtropical desert through temperate settings	400	35	120	6.33
Kaladit	Australia - Queensland	16.94	Australia: subtropical desert through temperate settings	250	35	165	6.68
Muluk	Australia - Northern Territory	13.58	Australia: subtropical desert through temperate settings	1000	40	445	
Larkla	Australia - Northern Territory	12.6	Australia: subtropical desert through temperate settings	3900	60	1560	
Groote Eylandt	Tasmania	41.38	Australia: subtropical desert through temperate settings	2600	60	595	6
Tasmanians-western	Tasmania	41.38	Australia: subtropical desert through temperate settings	3370	60	450	
Twi	Australia - Northern Territory	11.59	Australia: subtropical desert through temperate settings	7100	75	2662	7.4
Nungubuyu	Australia - Northern Territory	13.76	Australia: subtropical desert through temperate settings	7000	85	1610	
Badjalang	Australia - New South Wales	29.59	Australia: subtropical desert through temperate settings	6000	90	804	
Bunurong	Australia - Victoria	38.93	Australia: subtropical desert through temperate settings	7800	90	1953	
Karuna	Australia - South	34.56	Australia: subtropical desert through temperate settings	7200	100	1296	
Kariera	Australia - Western	20.88	Australia: subtropical desert through temperate settings	11700	130	1111	
Yir-yoront	Australia - Queensland	15	Australia: subtropical desert through temperate settings	1300	145	104	4.2
Worora	Australia - Western	15.32	Australia: subtropical desert through temperate settings	10400	150	1114	
Kakadu	Australia - Northern Territory	13.13	Australia: subtropical desert through temperate settings	6000	150	528	
Muringin-Yolngu	Australia - Northern Territory	12.93	Australia: subtropical desert through temperate settings	5000	160	588	
Tasmanians-eastern	Tasmania	42.62	Australia: subtropical desert through temperate settings	8570	165	700	
Ieldji-Forestriver	Australia - Western	15.55	Australia: subtropical desert through temperate settings	2600	218	442	
Wikmunkan	Australia - Queensland	13.47	Australia: subtropical desert through temperate settings	8300	238	1602	
Karwali	Australia - Queensland	24.41	Australia: subtropical desert through temperate settings	33000	275	660	
Aranda-northern	Australia - Northern Territory	23.7	Australia: subtropical desert through temperate settings	76700	285	2045	6.4
Pitlandjara	Australia - Northern Territory	26	Australia: subtropical desert through temperate settings	59800	289	358	6.96
Alawajara	Australia - South	27	Australia: subtropical desert through temperate settings	57200	298	572	6.26
Alawajara	Australia - Northern Territory	22.27	Australia: subtropical desert through temperate settings	46300	325	560	6.75
Pintubi	Australia - Northern Territory	22.86	Australia: subtropical desert through temperate settings	22900	325	343	5.8
Djuru	Australia - Western	19.16	Australia: subtropical desert through temperate settings	33800	350	1345	
Nakako	Australia - Western	27.31	Australia: subtropical desert through temperate settings	49400	360	429	8
Ngatadjara	Australia - Western	25.32	Australia: subtropical desert through temperate settings	78000	373	312	6.18

Idawongga	Australia - Western	22.93	Australia: subtropical desert through temperate settings	78700	380	354	6.94
Walbiri	Australia - Northern Territory	20	Australia: subtropical desert through temperate settings	137800	380	1598	4.51
Marudjara	Australia - Western	22.79	Australia: subtropical desert through temperate settings	22600	416	169	5.2
Acsugewi	California	40.75	North America: California and Northern Mexico	7250	0	1300	
Maad-Mountain	California	40.28	North America: California and Northern Mexico	8100	28	1000	
Wiwok	California	38	North America: California and Northern Mexico	4950	32	1212	
Wukuthmi	California	36.45	North America: California and Northern Mexico	380	50	92	
Shasta	California	41.62	North America: California and Northern Mexico	11700	55	2925	
Achnunawi	California	41.32	North America: California and Northern Mexico	9850	64	1700	
Tekeima	California	42.44	North America: California and Northern Mexico	7000	70	900	
North Foothill Yokuts	California	37.44	North America: California and Northern Mexico	940	70	360	
Diegueno	California	32.44	North America: California and Northern Mexico	16600	80	3000	
Klamath	Oregon	42.62	North America: California and Northern Mexico	8980	84	1200	
Yana	California	40.38	North America: California and Northern Mexico	5810	85	1820	
Monache	California	36.6	North America: California and Northern Mexico	1350	85	387	
Kiliwa	Mexico	31.44	North America: California and Northern Mexico	3640	90	445	
Seri	Mexico	29.59	North America: California and Northern Mexico	2160	95	550	5.53
Yuki-Coast	California	39.64	North America: California and Northern Mexico	1120	100	750	
Tubatulabal	California	36	North America: California and Northern Mexico	5800	110	1000	3.81
Chichimec	Mexico	22	North America: desert and desert scrub	33300	0	3000	
North Fork Paiute	California	37	North America: desert and desert scrub	2400	0	385	4.93
Bohogue North Shoshoni	Idaho	43	North America: desert and desert scrub	36540	0	380	
Bear Creek Paiute	Oregon	43.94	North America: desert and desert scrub	5400	0	60	
Ute-Wimianitci	Utah	37.67	North America: desert and desert scrub	15600	0	405	
Antelope Valley Shoshoni	Nevada	40.19	North America: desert and desert scrub	5840	0	78	4.37
Uncompahgre Ute	Utah	38.14	North America: desert and desert scrub	25640	0	1100	
Coahuilenos	Mexico	26	North America: desert and desert scrub	37300	0	600	
Ute-Timpanogas	Utah	40.22	North America: desert and desert scrub	13880	0	480	
Uintah Ute	Utah	40.47	North America: desert and desert scrub	23370	0	1750	3.06
Spring Valley Shoshoni	Nevada	39.19	North America: desert and desert scrub	6260	0	378	
White Knife Shoshoni	Nevada	41	North America: desert and desert scrub	4370	0	500	5
Wind River Shoshoni	Wyoming	43.19	North America: desert and desert scrub	96000	0	1500	
Karakawa	Texas	28.44	North America: desert and desert scrub	13500	0	2800	3.7
Pyramid Lake Paiute	Nevada	40	North America: desert and desert scrub	2620	33	485	
Rainroad Valley Shoshoni	Nevada	38.33	North America: desert and desert scrub	5840	90	250	
Catrail Paiute	Nevada	40.12	North America: desert and desert scrub	2180	90	481	
Ruby Valley Shoshoni (Wadaduka)	Nevada	40.36	North America: desert and desert scrub	3260	120	450	
Reese Valley Shoshoni	Nevada	39.25	North America: desert and desert scrub	2340	130	390	
Guaicura	Mexico	25	North America: desert and desert scrub	16290	135	977	
Kawalisu Shoshoni	California	35.37	North America: desert and desert scrub	4200	150	500	
Fish Lake Paiute	Nevada	37.69	North America: desert and desert scrub	2570	100	100	3.67
Honey Lake Paiute	California	40.27	North America: desert and desert scrub	3630	165	385	
Saline Valley Shoshoni	California	36.65	North America: desert and desert scrub	2800	180	65	
Deep Spring Paiute	California	37.28	North America: desert and desert scrub	650	195	23	
Washo	Nevada	39.08	North America: desert and desert scrub	12600	198	1877	
Antarianunts-Southern Paiute	Colorado	37.86	North America: desert and desert scrub	6800	200	234	
Kaibab Southern Paiute	Arizona	36.9	North America: desert and desert scrub	11460	200	425	
Surprise Valley Paiute	Nevada	41.5	North America: desert and desert scrub	2700	210	367	
Salmon-eater Shoshoni	Idaho	42.94	North America: desert and desert scrub	5790	210	400	
Mono Lake Paiute	California	38.11	North America: desert and desert scrub	2880	215	170	
Panamint Shoshoni	California	36.37	North America: desert and desert scrub	23600	220	500	
Koso Mountain Shoshoni	California	36.12	North America: desert and desert scrub	2590	220	222	
Death Valley aka Timbisha Shoshoni	California	36.52	North America: desert and desert scrub	3270	225	42	
Gosiute Shoshoni	Utah	39.88	North America: desert and desert scrub	25970	228	435	
Harney Valley Paiute	Oregon	43.25	North America: desert and desert scrub	16080	250	200	
Hukurdake Shoshoni	Utah	41.53	North America: desert and desert scrub	33760	260	1000	
Little Smoky Shoshoni	Nevada	39.33	North America: desert and desert scrub	5260	260	96	
Kawich Mountain Shoshoni	Nevada	37.92	North America: desert and desert scrub	5260	270	105	
Sheep-eater Shoshoni	Idaho	45.17	North America: desert and desert scrub	8810	270	550	

Grouse Creek Shoshoni	Utah	41.59	North America: desert and desert scrub	12700	315	200	4.38
Umatekla	Washington	45.49	Subarctic and continental midlatitude forests if North America and Asia	11000	0	1155	
Yakima	Washington	46.33	Subarctic and continental midlatitude forests if North America and Asia	14080	0	3780	
Mountain	Northwest Territories	63.68	Subarctic and continental midlatitude forests if North America and Asia	100000	0	780	4.72
Wegagon Ojibwa	Ontario	52.79	Subarctic and continental midlatitude forests if North America and Asia	49000	0	250	
Round Lake Ojibwa	Ontario	52.71	Subarctic and continental midlatitude forests if North America and Asia	14290	0	250	
Talitha	British Columbia	57.66	Subarctic and continental midlatitude forests if North America and Asia	70000	0	810	
Tan	Tukon	64.79	Subarctic and continental midlatitude forests if North America and Asia	35080	0	1000	4.5
Waswanip Cree	Quebec	49.78	Subarctic and continental midlatitude forests if North America and Asia	35800	0	150	5.65
Flathead	Montana	46.28	Subarctic and continental midlatitude forests if North America and Asia	53300	0	800	
Kalispel	Idaho	47.94	Subarctic and continental midlatitude forests if North America and Asia	66600	0	1000	3.62
Nez-perce	Idaho	46.25	Subarctic and continental midlatitude forests if North America and Asia	45000	0	4000	
Tenino	Washington	45.59	Subarctic and continental midlatitude forests if North America and Asia	7950	70	1510	
Saipoli	Washington	47.94	Subarctic and continental midlatitude forests if North America and Asia	6140	75	687	
Atlatwipiskat Cree	British Columbia	51.32	Subarctic and continental midlatitude forests if North America and Asia	5400	90	405	
Ojibwa-Kitchibuan	Ontario	53.62	Subarctic and continental midlatitude forests if North America and Asia	312000	215	4460	4.18
Micmac	Michigan	45.21	Subarctic and continental midlatitude forests if North America and Asia	60000	225	3000	3.5
Ojibwa-Northern Albany	New Brunswick	46.99	Subarctic and continental midlatitude forests if North America and Asia	101700	252	4393	5.45
Kaska	Ontario	51.22	Subarctic and continental midlatitude forests if North America and Asia	15740	275	225	
Coeur d'Alene	British Columbia	59.73	Subarctic and continental midlatitude forests if North America and Asia	60000	280	540	3.43
Montagnais	Idaho	47.42	Subarctic and continental midlatitude forests if North America and Asia	66600	300	1000	3.58
Rupert House Cree	Quebec	52.42	Subarctic and continental midlatitude forests if North America and Asia	660000	318	2700	6.6
Koyukon	Quebec	51.08	Subarctic and continental midlatitude forests if North America and Asia	50000	340	450	
Nipigon Ojibwa	Alaska	66.52	Subarctic and continental midlatitude forests if North America and Asia	182500	350	2000	
Rainy River Ojibwa	Ontario	49.34	Subarctic and continental midlatitude forests if North America and Asia	25500	350	221	
Saudene	Ontario	45.21	Subarctic and continental midlatitude forests if North America and Asia	19010	350	230	4.2
Nganasan	Northwest Territories	65.75	Subarctic and continental midlatitude forests if North America and Asia	150000	375	825	
North Saultraux	Russia	73.83	Subarctic and continental midlatitude forests if North America and Asia	190430	375	876	
Beaver	Ontario	52.28	Subarctic and continental midlatitude forests if North America and Asia	15420	375	185	
Selcani	Alberta	58.53	Subarctic and continental midlatitude forests if North America and Asia	194700	400	1000	4.12
Tukaghir	Alberta	56.14	Subarctic and continental midlatitude forests if North America and Asia	80000	400	656	
Naskapi	Russia	70	Subarctic and continental midlatitude forests if North America and Asia	386880	410	2360	4.08
Ket	Quebec	56.19	Subarctic and continental midlatitude forests if North America and Asia	96000	420	400	4.81
Slave	Russia	62	Subarctic and continental midlatitude forests if North America and Asia	62500	420	1000	
Tutchone	Northwest Territories	61.15	Subarctic and continental midlatitude forests if North America and Asia	245370	445	2454	
Hare	Alaska	61.24	Subarctic and continental midlatitude forests if North America and Asia	162000	450	1500	3.33
Dogrib	Northwest Territories	67.47	Subarctic and continental midlatitude forests if North America and Asia	173400	450	572	
Kutchin	Northwest Territories	63.85	Subarctic and continental midlatitude forests if North America and Asia	180900	450	1590	
Mistassini Cree	Tukon	65.94	Subarctic and continental midlatitude forests if North America and Asia	286100	450	4863	
Chipewyan aka Dene	Quebec	51.75	Subarctic and continental midlatitude forests if North America and Asia	77900	450	450	4.55
Xam	Saskatchewan	59.47	Subarctic and continental midlatitude forests if North America and Asia	619400	496	2850	4.19
Kung	South Africa	31.47	Tropical-subtropical Africa	12350	0	300	
Hadza	Botswana	20	Tropical-subtropical Africa	11000	75	726	5.08
Hukwe	Tanzania	3.82	Tropical-subtropical Africa	2500	80	600	4.5
G'wi	Zambia	16.93	Tropical-subtropical Africa	4300	125	124	
Iko	Botswana	22.46	Tropical-subtropical Africa	18000	270	528	2.92
Auni-knomani	Botswana	23.86	Tropical-subtropical Africa	11800	310	122	
Shompen	South Africa	27.37	Tropical-subtropical Africa	57000	352	364	
Onge	Nicarbar Islands	7	Tropical-subtropical Asia	860	0	342	3.62
Iawa	Andaman Islands	10.7	Tropical-subtropical Asia	980	40	393	3.11
North Island	Andaman Islands	12.19	Tropical-subtropical Asia	570	59	255	
Orogens	Andaman Islands	13.32	Tropical-subtropical Asia	1390	76	464	3.77
Chono	China	51.91	Tropical-subtropical Asia	3200	350	460	
Heta	Chile	45	Tropical-subtropical Asia	15390	0	2100	
Ala	Brazil	45	Tropical-subtropical South America	310	0	30	
Yaligau	Chile	23.52	Tropical-subtropical South America	22690	0	3400	4.8
Nukak	Argentina	49.55	Tropical-subtropical South America	8600	90	2500	3.5
Akurtyo	Colombia	60	Tropical-subtropical South America	3010	225	282	3.99
	Suriname	2.56	Tropical-subtropical South America	1640	285	66	
		3					

Ona	Argentina	53.9	Tropical-subtropical South America	48100	320	3497	
Tehuelche	Argentina	46	Tropical-subtropical South America	475300	360	9000	5

Appendix E

Stochastic to deterministic equation of diffusion

In Chapter 4, Section 4.2.1.2, we obtain an analytical solution to the stochastic equation of diffusion for one particle diffusing in a bounded lattice. Here we show how the stochastic representation, which starts in Equation (4.4), is related to the traditional deterministic equation of diffusion.

We define the joint probability of population A_k having n_k families, $k = 1, \dots, K$ at time t as $p(\vec{n}, t)$, where $\vec{n} = [n_1, n_2, \dots, n_K]$. K being the number of compartments in the one-dimensional domain. We define right-jumping and left-jumping operators:

$$\mathcal{R}_i \vec{n} = [\dots, n_i + 1, n_{i+1} - 1, \dots],$$

$$\mathcal{L}_i \vec{n} = [\dots, n_{i-1} - 1, n_i + 1, \dots].$$

The evolution of the stochastic model described by Figure 4.3 will be:

$$\begin{aligned} p(\vec{n}; t + \Delta t) = & \sum_{i=1}^{K-1} d\Delta t (n_i + 1) p(\mathcal{R}_i \vec{n}; t) \\ & + \sum_{i=2}^K d\Delta t (n_i + 1) p(\mathcal{L}_i \vec{n}; t) \\ & + p(\vec{n}; t) \left(1 - \sum_{i=1}^{K-1} d\Delta t n_i - \sum_{i=2}^K d\Delta t n_i \right), \end{aligned}$$

where the first term on the right-hand side describes the probability of a particle being in the left compartment at time t and jumping right in Δt , the second term describes the probability of a particle being in the right compartment at time t and jumping left in Δt , and the third term is the probability of the system being in the wanted configuration at time t and not jumping right or left in Δt . With algebraic manipulation we can obtain:

$$\begin{aligned} \frac{p(\vec{n}; t + \Delta t) - p(\vec{n}; t)}{\Delta t} = & d \sum_{i=1}^{K-1} [(n_i + 1)p(\mathcal{R}_i \vec{n}; t) - n_i p(\vec{n}; t)] \\ & + d \sum_{i=2}^K [(n_i + 1)p(\mathcal{L}_i \vec{n}; t) - n_i p(\vec{n}; t)]. \end{aligned}$$

If we take the limit where Δt tends to 0, we have the continuous-time master equation:

$$\begin{aligned} \frac{dp(\vec{n})}{dt} = & d \sum_{i=1}^{K-1} [(n_i + 1)p(\mathcal{R}_i \vec{n}) - n_i p(\vec{n})] \\ & + d \sum_{i=2}^K [(n_i + 1)p(\mathcal{L}_i \vec{n}) - n_i p(\vec{n})]. \end{aligned} \quad (\text{E.1})$$

The mean of the i^{th} compartment will be:

$$M_i(t) = \sum_{\vec{n}} n_i p(\vec{n}; t) \equiv \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_K=0}^{\infty} n_i p(\vec{n}; t), \quad (\text{E.2})$$

for each $i = 1, \dots, K$. The evolution equation of the mean can be obtained by multiplying Equation (E.1) and the correspondent reflective boundary conditions by n_i and summing over $\vec{n}$, as seen in Erban et al. [73] to arrive at the system of equations:

$$\frac{dM_1}{dt} = d(M_2 - M_1), \quad (\text{E.3})$$

$$\frac{dM_i}{dt} = d(M_{i-1} - 2M_i + M_{i+1}), \quad i = 2, \dots, K-1, \quad (\text{E.4})$$

$$\frac{dM_K}{dt} = d(M_{K-1} - M_K). \quad (\text{E.5})$$

Because the classical deterministic equation for diffusion is usually given in terms of concentration, we will approximate the concentration $c(x_i, t)$ as the mean over the compartment length: $c(x_i, t) \approx M_i(t)/h$. By dividing Equation (E.4) by h and doing a Taylor expansion on the right-hand side, we obtain:

$$\frac{\partial c(x_i, t)}{\partial t} \approx dh^2 \frac{\partial^2 c(x_i, t)}{\partial x^2}. \quad (\text{E.6})$$

This is equivalent to the classical deterministic equation for diffusion if we take $d = D/h^2$ as we have done in Chapter 4. The method commonly used to solve the diffusion equation in the continuum limit is the separation of variables. An analogous procedure would allow us to consider more species.

Appendix F

One-dimensional case of the Immediate Meeting Model

F.1 Parameter estimation

We will estimate realistic parameter values for the one-dimensional simulations of the Immediate Meeting Model. Parameter values for these simulations were taken from Pearce [2]. For clarity we expand on the origins of each parameter. Grove et al. [3, Table 5.01] (found in Appendix D) collected a dataset corresponding to 178 different groups of hunter-gatherers in different locations around the world, collected from different sources. This dataset contains information such as: the general biome (or climatic area) where the society lives, the area occupied by the population, the estimated total displacement in a year due to the migrations, the number of members in the population, the average number of members in a family, etc. From this information we can see that the mean area of occupation for a society is about 50,380 km². We can also see that 90% of the areas are under 160,000 km², we choose to use this larger area so as to keep the simulation closer to the lower bound on minimum families met in the worst case scenario. Because the area is

two-dimensional and we are working in one dimension, we fix our domain to be 400 km wide. For the one-dimensional case, we assumed a detection distance of 0.5 km.

Also, by dividing each population over the number of members in a family, then over the area of each society we can determine the density of families for each group. The mean for this number is 0.02 families per km². Instead of choosing a family number from the data table, we use this mean density of families obtained and multiply it by the chosen area (160,000 km²). This gives us 3,200 families in this space. However we have to take into account that this calculation is for a two-dimensional area, whereas we first work with a one-dimensional line. Taking the square root of the 3,200 families gives 56 families, which is fortunate since this means that an individual has the potential to meet 30 families, thus achieving its 150-individuals layer, the active network.

To calculate the diffusion parameter, D , we need the speed at which the hunter-gatherers move each day as they forage for food. Layton et al. [111] reports foraging distances from camp of 2 to 25 km/day: The Nukak population in tropical areas forages a maximum of 2 to 11 km/day and the hunters a range of 3 to 16 km/day [113, p 172-173]. This is just the distance away from camp, so that distance should approximately be doubled to 6 km/day, to keep within the lower limit. I erred and ran the stochastic simulations with 5.3 km/day as the speed. This will have to be improved, but works as a conservative limit. The one-dimensional diffusion coefficient can now be calculated as follows [148, 157]:

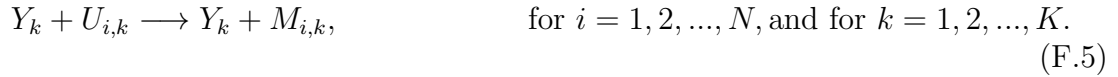
$$D = \frac{\Delta x^2}{2\Delta t} \tag{F.1}$$

where Δx is 5.3 km and Δt is a day. Thus $D = 14 \text{ km}^2/\text{day}$.

F.2 Simulations in one dimension

As seen in Section 4.2.1, not only will each family have an index identifying it, $i = 1, \dots, N$, but they will also have a location index $k = 1, \dots, K$, given a one-dimensional domain divided into K compartments. Throughout the simulation all of these families diffuse and the unmet species can convert to and from the met species. Thus we have $2N + 1$ families spread over K spatial compartments in one dimension. The algorithm will keep track of $K \times (2N + 1)$ different “populations”.

In total, the chemical reactions will be

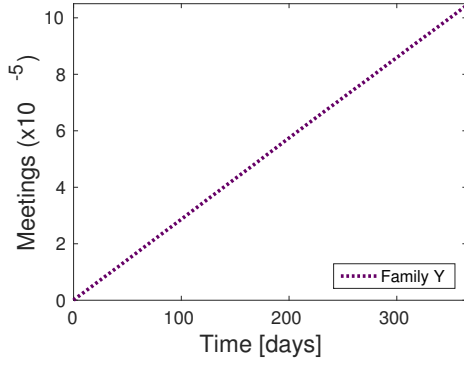


We coded these reactions into the Gillespie SSA and obtained position and number of meetings for each family every day for a year.

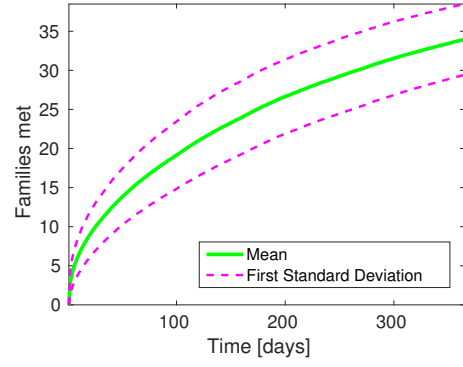
For our results we have three main graphs:

1. The total number of meetings throughout the year.
2. The total number of new meetings throughout the year.
3. The number of times each family was met by Y .

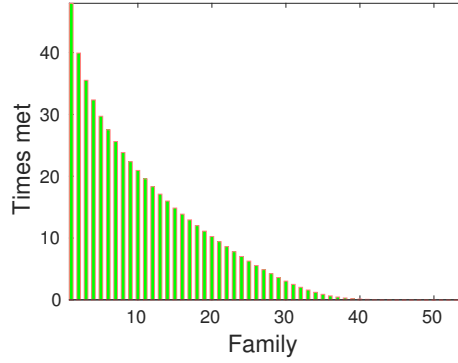
The first graph is always a straight line as long as the forgetting time is small compared to a year. This is as expected because, according to the ideal gas model [126], the number of encounters an individual experiences when moving randomly is directly proportional to the individual's speed, their detection distance, the density of individuals, and time. Since the former three entities are constant in our model, the number of total meetings will be directly proportional to time only, assuming memory plays a proportionally small role. The second graph is a measure of which families have been met throughout the year, counting only the first time Y met them. This graph tells us if our Y family has met 150 people and how long it took for them to meet. This graph can initially be approximated with a line, much like the ideal gas model, but its slope starts decreasing as soon as all nearby families have been met and family Y has to go further to find new families. The curve must asymptote to a maximum where all available families have been met, and this will happen sooner or later depending on the combination of all parameter values. For the third graph we order the met families from most met to least met in each run and then average over all the runs, obtaining an indicator of how many times a person in family Y will meet with each other person. This graph shows that a small number of families will be met several times, whilst most families will only be met a few times. These three graphs are illustrated in Figure F.1.



(a) Total number of meetings.



(b) Total number of new meetings.



(c) Number of times each family was met by Y at $t = 365$ days.

Figure F.1: Example solution from the Gillespie SSA for 1D for the model described in 4.2.2, IMM, average of 2,000 runs. Parameter values used were: 54 unmet families initially, forgetting time of a day, length of 400 km, detection distance of 0.5 km, and diffusion coefficient of $9 \text{ km}^2/\text{day}$.

F.3 One-dimensional results analysis

Parameter values for this section were estimated as explained in Appendix F.1. Note that, presently, we assume a detection distance of 0.5 km.

After 2,000 simulations for 54 families in a 400 km domain with a speed of 5.3 km/day, we can see from Figure F.2 that the number of families family Y meets over the year is given approximately by:

$$\text{Families met} = 1.6t^{0.5}. \quad (\text{F.6})$$

This power law was approximated using the MATLAB curve-adjusting package.

Note, however, that the simulation must be bounded because of the finite number of families available to meet. Also, we can see from Figure F.2 that family Y has clearly met the 15-individuals layer (by having met more than 3 families) within the first month. The 50-individual layer is met within the second month, and the 150-individual layer is approximately covered by the end of the year. This means it is plausible that an individual can meet the whole of his/her active network within a year by merely foraging randomly for this set of parameter values in one dimension.

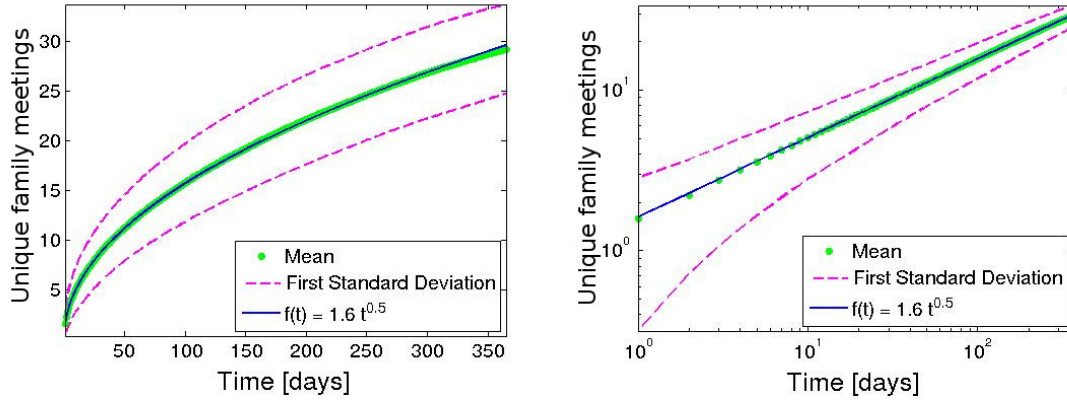


Figure F.2: The dotted green line represents the mean number of different families the family Y has met throughout the year. The dashed magenta lines represent ± 1 standard deviation. The solid blue line is the power-law approximation obtained using MATLAB's curve-fitting package. Parameter values are 54 families in a 400 km domain with a speed of 5.3 km/day, and averaged from 2,000 realisations.

Figure F.3 shows, for each day, how many families family Y met for the very first time. On average, family Y meets more than 1 new family on day one, but this figure quickly drops as a power law. The fact that, on day one, the number of new families met is bigger than 1 depends on the density of U families initially. In at least one of the realisations, family Y met 8 unmet families in day 1. This average

falls below 1 new meeting for the second day. This behaviour can be explained by the fact that at the start of the realisation every family around Y is a family never met before by Y , but as the families nearby are met, family Y has to go much further to meet new families.

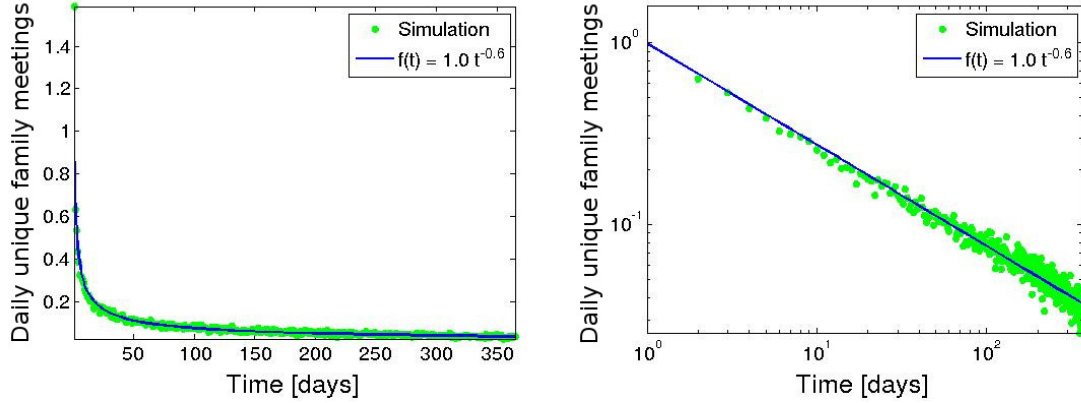


Figure F.3: Number of new families that family Y meets each day as an average over 2,000 realisations. These are families family Y had not met before in the year and so represents the number of uniquely met families each day. The solid blue line is the approximation obtained using MATLAB’s curve-fitting package. Parameter values are 54 families in a 400 km domain with a speed of 5.3 km/day.

From Figure F.3 we obtain an estimating function that could be used to predict how many days it would take to meet further layers in the social organization. For example, it would take 2,963 days, or 8.1 years, before an individual could meet the whole 500-individual layer just by diffusion in one dimension.

In Figure F.4 we rank the families in terms of their number of meetings. Explicitly, for each realisation we sort the families in order of how many times they have met Y . A number is given to each family in the realisation, ‘Family 1’ being the most met, and ‘Family 53’ the least met. Having done this for every realisation, we average over all simulations. This result shows that all the personal network layers

are achieved right up to the active network of 150 individuals. The result also tells us how many times family Y meets with these different layers.

Since we assume that a family is 5 people we satisfy the clique social layer meetings on a daily basis. To achieve the number of meetings needed for the sympathy group, the 15-individual layer, each member of family Y has to meet 2 other families. Figure F.4 suggests Y can see 2 other families at least 48 times in a year, that is a little less than once a week. The affiliation group (or residential band) is covered when family Y meets 9 other families, and from our results, these are met at least 25 times in a year, which is equivalent to twice a month. The final layer covered corresponds to the active network, of 150 individuals, and so family Y would have to meet 29 other families for this layer to be covered. Figure F.4 not only suggests that this is possible but that all 150 individuals can be met at least twice a year. This could correspond to special occasions such as solstices for primitive humans, or Christmas and birthdays for modern day humans.

Family Y meets up to 31 families at least once a year on average, but the remaining 22 families are met less than once a year. From those remaining 22 families, 7 families are not met at all in our 2,000 simulations.

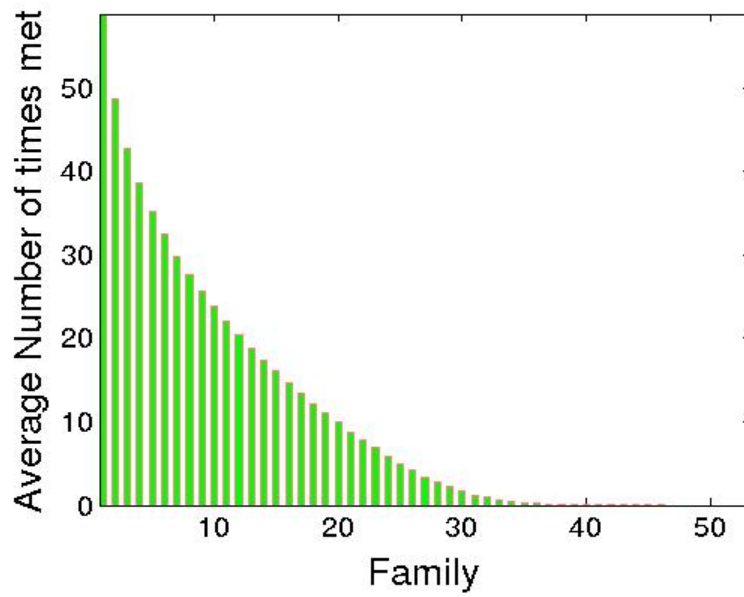


Figure F.4: Ordered by average number of times they meet, we see here how many times family Y meets with other families throughout the year. Parameter values are 54 families in a 400 km domain, and 2,000 realisations.

These results are for a one-dimensional space, which does not represent most of the hunter-gatherers we want to describe, but it could represent the movement of societies that live along a riverbank, a mountain range, or a particular marked path.

Appendix G

Parameter estimation for all the biome-specific cases of the Immediate Meeting Model

We calculate here the parameter values used for the application of the IMM to each biome, in Section 4.4.1.1, to obtain an estimate of maximum and minimum boundaries (which we call BiomeMax and BiomeMin respectively) and a possible intermediate case (which we name BiomeMid) of number of families met in a year. For the BiomeMax number of families met we used the largest family size in the biome, the fastest speed, the biggest concentration of people, and the minimum area. For the BiomeMin, the opposite extremes were chosen. Whereas for the BiomeMid case, a mean of the given family sizes is used (set to 5 when data was not available), an average of given areas of occupancy, an average of calculated population concentration, and a speed in between maximum and minimum values are used. The speed was obtained as explained in the next paragraph, but a division over 3 was sometimes used to achieve the value in between the extremes.

The distance in a day to use for the diffusion coefficient was found using a combination of Layton et al. [111, p 1230], Table IV, and Grove et al. [3, Table

5.01], (as seen in Appendix D). For each biome the distance walked was obtained as follows:

- For Arctic: Maximum and minimum were taken from Grove et al. [3]. For the BiomeMid case we took the total displacement over the number of movements, divided by 3, and rounded to the closest integer.
- For Australia: Maximum and minimum were taken from Layton et al. [111], taking hunters from the Nukak population in ‘Tropical’ for maximum and foragers for minimum. For the BiomeMid case, using Grove et al. [3], we took the total displacement over the number of movements, divided by 3, and rounded to the closest integer. Data where the total displacement in a year was zero was not taken into account.
- For North America: Minimum and BiomeMid cases were taken from Grove et al. [3], calculating total displacement over the number of times moved in a year, divided by 3, and rounded to the closest integer. Data where the total displacement in a year was zero was not taken into account. Maximum was taken from Layton et al. [111] for the ‘semi-desert’.
- For Subarctic and continental midlatitude forests of North America and Asia: Minimum and BiomeMid cases were taken from Grove et al. [3], calculating total displacement over the number of times moved in a year, divided by 3, and rounded to the closest integer. Again, data where the total displacement in a year was zero was not taken into account. Maximum was taken from Layton et al. [111], for ‘Boreal forest’.

- For tropical and subtropical Africa: Maximum and BiomeMid cases were both the same, the Aka population in ‘Tropical’, from Layton et al. [111]. Minimum was taken from Grove et al. [3], as the total displacement in a year over the number of times moved, divided by 3, and rounded to the closest integer. Data where the total displacement in a year was zero was not taken into account.
- For tropical and subtropical Asia and tropical and subtropical South America: data was taken from Layton et al. [111]. We used hunters from the Nukak population in ‘Tropical’ for the maximum, foragers for the minimum, and the average of both for the BiomeMid case.

For the number of initially unmet families, we divided the population of each data entry in Grove et al. [3] over its mean family size (or over 5 when the mean family size was not available), then rounded to the closest integer. The maximum for each biome was used for the BiomeMax case, the minimum of each biome was used for the BiomeMin case, and an average was taken for the BiomeMid case. Horizontal and vertical lengths were calculated as the square root of the total area, then rounded to the closest integer.

Appendix H

Results for BiomeMin, MiomeMid and BiomeMax cases for each biome, Immediate Meeting Model.

This appendix contains more example results of the work done in Section 4.4.1.1. Results of the total number of encounters in a year, the number of new families met by the end of the year, the number of times each family was met by Y , and 9 example paths are shown for the Arctic, whilst only the first three results are shown for Australia, the subarctic and continental midlatitude forests of North America and Asia, tropical and subtropical Africa, and tropical and subtropical South America. More examples of paths are not shown because of the similitude with the ones in the Arctic and the low information content.

H.1 Arctic

Results are given here for the BiomeMin, BiomeMid and BiomeMax cases for the Arctic biome. A two-dimensional IMM with an average of 4.4 people per family and an encounter distance of 1 km was used. For the BiomeMin case we had 25 initially unmet families in an area of 370,881 km², and a diffusion coefficient of 9

km^2/day , with 100 simulations. For the BiomeMid case we had 318 initially unmet families in an area of $81,225 \text{ km}^2$, and a diffusion coefficient of $25.8 \text{ km}^2/\text{day}$, with 100 simulations too. For the BiomeMax case we had 1,595 initially unmet families in an area of $18,769 \text{ km}^2$, and a diffusion coefficient of $132.3 \text{ km}^2/\text{day}$, with 26 simulations.

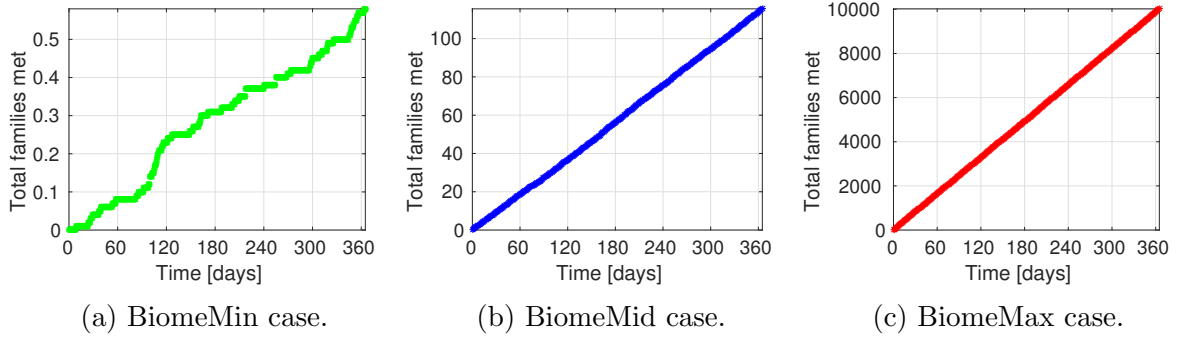


Figure H.1: Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case had 1 meeting, BiomeMid case had 116 meetings, and BiomeMax case had 10,021 meetings in average by the end of the year.

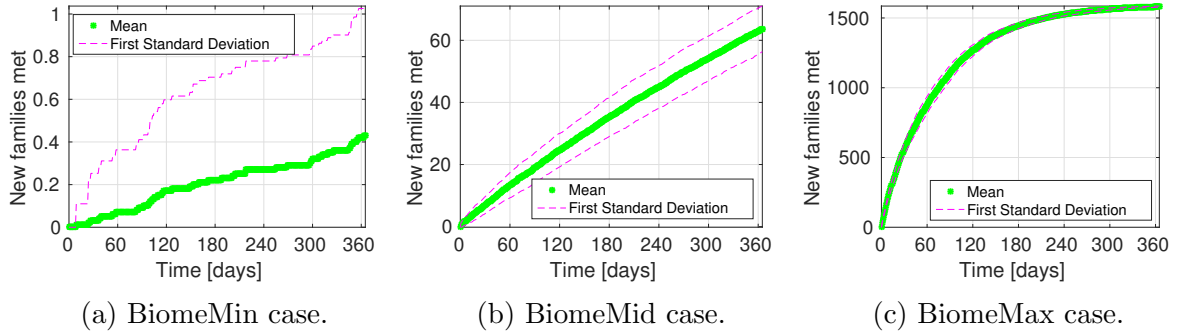


Figure H.2: Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case met 0.43 new families, BiomeMid case met 63.69 new families, and BiomeMax case met 1,581.3 new families in average by the end of the year.

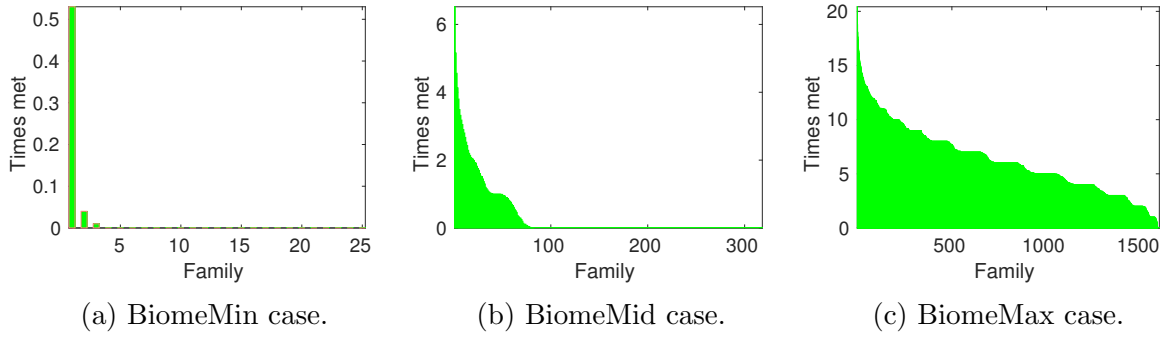


Figure H.3: Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. The most met family was met 0.53 times in the BiomeMin case, 6.52 times in the BiomeMid case, and 20.42 times in the BiomeMax case in average by the end of the year.

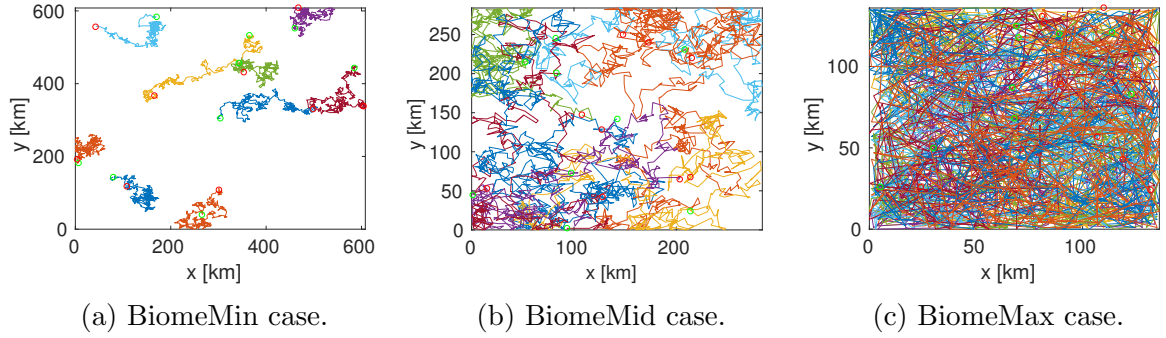


Figure H.4: Example of 9 paths for each of the three cases for the Arctic biome. See text for parameter values.

H.2 Australia: subtropical desert through temperate settings

Results are given here for the BiomeMin, BiomeMid and BiomeMax cases for the Australia biome, where ‘subtropical desert through temperate settings’ means the biome here considered spans climates from subtropical deserts through to mild climates. A two-dimensional IMM with an average of 6.3 people per family and an encounter distance of 1 km was used. For the BiomeMin case we had 6 initially

unmet families in an area of 137,641 km², and a diffusion coefficient of 1 km²/day, with 100 simulations. For the BiomeMid case we had 167 initially unmet families in an area of 14,884 km², and a diffusion coefficient of 5.8 km²/day, with 100 simulations too. For the BiomeMax case we had 718 initially unmet families in an area of 81 km², and a diffusion coefficient of 64 km²/day, with 55 simulations.

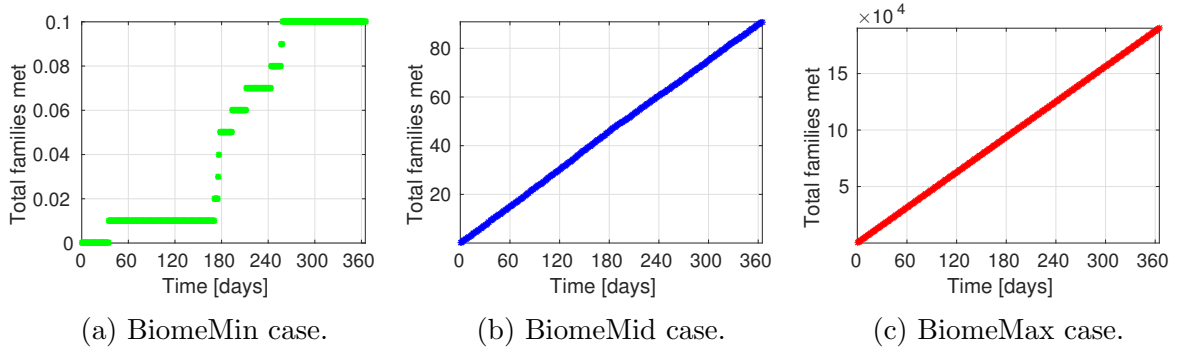


Figure H.5: Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the Arctic. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 91 meetings, and BiomeMax case had 190,158 meetings in average by the end of the year.

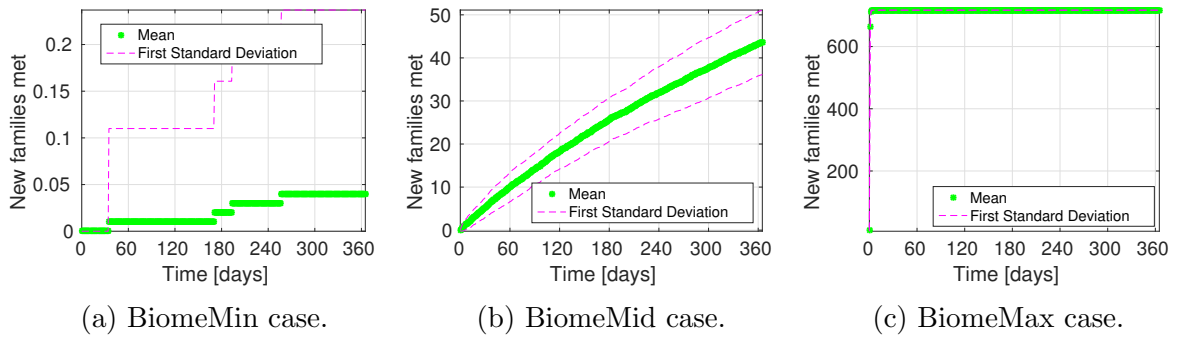


Figure H.6: Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for Australia. See text for parameter values. BiomeMin case met 0.40 new families, BiomeMid case met 43.64 new families, and BiomeMax case met 717 new families in average by the end of the year.

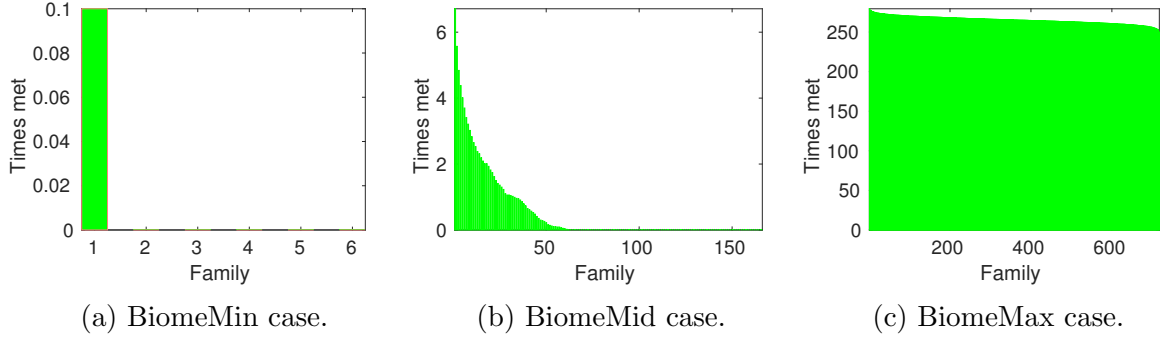


Figure H.7: Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for Australia. See text for parameter values. The most met family was met 0.10 times in the BiomeMin case, 6.71 times in the BiomeMid case, and 279.38 times in the BiomeMax case in average by the end of the year.

H.3 Subarctic and continental midlatitude forests of North America and Asia

Results are given here for the BiomeMin, BiomeMid and BiomeMax cases for the subarctic and continental midlatitude forests of North America and Asia. A two-dimensional IMM with an average of 4.4 people per family and an encounter distance of 1 km was used. For the BiomeMin case we had 26 initially unmet families in an area of 660,156 km², and a diffusion coefficient of 3.8 km²/day, with 100 simulations. For the BiomeMid case we had 308 initially unmet families in an area of 93,330 km², and a diffusion coefficient of 17 km²/day, with 100 simulations too. For the BiomeMax case we had 1,066 initially unmet families in an area of 5,402 km², and a diffusion coefficient of 56.3 km²/day, with 15 simulations.

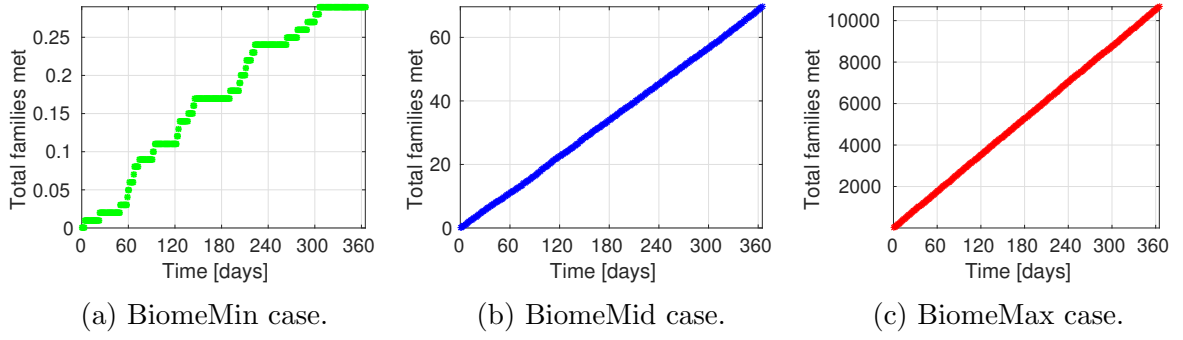


Figure H.8: Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 70 meetings, and BiomeMax case had 10,668 meetings in average by the end of the year.

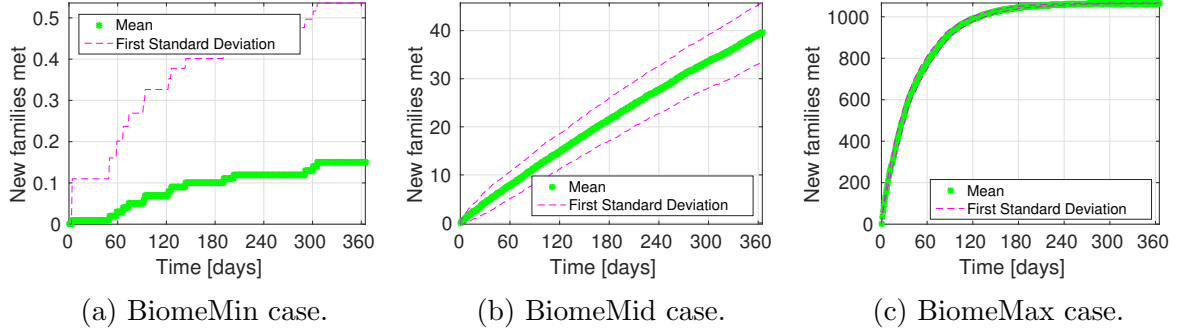


Figure H.9: Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. BiomeMin case met 0.15 new families, BiomeMid case met 39.64 new families, and BiomeMax case met 1,065.4 new families in average by the end of the year.

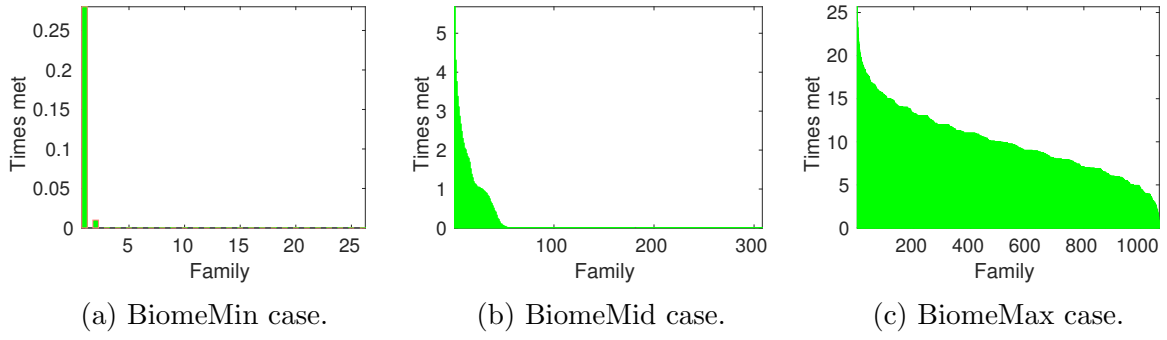


Figure H.10: Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for subarctic and continental midlatitude forests of North America and Asia. See text for parameter values. The most met family was met 0.28 times in the BiomeMin case, 5.58 times in the BiomeMid case, and 25.67 times in the BiomeMax case in average by the end of the year.

H.4 Tropical and subtropical Africa

Results are given here for the BiomeMin, BiomeMid and BiomeMax cases for the tropical and subtropical Africa biome. A two-dimensional IMM with an average of 4.2 people per family and an encounter distance of 1 km was used. For the BiomeMin case we had 23 initially unmet families in an area of 57,121 km², and a diffusion coefficient of 3.6 km²/day, with 100 simulations. For the BiomeMid case we had 90 initially unmet families in an area of 13,456 km², and a diffusion coefficient of 132.3 km²/day, with 100 simulations too. For the BiomeMax case we had 180 initially unmet families in an area of 2,500 km², and a diffusion coefficient of 132.3 km²/day, with 89 simulations.

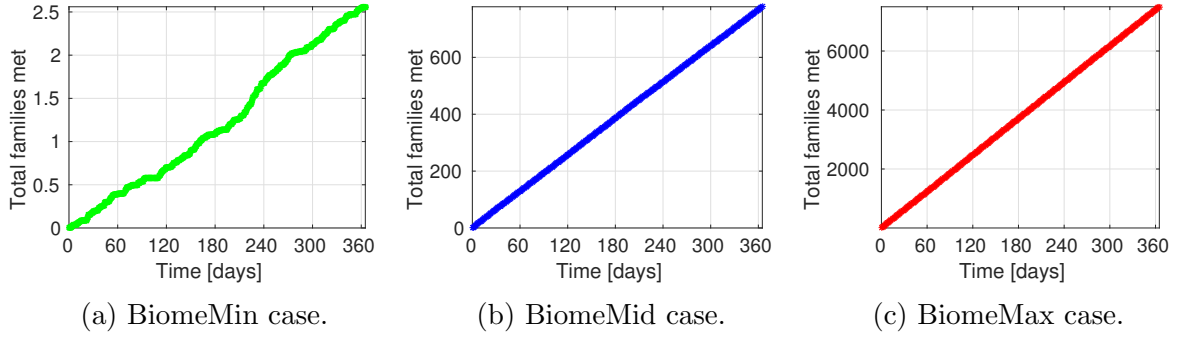


Figure H.11: Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the tropical and subtropical Africa. See text for parameter values. BiomeMin case had 3 meetings, BiomeMid case had 778 meetings, and BiomeMax case had 7, 500 meetings in average by the end of the year.

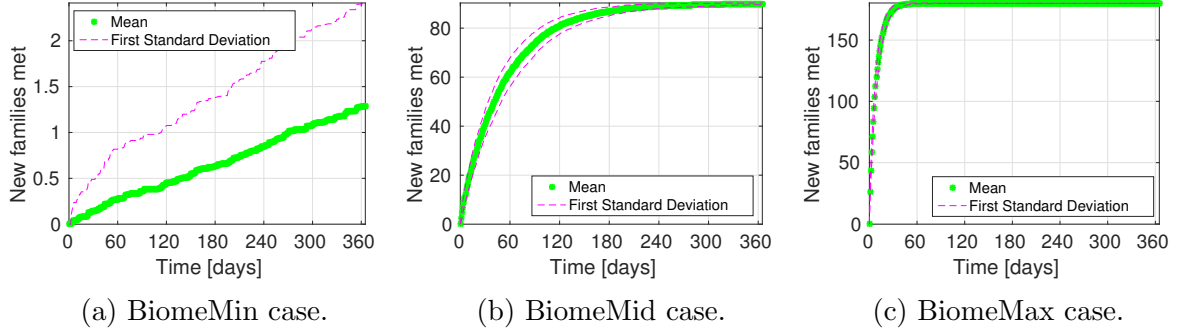


Figure H.12: Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical Africa. See text for parameter values. BiomeMin case met 1.29 new families, BiomeMid case met 89.89 new families, and BiomeMax case met 180 new families in average by the end of the year.

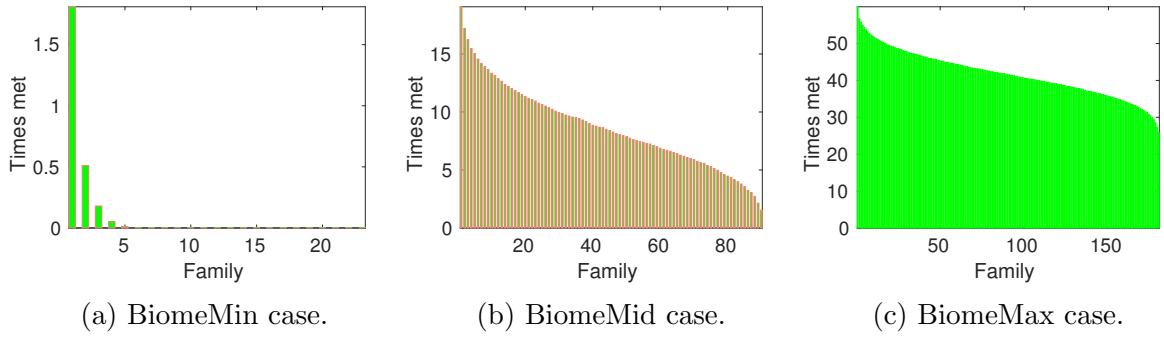


Figure H.13: Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical Africa. See text for parameter values. The most met family was met 1.81 times in the BiomeMin case, 19.07 times in the BiomeMid case, and 59.97 times in the BiomeMax case in average by the end of the year.

H.5 Tropical and subtropical South America

Results are given here for the BiomeMin, BiomeMid and BiomeMax cases for the tropical and subtropical South America biome. A two-dimensional IMM with an average of 4.3 people per family and an encounter distance of 1 km was used. For the BiomeMin case we had 5 initially unmet families in an area of 476,100 km², and a diffusion coefficient of 1 km²/day, with 100 simulations. For the BiomeMid case we had 553 initially unmet families in an area of 30,276 km², and a diffusion coefficient of 15 km²/day, with 100 simulations too. For the BiomeMax case we had 1,799 initially unmet families in an area of 324 km², and a diffusion coefficient of 64 km²/day, with 19 simulations.

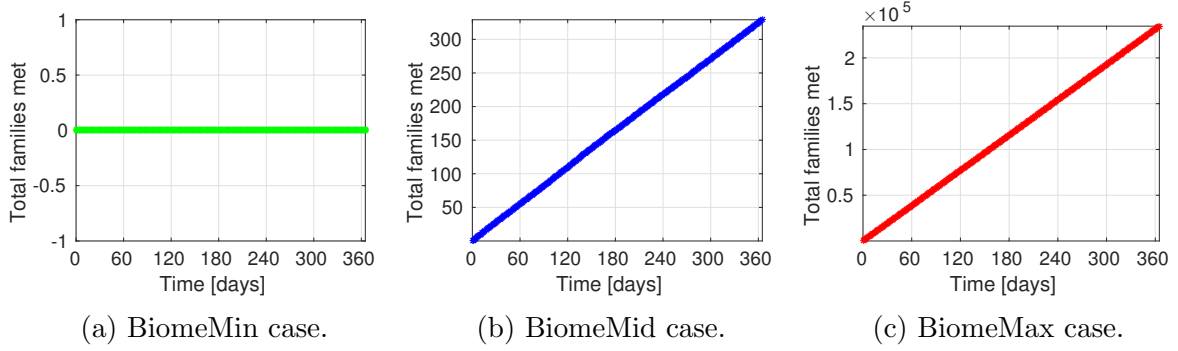


Figure H.14: Total number of families met, regardless of new or previously met, over the year for the BiomeMin, BiomeMid, and BiomeMax case for the tropical and subtropical South America. See text for parameter values. BiomeMin case had 0 meetings, BiomeMid case had 329 meetings, and BiomeMax case had 234,720 meetings in average by the end of the year.

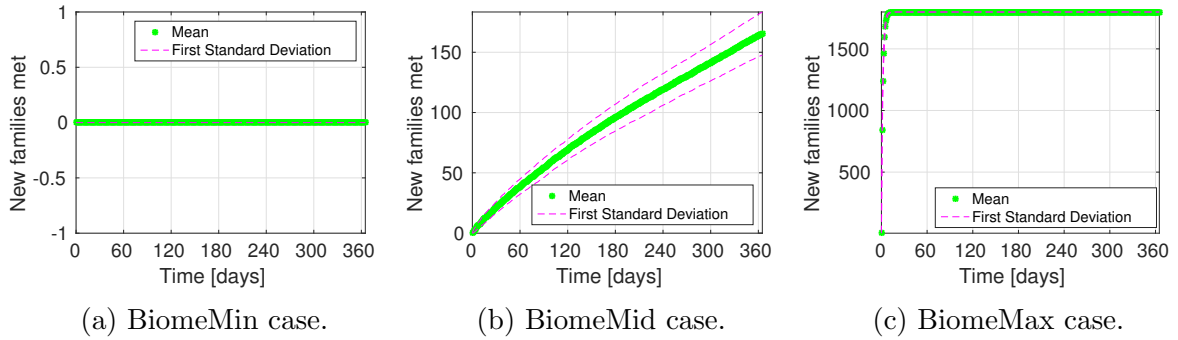


Figure H.15: Number of new families met over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical South America. See text for parameter values. BiomeMin case met 0 new families, BiomeMid case met 165.48 new families, and BiomeMax case met 1,799 new families in average by the end of the year.

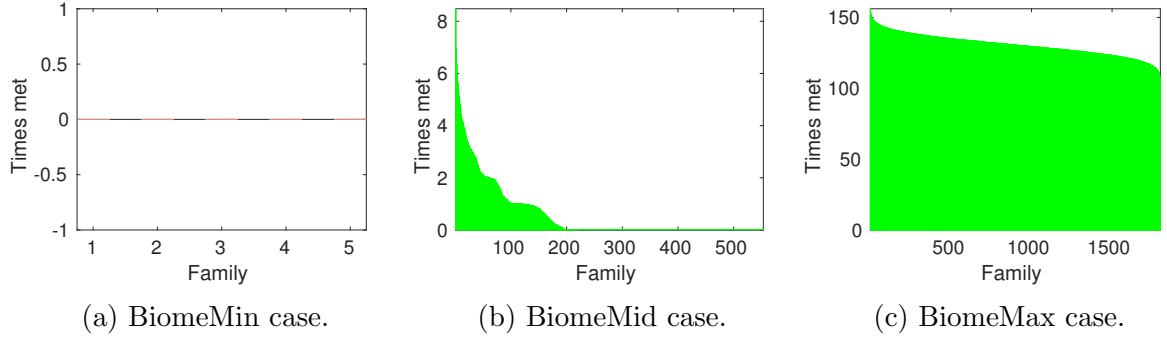


Figure H.16: Number times each family was met by family Y over the year for the BiomeMin, BiomeMid, and BiomeMax case for tropical and subtropical South America. See text for parameter values. The most met family was met 0 times in the BiomeMin case, 8.48 times in the BiomeMid case, and 156.11 times in the BiomeMax case in average by the end of the year.

Bibliography

- [1] J. Goyette, C. Solís Salas, N. Coker-Gordon, M. Bridge, S. A. Isaacson, J. Al-lard, and O. Dushek. Biophysical assay for tethered signaling reactions reveals tether-controlled activity for the phosphatase SHP-1. *Science Advances*, 3(3), 2017.
- [2] E. Pearce. *The Effects of Latitude on Hominin Social Network Maintenance*. DPhil Thesis, University of Oxford, 2013.
- [3] M. Grove, E. Pearce, and R. I. M. Dunbar. Fission-fusion and the evolution of hominin social systems. *Journal of Human Evolution*, 62(2):191–200, 2012.
- [4] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contribu-tions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.
- [5] S. S. Taylor, I. G. Jamieson, and D. P. Armstrong. Successful island rein-troductions of new zealand robins and saddlebacks with small numbers of founders. *Animal Conservation*, 8(4):415–420, 2005.
- [6] J. N. Carey, E. L. Metttert, M. Roggiani, K. S. Myers, P. J. Kiley, and M. Gou-lian. Regulated stochasticity in a bacterial signaling network permits tolerance to a rapid environmental change. *Cell*, 173(1):196 – 207.e14, 2018.
- [7] R. Baker. Stochastic modelling of biological processes, Lecture Notes. Tech-nical report, University of Oxford, 2017.
- [8] O. Acuto, V. D. Bartolo, and F. Michel. Tailoring T-cell receptor signals by proximal negative feedback mechanisms. *Nature Reviews Immunology*, 8(9): 699–712, 2008.
- [9] A. Goldbeter. A model for circadian oscillations in the *Drosophila* period protein (PER). *Proceedings of the Royal Society B*, 261(1362):319–324, 1995.
- [10] A. Goldbeter and M. J. Berridge. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cam-bridge University Press, 1997.
- [11] V. Ajraldi, M. Pittavino, and E. Venturino. Modeling herd behavior in pop-ulation systems. *Nonlinear Analysis: Real World Applications*, 12(4):2319 – 2338, 2011.

- [12] J. Roy, D. Barman, and S. Alam. Role of fear in a predator-prey system with ratio-dependent functional response in deterministic and stochastic environment. *Biosystems*, 197:104176, 2020.
- [13] O. Wanner and P. Reichert. Mathematical modeling of mixed-culture biofilms. *Biotechnology and Bioengineering*, 49(2):172–184, 1996.
- [14] L. S. Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, 2014.
- [15] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.
- [16] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.
- [17] S. Kar, W. T. Baumann, M. R. Paul, and J. J. Tyson. Exploring the roles of noise in the eukaryotic cell cycle. *Proceedings of the National Academy of Sciences*, 106(16):6471–6476, 2009.
- [18] G. Gardiner. *Handbook of Stochastic Processes for Physics, Chemistry and Natural Sciences*. Springer Verlag, 2nd edition, 1985.
- [19] N. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier Science, 1992.
- [20] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54:1–26, 2007.
- [21] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- [22] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.
- [23] Y. Cao, H. Li, and L. Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The Journal of Chemical Physics*, 2004.
- [24] C. Dittamo and D. Cangelosi. Optimized parallel implementation of gillespie’s first reaction method on graphics processing units. In *2009 International Conference on Computer Modeling and Simulation*, pages 156–161, 2009.
- [25] G. Klingbeil, R. Erban, M. Giles, and P. K. Maini. Fat versus thin threading approach on gpus: Application to stochastic simulation of chemical reactions. *IEEE Transactions on Parallel and Distributed Systems*, 23(2):280–287, 2012.

- [26] G. Klingbeil, R. Erban, M. Giles, and P. K. Maini. STOCHSIMGPU: parallel stochastic simulation for the Systems Biology Toolbox 2 for MATLAB. *Bioinformatics*, 27(8):1170–1171, 2011.
- [27] D. L. DeAngelis and S. Yurek. Spatially explicit modeling in ecology: A review. *Ecosystems*, 20(2):284–300, 2017.
- [28] S. D. Fretwell. On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica*, 19:45–52, 1969.
- [29] R. S. Cantrell and C. Cosner. Spatial heterogeneity and critical patch size: Area effects via diffusion in closed environments. *Journal of Theoretical Biology*, 209(2):161 – 171, 2001.
- [30] J. Skellam. Random dispersal in theoretical populations. *Bulletin of Mathematical Biology*, 53(1):135 – 165, 1991.
- [31] A. Hastings. Models of spatial spread: A synthesis. *Biological Conservation*, 78(1):143 – 148, 1996. Invasion Biology.
- [32] J. Roughgarden. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*, Chapter 13: Production functions from ecological populations: a survey with emphasis on spatially implicit models, pages 296–316. Princeton University Press, 1997.
- [33] J. K. Actor. *Introductory Immunology: Basic Concepts for Interdisciplinary Applications*. Elsevier Science, 2nd edition, 2019.
- [34] Immunology applications, 2021. URL `\url{https://chem.libretexts.org/@go/page/8706}`.
- [35] J. B. Bassingthwaite, H. J. Chizeck, L. E. Atlas, and H. Qian. Multiscale modeling of cardiac cellular energetics. *Annals of the New York Academy of Sciences*, 1047:395–424, 2005.
- [36] Y. Choe. The role of temporal parameters in a thalamocortical model of analogy. *IEEE Transactions on Neural Networks*, 15(5):1071–1082, 2004.
- [37] M. Breakspear and C. J. Stam. Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society B*, 360: 1051–1074, 2005.
- [38] T. Alarcón, H. M. Byrne, and P. K. Maini. Towards whole-organ modelling of tumour growth. *Progress in Biophysics and Molecular Biology*, 85(2):451 – 472, 2004.
- [39] S. M. Peirce and T. C. Skalak. Microvascular remodeling: A complex continuum spanning angiogenesis to arteriogenesis. *Microcirculation*, 10(1):99–111, 2003.

- [40] S. McDougall, J. Dallon, J. Sherratt, and P. K. Maini. Fibroblast migration and collagen deposition during dermal wound healing: mathematical modelling and clinical implications. *Philosophical Transactions of the Royal Society A*, 364(1843):1385–1405, 2006.
- [41] B. D. Cumming, D. L. S. McElwain, and Z. Upton. A mathematical model of wound healing and subsequent scarring. *Journal of the Royal Society Interface*, 7(42):19–34, 2010.
- [42] T. Sun, S. Adra, R. Smallwood, M. Holcombe, and S. MacNeil. Exploring hypotheses of the actions of TGF- β 1 in epidermal wound healing using a 3D computational multiscale model of the human epidermis. *PLoS ONE*, 4(12):e8515, 2009.
- [43] W. A. Comisar, S. X. Hsiong, H.-J. Kong, D. J. Mooney, and J. J. Linderman. Multi-scale modeling to predict ligand presentation within rgd nanopatterned hydrogels. *Biomaterials*, 27(10):2322 – 2329, 2006.
- [44] A. Cappuccio, P. Tieri, and F. Castiglione. Multiscale modelling in immunology: a review. *Briefings in Bioinformatics*, 17(3):408–418, 2015.
- [45] G. Gaud, R. Lesourne, and P. E. Love. Regulatory mechanisms in T cell receptor signalling. *Nature Reviews Immunology*, 18:485–497, 2018.
- [46] J. Tang, H.-Y. Yin, and J.-L. Zhang. Chapter 1: Luminescent zinc complexes as bioprobes for imaging molecular events in live cells. In K. K.-W. Lo, editor, *Inorganic and Organometallic Transition Metal Complexes with Biological Molecules and Living Cells*, pages 1 – 53. Academic Press, 2017.
- [47] N. Jonniya. Investigating phosphorylation-induced conformational changes in WNK1 kinase by molecular dynamics simulations. *ACS Omega*, 4(17):17404–17416, 2019.
- [48] F. Bornancin and P. J. Parker. Phosphorylation of threonine 638 critically controls the dephosphorylation and inactivation of protein kinase α . *Current Biology*, 6(9):1114 – 1123, 1996.
- [49] A. Enjalbert and C. Le Pechon-Vallee. Protein kinases. In H. L. Henry and A. W. Norman, editors, *Encyclopedia of Hormones*, pages 277 – 285. Academic Press, New York, 2003.
- [50] R. P. Bhattacharyya, A. Reményi, B. J. Yeh, and W. A. Lim. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual Review of Biochemistry*, 75(1):655–680, 2006. PMID: 16756506.
- [51] A. S. Shaw and E. L. Filbert. Scaffold proteins and immune-cell signalling. *Nature Reviews Immunology*, 9:47–56, 2009.

- [52] C. J. Bashor, N. C. Helman, S. Yan, and W. A. Lim. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science*, 319(5869):1539–1543, 2008.
- [53] A. P. Won, J. E. Garbarino, and W. A. Lim. Recruitment interactions can override catalytic interactions in determining the functional identity of a protein kinase. *Proceedings of the National Academy of Sciences*, 108(24):9809–9814, 2011.
- [54] Y. Zhang, L. Clemens, J. Goyette, J. Allard, O. Dushek, and S. A. Isaacson. The influence of molecular reach and diffusivity on the efficacy of membrane-confined reactions. *Biophysical Journal*, 117(7):1189 – 1201, 2019.
- [55] N. S. Hopkins, Ákos Östör, F. Ghannam, E. A. Smith, R. A. Paul, C. F. Keyes, G. Feeley-Harnik, J. H. Hill, U. Hannerz, R. H. Tuttle, M. J. Aronoff, A. J. Kuper, P. Jeganathan, D. K. Robotham, C. L. Delaney, S. L. Hanchett, A. L. Kolata, C. Lomnitz, M. M. Horowitz, C. M. Leslie, J. R. Spencer, E. Schildkrout, A. Seeger, P. V. Esterik, H. Varenne, J. W. Fernandez, and R. W. Nicholas. Anthropology. Encyclopædia Britannica, 2020. URL `\url{https://www.britannica.com/science/anthropology}`.
- [56] C. McCourt, editor. *Anthropology in Action*, volume 27. Berghahn Journals, 2020.
- [57] R. I. M. Dunbar. The anatomy of friendship. *Trends in Cognitive Sciences*, 22(1):32 – 51, 2018.
- [58] C. Crockford, R. M. Wittig, P. L. Whitten, R. M. Seyfarth, and D. L. Cheney. Social stressors and coping mechanisms in wild female baboons (*papio hamadryas ursinus*). *Hormones and Behavior*, 53(1):254 – 265, 2008.
- [59] R. M. Wittig, C. Crockford, J. Lehmann, P. L. Whitten, R. M. Seyfarth, and D. L. Cheney. Focused grooming networks and stress alleviation in wild female baboons. *Hormones and Behavior*, 54(1):170 – 177, 2008.
- [60] J. B. Silk, S. C. Alberts, and J. Altmann. Social bonds of female baboons enhance infant survival. *Science*, 302(5648):1231–1234, 2003.
- [61] J. B. Silk, J. C. Beehner, T. J. Bergman, C. Crockford, A. L. Engh, L. R. Moscovice, R. M. Wittig, R. M. Seyfarth, and D. L. Cheney. Strong and consistent social bonds enhance the longevity of female baboons. *Current Biology*, 20(15):1359 – 1361, 2010.
- [62] E. Z. Cameron, T. H. Setsaas, and W. L. Linklater. Social bonds between unrelated females increase reproductive success in feral horses. *Proceedings of the National Academy of Sciences*, 106(33):13850–13853, 2009.

- [63] C. M. Nuñez, J. S. Adelman, and D. I. Rubenstein. Sociality increases juvenile survival after a catastrophic event in the feral horse (*Equus caballus*). *Behavioral Ecology*, 26(1):138–147, 2014.
- [64] R. I. M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493, 1992.
- [65] R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–694, 1993.
- [66] R. A. Hill and R. I. M. Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, 2003.
- [67] C. R. Weston and R. J. Davis. Signaling specificity– a complex affair. *Science*, 292(5526):2439–2440, 2001.
- [68] T. Yokosuka, M. Takamatsu, W. Kobayashi-Imanishi, A. Hashimoto-Tane, M. Azuma, and T. Saito. Programmed cell death 1 forms negative costimulatory microclusters that directly inhibit T cell receptor signaling by recruiting phosphatase SHP2. *Journal of Experimental Medicine*, 209(6):1201–1217, 2012.
- [69] J. Zhang, A. K. Somani, and K. A. Siminovitch. Roles of the SHP-1 tyrosine phosphatase in the negative regulation of cell signalling. *Seminars in immunology*, 12(4):361–78, 2000.
- [70] P. Pattnaik. Surface plasmon resonance. *Applied Biochemistry and Biotechnology*, 126(2):79–92, 2005.
- [71] C. T. Campbell and G. Kim. SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. *Biomaterials*, 28(15):2380 – 2392, 2007. Imaging Techniques for Biomaterials Characterization.
- [72] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404 – 425, 1992.
- [73] R. Erban, J. Chapman, and P. K. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. Technical report, University of Oxford, 2007.
- [74] P. J. Flory. *Principles of Polymer Chemistry*. Baker lectures 1948. Cornell University Press, 1953.
- [75] H. Yamakawa. *Modern Theory of Polymer Solutions*. Harper’s chemistry series. Harper & Row, 1971.
- [76] J. Mai, V. N. Kuzovkov, and W. von Niessen. A theoretical stochastic model for the $A + 1/2B_2 \rightarrow 0$ reaction. *The Journal of Chemical Physics*, 98(12):10017, 1993.

- [77] H. Schnörer, V. Kuzovkov, and A. Blumen. Segregation in annihilation reactions without diffusion: Analysis of correlations. *Physical Review Letters*, 63(7):805–808, 1989.
- [78] J. Mai, V. N. Kuzovkov, and W. von Niessen. Stochastic model for the A+B2 surface reaction: Island formation and complete segregation. *The Journal of Chemical Physics*, 100(8):6073, 1994.
- [79] J. Mai, V. Kuzovkov, and W. von Niessen. A general stochastic model for the description of surface reaction systems. *Physica A: Statistical Mechanics and its Applications*, 203(2):298–315, 1994.
- [80] J. C. Crocker and E. R. Weeks. Particle tracking using IDL, 2008. URL `\url{http://www.physics.emory.edu/faculty/weeks/idl/gofr.html}`.
- [81] V. Kuzovkov and E. Kotomin. Kinetics of bimolecular reactions in condensed media: critical phenomena and microscopic self-organisation. *Reports on Progress in Physics*, 1479, 1988.
- [82] E. Kotomin and V. Kuzovkov. Some problems of recombination kinetics I. *Chemical Physics*, 76:479–487, 1983.
- [83] R. Balescu. Equilibrium and nonequilibrium statistical mechanics. *NASA STI/Recon Technical Report A*, 76:32809, 1975.
- [84] E. Kotomin, V. Kuzovkov, W. Frank, and A. Seeger. Reaction kinetics beyond rate equations: a correlation-function study of the effects of space dimension and reactant mobilities on the bimolecular annihilation reaction. *Journal of Physics A: Mathematical and General*, 27(5):1453–1462, 1994.
- [85] H. Schnörer, V. Kuzovkov, and A. Blumen. Bimolecular annihilation reactions with immobile reactants. *The Journal of Chemical Physics*, 92(4):2310, 1990.
- [86] K. Dettmann, G. Leibfried, and K. Schroeder. Spontaneous recombination of frenkel pairs for electron irradiation. *physica status solidi (b)*, 22(2):423–432, 1967.
- [87] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300, 1935.
- [88] A. Singer. Maximum entropy formulation of the Kirkwood superposition approximation. *The Journal of Chemical Physics*, 121(8):3657–66, 2004.
- [89] T. F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6(2):418–445, 1994.
- [90] T. F. Coleman and Y. Li. On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds. *Mathematical Programming*, 67(1):189–224, 1994.

- [91] L. Clemens, M. Kutuzov, K. V. Bayer, J. Goyette, J. Allard, and O. Dushek. Determination of the molecular reach of the protein tyrosine phosphatase SHP-1. *bioRxiv*, 2020.
- [92] Q. Yan, T. Barros, P. R. Visperas, S. Deindl, T. A. Kadlecsek, A. Weiss, and J. Kuriyan. Structural basis for activation of zap-70 by phosphorylation of the sh2-kinase linker. *Molecular and Cellular Biology*, 33(11):2188–2201, 2013.
- [93] I. Letunic, S. Khedkar, and P. Bork. SMART (simple modular architecture research tool), 2019. URL `\url{https://smart.embl.de}`.
- [94] I. Letunic, S. Khedkar, and P. Bork. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1):D458–D460, 2020.
- [95] R. G. Huber, H. Fan, and P. J. Bond. The structural basis for activation and inhibition of zap-70 kinase domain. *PLOS Computational Biology*, 11(10):1–16, 2015.
- [96] C. A. Cartwright, M. P. Kamps, A. I. Meisler, J. M. Pipas, and W. Eckhart. pp60c-src activation in human colon carcinoma. *The Journal of Clinical Investigation*, 83(6):2025–2033, 1989.
- [97] E. Turro, D. Greene, A. Wijgaerts, C. Thys, C. Lentaigne, T. K. Bariana, S. K. Westbury, A. M. Kelly, D. Selleslag, J. C. Stephens, S. Papadia, I. Simeoni, C. J. Penkett, S. Ashford, A. Attwood, S. Austin, T. Bakchoul, P. Collins, S. V. V. Deevi, R. Favier, M. Kostadima, M. P. Lambert, M. Mathias, C. M. Millar, K. Peerlinck, D. J. Perry, S. Schulman, D. Whitehorn, C. Wittevrongel, B.-B. Consortium, M. D. Maeyer, A. Rendon, K. Gomez, W. N. Erber, A. D. Mumford, P. Nurden, K. Stirrups, J. R. Bradley, F. L. Raymond, M. A. Laffan, C. V. Geet, S. Richardson, K. Freson, and W. H. Ouwehand. A dominant gain-of-function mutation in universal tyrosine kinase src causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. *Science Translational Medicine*, 8(328):328ra30, 2016.
- [98] V. Pede, A. Rombout, B. Verhasselt, and J. Philippé. The biological relevance of ZAP-70 in CLL. In P. Oppezzo, editor, *Chronic Lymphocytic Leukemia*, Chapter 7. IntechOpen, Rijeka, 2012.
- [99] A. Satoh. Chapter 10 - typical properties of colloidal dispersions calculable by molecular-microsimulations. In A. Satoh, editor, *Introduction to Molecular-Microsimulation of Colloidal Dispersions*, volume 17 of *Studies in Interface Science*, pages 153 – 159. Elsevier, 2003.
- [100] T. Nishida. The social group of wild chimpanzees in the Mahali Mountains. *Primates*, 9(3):167–224, 1968.
- [101] V. F. Reynolds and F. Reynolds. Chimpanzees of the budongo forest. *I. DeVore (Ed.), Primate Behavior. Field Studies of Monkeys and Apes, New York etc. (Holt, Rinehart and Winston)*, pages 368–424, 1965.

- [102] J. van Lawick-Goodall. Mother-offspring relationships in free-ranging chimpanzees. *Primate Ethology (Weidenfeld & Nicolson)*, pages 287–346, 1967.
- [103] D. Stokols and I. Altman. *Handbook of Environmental Psychology*. John Wiley and Sons, Inc., 1987. N1 = Monograph, CY = New York [etc.], AV = Wageningen UR Library.
- [104] K. R. Dixon and J. A. Chapman. Harmonic mean measure of animal activity areas. *Ecology*, 61(5):pp. 1040–1044, 1980.
- [105] A. M. Adams, S. Madhavan, and D. Simon. Women’s social networks and child survival in Mali. *Social Science & Medicine*, 54(2):165 – 178, 2002.
- [106] S. Domínguez and T. Arford. It is all about who you know: Social capital and health in low-income communities. *Health Sociology Review*, Vol. 19, No. 1(hesr.2010.19.1.114):pp. 114–129, 2010.
- [107] J. Holt-Lunstad, T. B. Smith, and J. B. Layton. Social relationships and mortality risk: A meta-analytic review. *PLoS Med*, 7(7):e1000316, 2010.
- [108] S. M. Kana’iaupuni, K. M. Donato, T. Thompson-Colon, and M. Stainback. Counting on kin: Social networks, social support, and child health status. *Social Forces*, 83:1137–1164, 2005.
- [109] T. E. Strandberg, J. Sirola, K. H. Pitkala, R. S. Tilvis, A. Y. Strandberg, and S. Stenholm. Association of midlife obesity and cardiovascular risk with old age frailty: a 26-year follow-up of initially healthy men. *International Journal of Obesity*, 2012.
- [110] K. R. Hill, R. S. Walker, M. Božičević, J. Eder, T. Headland, B. Hewlett, A. M. Hurtado, F. Marlowe, P. Wiessner, and B. Wood. Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, 331(6022):1286–1289, 2011.
- [111] R. Layton, S. O’Hara, and A. Bilsborough. Antiquity and social functions of multilevel social organization among human hunter-gatherers. *International Journal of Primatology*, 33(5):1215–1245, 2012.
- [112] C. Boesch and H. Boesch-Achermann. *The Chimpanzees of the Tai Forest: Behavioural Ecology and Evolution*. New York: Oxford University Press, 2000.
- [113] G. Politis. *Nukak: Ethnoarcheology of an Amazonian People*. New York: Routledge, 2007.
- [114] L. C. Aiello and P. Wheeler. The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology*, 36(2):pp. 199–221, 1995.

- [115] W. X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar. Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B*, 272:439–444, 2005.
- [116] R. I. M. Dunbar and M. Spoors. Social networks, support cliques, and kinship. *Human Nature*, 6(3):273–290, 1995.
- [117] L. R. Binford. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-Gatherer and Environmental Data Sets*. University of California Press, 2001.
- [118] E. Pearce and T. Moutsiou. Using obsidian transfer distances to explore social network maintenance in late Pleistocene hunter–gatherers. *Journal of anthropological archaeology*, 36:12–20, 2014.
- [119] F. A. Hassan. Demographic archaeology. *Studies in Archaeology*, 1981.
- [120] P. Turchin. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*, volume 1. Sinauer Associates Sunderland, 1998.
- [121] E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834, 2008.
- [122] G. M. Viswanathan, M. G. E. Da Luz, E. P. Raposo, and H. E. Stanley. *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters*. Cambridge University Press, 2011.
- [123] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.
- [124] X.-W. Wang, X.-P. Han, and B.-H. Wang. Correlations and scaling laws in human mobility. *PLoS ONE*, 2014.
- [125] D. Boyer, M. C. Crofoot, and P. D. Walsh. Non-random walks in monkeys and humans. *Journal of The Royal Society Interface*, 9(70):842–847, 2011.
- [126] J. M. C. Hutchinson and P. M. Waser. Use, misuse and extensions of ”ideal gas” models of animal encounter. *Biological Reviews of the Cambridge Philosophical Society*, 82(3):335–59, 2007.
- [127] L. Barrett and C. B. Lowen. Random walks and the gas model: spacing behaviour of grey-cheeked mangabeys. *Functional Ecology*, 12(6):857–865, 1998.
- [128] H. Sugiura, C. Saito, S. Sato, N. Agetsuma, H. Takahashi, T. Tanaka, T. Furuchi, and Y. Takahata. Variation in intergroup encounters in two populations of japanese macaques. *International Journal of Primatology*, 21(3):519–535, 2000.

- [129] P. H. Crowley, S. E. Travers, M. C. Linton, S. L. Cohn, A. Sih, and R. C. Sargent. Mate density, predation risk, and the seasonal sequence of mate choices: A dynamic game. *The American Naturalist*, 137(4):pp. 567–596, 1991.
- [130] J. M. Rowcliffe, G. Cowlshaw, and J. Long. A model of human hunting impacts in multi-prey communities. *Journal of Applied Ecology*, 40(5):872–889, 2003.
- [131] W. Jetz, C. Carbone, J. Fulford, and J. H. Brown. The scaling of animal space use. *Science*, 306(5694):266–268, 2004.
- [132] J. C. Mitani, G. F. Grether, P. S. Rodman, and D. Priatna. Association among wild orang-utans: sociality, passive aggregations or chance? *Animal Behaviour*, 42(1):33 – 46, 1991.
- [133] P. M. Waser. Spatial associations and social interactions in a “solitary” ungulate: the bushbuck *tragelaphus scriptus* (pallas). *Zeitschrift für Tierpsychologie*, 37(1):24–36, 1975.
- [134] P. M. Waser. *Cercocebus albigena*: Site attachment, avoidance, and inter-group spacing. *The American Naturalist*, 110(976):pp. 911–935, 1976.
- [135] O. Schülke and P. M. Kappeler. So near and yet so far: territorial pairs but low cohesion between pair partners in a nocturnal lemur, *phaner furcifer*. *Animal Behaviour*, 65(2):331 – 343, 2003.
- [136] S. Gursky. Associations between adult spectral tarsiers. *American Journal of Physical Anthropology*, 128(1):74–83, 2005.
- [137] D. Bray, M. D. Levin, and C. J. Morton-Firth. Receptor clustering as a cellular mechanism to control sensitivity. *Nature*, page 74 175, 1998.
- [138] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 2002.
- [139] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.
- [140] H. G. Othmer, S. R. Dunbar, and W. Alt. Models of dispersal in biological systems. *Journal of Mathematical Biology*, 26(3):263–298, 1988.
- [141] S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology*, 1(3-4):137–51, 2004.
- [142] S. Lampoudi, D. T. Gillespie, and L. R. Petzold. The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems. *The Journal of Chemical Physics*, 130(9):094104, 2009.

- [143] R. A. Epstein. *The Theory of Gambling and Statistical Logic*. Elsevier/Academic Press, 2 edition, 2013.
- [144] W. C. Yueh. Eigenvalues of several tridiagonal matrices. *Applied Mathematics E - Notes*, 5:66–74, 2005.
- [145] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15:1–89, 1943.
- [146] T. Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, 1990.
- [147] J. Campbell. The smm model as a boundary value problem using the discrete diffusion equation. *Theoretical Population Biology*, 72(4):539 – 546, 2007.
- [148] A. Okubo, P. K. Maini, M. H. Williamson, and J. D. Murray. On the spatial spread of the grey squirrel in Britain. *Proc. R. Soc. of Lond. B: Biol. Sci.*, 238(1291):113–125, 1989.
- [149] H. A. David and H. N. Nagaraja. *Encyclopedia of Statistical Sciences*, Chapter Order Statistics. American Cancer Society, 2006.
- [150] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*, volume 54 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, New York, 1992.
- [151] L. Giuggioli and V. M. Kenkre. Consequences of animal interactions on their dynamics: emergence of home ranges and territoriality. *Movement ecology*, 2(1):20, 2014.
- [152] H. P. Possingham and A. I. Houston. Optimal patch use by a territorial forager. *Journal of Theoretical Biology*, 145(3):343–353, 1990.
- [153] E. Pearce, A. Shuttleworth, M. Grove, and R. H. Layton. *Lucy to Language: The Benchmark Papers*, Chapter Chapter 16: The Costs of Being a High-latitude Hominin. Oxford University Press, 2014.
- [154] B. Lufkin. How online gaming has become a social lifeline, 2020. URL `\url{https://www.bbc.com/worklife/article/20201215-how-online-gaming-has-become-a-social-lifeline}`.
- [155] R. E. Baker, C. A. Yates, and R. Erban. From microscopic to macroscopic descriptions of cellmigration on growing domains. *Bulletin of Mathematical Biolog*, 72:719–762, 2010.
- [156] S. Sobolev. Discrete space-time model for heat conduction: Application to size-dependent thermal conductivity in nano-films. *International Journal of Heat and Mass Transfer*, 108:933 – 939, 2017.
- [157] R. Cantrell and C. Cosner. *Spatial Ecology via Reaction Diffusion Equations*. ISBN: 0-471-49301-5. John Wiley & Sons, Ltd., 2003.