Modelling of the Invasion Dynamics of *Plasmodium falciparum* Merozoites into Red Blood Cells

by
Maynard Meiring

Thesis presented in fulfilment of the requirements for the degree of
Master of Science in the Faculty of Science
at Stellenbosch University

Supervisor: Prof Jacky L. Snoep
Co-supervisor: Dr. Dawie D. van Niekerk

March 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2016

Abstract

BACKGROUND The physiological symptoms of *P. falciparum* infection are associated with a 48 hour replication cycle whereby parasites (merozoites) invade susceptible red blood cells (RBCs), develop and multiply within these now infected RBCs and eventually rupture them, releasing 8 to 32 daughter merozoites which are free to diffuse through their environment and may subsequently invade other available RBCs and begin a new replication cycle.

OBJECTIVES To develop a novel, parameterised mechanistic mathematical model which describes the dynamics of merozoite invasion during the initial stages of malarial infection. Furthermore, to determine if the dispersal of daughter merozoites can be modelled by a random walk model and extrapolated to a diffusion process.

APPROACH Two novel mathematical models were constructed to describe the dynamics of the aforementioned invasion process – a deterministic model consisting of four coupled ordinary differential equations and a stochastic, random walk model. In addition, several forms of the mechanistic model were built to accommodate multiply-invaded RBCs. These parasitised RBCs are those which have been invaded by more than one merozoite; a phenomenon largely absent from modelling exercises seen in the literature. The novelty of the models pertains to the inclusion of a "merozoite-erythrocyte complex" or MEC. This is an unstable, intermediary species which forms after the initial interaction between a free merozoite and susceptible RBC but before successful invasion may occur.

CONCLUSION The most significant conclusion gained from this research was that while the mathematical models were constructed in a theoretically sound manner in which all interactions were based on observable phenomena, the data required to robustly validate these models and acquire accurate estimates for the parameters was not available.

Acknowledgements

I would like to thank my supervisor, Prof Jacky Snoep and Dr. Dawie van Niekerk from the biochemistry department at the University of Stellenbosch for guiding me through the entire duration of this project.

I would also like to thank SACEMA and SARCHI: Mathematical Modelling of Health and Epidemiology for funding this research and all those involved who provided invaluable insights and technical help without which this thesis would not have been possible.

Contents

List of Figures

List of Tables

1. Introduction

Malaria is a disease caused by a number of parasite species belonging to the *Plasmodium* genus. All of the parasite species infect and multiply within the host's red blood cells (RBCs) and two of these in particular have been identified as having a significant impact on the human population; namely *P. falciparum* and *P. vivax*. The former has been the cause of up to 500 million clinical cases and around 1.2 million deaths per year, mainly affecting the tropical and sub-tropical regions of the world [1]. However, the latter is more widespread throughout the 104 countries or regions where malaria has been defined to be endemic [2, 3]. *P. falciparum* is also the only species currently known to cause severe infection in humans, in addition to so-called uncomplicated infection which is caused by all *Plasmodium* species. Severe malaria is multi-factorial and results in many life-threatening complications such as severe anaemia, metabolic dysfunction, enlargement of the spleen and excessive accumulation of infected RBC in the microvasculature and vital organs of the host [4, 5, 6]. It is estimated that one out of every 20 cases of *P. falciparum* infection result in severe illness [2]. Uncomplicated infection is a much milder form of the disease with the general definition including presence of infected RBCs and onset of fever with no complications of sub-categories of severe infection [7].

The bite of a female *Anopheles* mosquito infected with the *Plasmodium* parasite initiates infection. This bite releases sporozoites into the bloodstream which enter the liver and develop into hepatic schizonts, each containing many thousand malarial parasites. These hepatic schizonts eventually burst releasing free parasites (formally called merozoites) into the in order for these merozoites to invade any susceptible uninfected RBCs (uRBCs) [8]. Once a parasite has successfully invaded a RBC – from here on referred to as a parasitised RBC (pRBC) – it develops through several stages where various biochemical alterations are made within the host cell [9]. This phase terminates when a mature infected RBC (schizont) ruptures and releases new parasites into the bloodstream which initiates another replication cycle; the aforementioned rupturing event is known as schizogony. The time which merozoites spend in circulation after being released from either hepatic or erythrocytic schizonts and before invading a susceptible RBC is termed the transition period of invasion; it is also the only time when merozoites are extracellular relative to any other type of cell, host or parasitic. It is noteworthy that although sporozoites have to travel through the bloodstream before invading a liver cell the transition period described above does not apply to these parasitic forms since they do not form part of the erythrocytic replication cycle.

While the main result of this cycle is to increase the number of merozoites within the human host, the

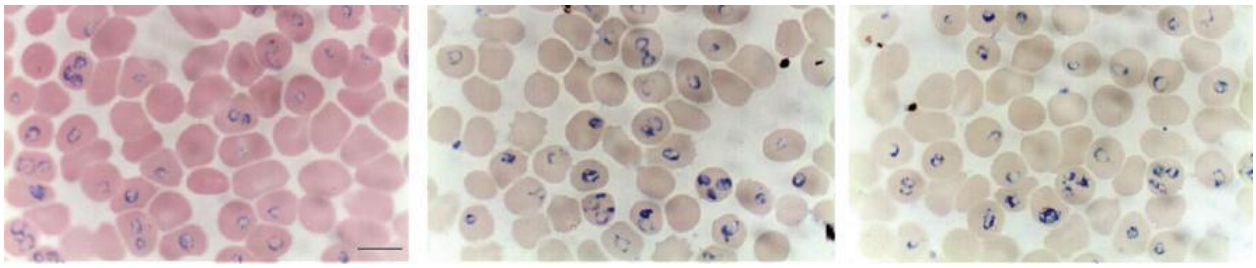

Figure 1.1: Examples of experimental cultures containing uninfected and parasitised RBCs. Three photomicrographs showing experimental cultures of pRBCs invaded by *P. falciparum* merozoites a few hours after the transition period. Singly- and multiply-invaded pRBCs can be seen in all three examples (the number of rings in each pRBC indicates the number of merozoites which invaded that particular RBC). The culture in the first image was stained with Wright's eosin methylene blue solution while the second and third were stained with 3 % and 10 % Giemsa, respectively. The scale bar in the bottom-right of the first image denotes 10 µm. Image was adapted from [10].

secondary or peripheral role is to produce gametocytes, the sexual form of the erythrocytic parasite. As with sporozoites, the aforementioned transition period does not apply to these sexual forms. Gametocytes simply undergo a series of developments over several days after which they may potentially be taken up by a female mosquito during a bloodmeal thereby entering the insect stage of the *Plasmodium* life cycle. This is how the disease is transmitted from host to vector. The complete life cycle of the parasite consists of an insect and human stage; for the human stage one can distinguish between the liver and blood forms of the parasite.

Besides the considerable number of epidemiological studies and investigations into this deadly disease where the objectives have primarily been to quantify transmission rates and assess the effectiveness of various intervention strategies within select geographical locations, a vast amount of research has been dedicated to understanding the life cycle of the parasite inside a human host as outlined earlier [11, 9, 12, 13, 3, 14, 15, 5, 16]. The current work is primarily dedicated to one particular area of this broad topic: construction of a mathematical model describing the invasion process of the transition period, i.e the period starting from the release of merozoites from either hepatic or erythrocytic schizonts up to the point where each merozoite has either successfully invaded a susceptible RBC or has perished due to natural decay.

The secondary focus – related to the first – is the phenomenon of multiply-invaded pRBCs, i.e. RBCs which have been invaded by more than one merozoite. This is not to be confused with co-infection where the host is infected by two or more different strains of the same *Plasmodium* species; a multiply-invaded pRBC is simply a RBC harbouring more than one merozoite. Figure 1.1 shows three examples of *P.*

falciparum experimental cultures where singly- and multiply-invaded pRBCs can be seen. Essentially all the studies on the topic of multiple invasions have looked at how they are associated or correlated to other measures of disease severity, but a definitive mathematical model describing their dynamics has yet to be constructed. Importantly, the effects of the immune system were not taken into account since these effects only become apparent several weeks after the initial infection and the models presented here are only concerned with naive infections.

1.1 Problem Statements

Below are the specific research questions and problem statements driven by available data and informed hypotheses about the erythrocytic replication cycle, the different types of cells present during this phase as well as the density and influence of multiply-invaded pRBCs on the overall level of infection.

i Can a plausible, mechanistic mathematical model describing the dynamics of invasion of merozoites into RBCs be constructed?

ii What are the most important biological factors affecting the propagation of merozoites?

iii What effect do these factors have on the number of multiply-invaded pRBCs?

iv What effect do physical and spatial constraints have on the dynamics of invasion?

The following approach was used to address these questions which do inevitably overlap.

Extensive literature review Many papers and articles in various journals were studied in order to grasp the scope of the known biology of within-host malarial infection. Insights were gained from a detailed literature study of a variety of perspectives including purely empirical assessments as well as theoretical arguments based on the known biology.

Collection of data A range of relevant data pertaining the research objective were collected; special attention was paid to considering in what type of experimental or clinical conditions the data were collected. The scarcity of certain types of data was well-noted and the reasons for this are detailed in the relevant chapters.

Construction of deterministic mathematical models A set of deterministic mathematical models were constructed in an attempt to explain the available data by relating each parameter to an

experimentally observable reaction or mechanism.

Construction of a stochastic model A stochastic model was built in an attempt to simulate the behaviour of newly-released merozoites from one or more rupturing schizonts in standard *in vitro* cultures. In addition, how a merozoite interacts with a RBC was also modelled within this framework. The main purpose of this model was to have a second approach of explaining the available data (the first being the deterministic models). One could also determine whether this stochastic model could predict the data pertaining to severity of infection and multiply-invaded pRBCs when using the best-fit parameters from the deterministic model. In addition, a stochastic model is better suited to describe certain scenarios than a deterministic one as will be discussed in section 1.3.

1.2 Literature Review & Motivation

The dynamics of Plasmodium infection and the phenomenon of multiply-invaded pRBCs have both been studied extensively by many researchers. As a result, several hypotheses have been proposed to explain the existence of multiply-invaded pRBCs as well as the effects they may have on the host and parasite populations [14, 17, 11, 18, 19, 16, 20]. This section reviews the current knowledge and theories regarding the different factors which influence the number of singly- and multiply-invaded pRBCs. Determining which host and parasitic factors are the most relevant to the invasion process was essential; this also informed the exclusion criteria regarding the literature used in this work. Specifically, the immune system of the host was not taken into account since it is irrelevant to the dynamics of invasion during the initial stages of infection.

1.2.1 Hematocrit & Erythrocyte Populations

The biggest determinant of parasite growth was found to be the conditions in which they were allowed to grow. A clear example of these differing conditions is whether they were grown in *in vitro* cultures or were collected from consenting patients. The conditions within these two environments are markedly different when one considers the availability of resources and metabolites, the amount of mixing and several other factors which may influence how well the parasites may grow.

Two of the most important factors are the concentration of RBCs (in cells/µl) and the hematocrit. The hematocrit of any solution is defined as the percentage of that solution which consists of RBCs. The

extent to which the hematocrit differs between *in vitro* and *in vivo* environments can be as much as two orders of magnitude with the latter kept at $\sim 45\,\%$ by homoeostatic mechanisms while the former typically ranges from $0.2\,\%$ to $5\,\%$ during experimental procedures [21, 22, 23, 24, 25]. The relationship between the concentration of RBCs and the hematocrit is intuitive since a higher concentration means there are more RBCs per unit volume which translates into RBCs making up a larger proportion of the total solution.

The crux of the matter is that the concentration of RBCs, the hematocrit and the absolute number of RBCs are all important determinants of the number of pRBCs after any replication cycle. The hematocrit in particular has been shown to correlate negatively with the fold increase in invaded cells per cycle and the re-invasion parasitemia in a consistent, statistically significant manner [14, 26, 24]. The higher the hematocrit of the experimental culture, the lower the multiplication rate of invaded cells was observed to be.[1] The fact that the multiplication rate was lower does not necessarily result in a fewer number of pRBCs, i.e the absolute number of pRBCs involved. For example, suppose one begins with a $1\,\%$ parasitemia and after one replication cycle the re-invasion parasitemia is $5\,\%$. This scenario gives a multiplication rate of 5 (final/initial parasitemia). Now suppose one performs a second experiment with the same initial parasitemia of $1\,\%$, but with twice the hematocrit of the first experiment. Suppose that after a single replication cycle one observes a parasitemia of $2.5\,\%$. This experiment results in a multiplication rate of 2.5 (half that of the first experiment) but an identical number of pRBCs because there are twice the number of RBCs in this second experiment. Several examples similar to that just described were observed by Mata-Cantero and colleagues [24].

While the hematocrit and concentration describe the quantity of RBCs, the "quality" of the RBC population is equally important and refers to the physical or temporal characteristics. These factors include their density (mass/volume), blood group and the time they have been in circulation. These form part of the more general innate resistance characteristics of the host [27]. It has been shown, for example, that the more dehydrated (more dense) a RBC is, the more difficult it is for a merozoite to invade that particular RBC [28]. Similarly, pRBCs which are dehydrated to some degree or have certain polymorphisms have been observed to rupture insufficiently (merozoites are less forcefully ejected leading to a more concentrated distribution of parasites), produce non-viable merozoites or exhibit so-called "egress abortion" whereby the pRBC does not rupture at all. [29]. An example of a morphological

[1] This multiplication rate is defined as the ratio of the number of invaded cells in the current cycle and the number in the previous cycle.

abnormality (polymorphism) is the distinct sickle shape of RBCs seen in patients with sickle cell anaemia, a well-known protective condition against the development of severe malaria. The two potential scenarios of abnormal egress just described will directly affect the number of pRBCs in subsequent replication cycles and by extension the number of multiply-invaded pRBCs [28].

Now that the potential effects of the quantity and quality of the RBC population have been discussed, a more detailed look at the initial parasitemia (the percentage of pRBCs prior to schizogony) is required since it is arguably the most intuitive contributing factor in determining the number of pRBCs after a particular replication cycle. Two types of invasion studies are regularly carried out: single-cycle assays or long-term cultures. The first is only concerned with a single replication cycle while the latter may be run for several weeks; this major difference between the two assays is due to their respective goals. The main purpose of a single-cycle assay is to determine the maximum multiplication rate of a particular isolate; for this reason the initial parasitemia is usually kept low to ensure that the number of susceptible RBCs is not a limiting factor.[2] Long-term assays are employed to maximise parasite yield after two or more replication cycles depending on the desired yield.

Both types of assays may be further categorised into those which normalise all isolates to a uniform initial parasitemia and subsequently analyse the multiplication rates of each one or, those which dilute each isolate by the same factor thereby obtaining a range of initial parasitemias. The most used solution to dilute experimental cultures is complete RPMI 1640 medium which attempts to mimic the conditions seen *in vivo* regarding amino acids, gas compositions and other essential nutrients required to support the growth of eukaryotic cells. Assays of this second type are used to determine the relationship between the initial and final parasitemias of a collection of isolates, usually from the same geographical region; the literature shows a predominantly linear relationship up to some maximum initial parasitemia [14, 16, 17, 24]. The normalised set up investigates differences in the intrinsic multiplication rate of several isolates. More specifically, normalising isolates before allowing invasion to take place leads one to consider the so-called invasion efficiency which is defined as the ratio of the number of merozoites which successfully invaded and the number of merozoites which were present before invasion took place. The efficiency with which merozoites invade is directly related to the proportion of multiply-invaded pRBCs since it considers the absolute number of merozoites as opposed to the number of pRBCs which makes this quantity an important factor to consider. One particular study observed a marked increase in invasion efficiency when using low initial merozoite to uRBC ratios [19]. This ratio may be labelled the

[2] Single-cycle assays typically use an initial parasitemia of between 0.5 % and 2 %.

initial merozoite parasitemia and is different from the usual definition of parasitemia which only takes into account the number of schizonts.[3]

A second, more common definition of invasion efficiency pertains to the population of uRBCs. It aims to quantify the degree of inhibition of invasion associated with a specific anti-malarial drug or natural polymorphism present within the uRBC population being considered. The amount of inhibition is then analysed relative to the invasion efficiency of untreated or "normal" uRBCs [17, 30, 31]. As an example, consider that merozoites exploit an array of receptor-ligand interactions with host RBCs in order to successfully invade. One of the more common RBC receptors have proteins which are sensitive to treatment with trypsin which makes those receptors unable to accommodate the associated ligand from a potentially invasive merozoite. In one particular study the trypsin-treated uRBCs were observed to be 85 % less susceptible to invasion when compared to untreated uRBCs. In other words, the invasion efficiency of trypsin-treated uRBCs was only 15 % as strong as the efficiency of untreated uRBCs [17]. This definition and its use are beyond the scope of this work however. The defining difference between this definition and the one used in the current work is that the former looks at the factors surrounding the susceptibility of uRBCs to invasion while the latter focuses on how effective merozoites are at invading susceptible RBCs.

1.2.2 Diffusion, Suspension Cultures & Multiply-invaded pRBCs

The effects of the concentration of RBCs, hematocrit, the initial parasitemia as well as the concept of the invasion efficiency have all been discussed in the context of how they affect the number of pRBCs after one or more replication cycles. However, in addition to studies comparing *in vitro* and *in vivo* conditions, several have investigated which experimental factors resulted in the highest yield of pRBCs after a pre-decided number of replication cycles [32, 20, 13]. The general finding was to use so-called suspension cultures (cultures which are shaken or agitated in some manner usually by placing them on a rotating platform set at a pre-decided speed) as opposed to static cultures (those which are not shaken), at a low hematocrit and perform regular medium changes, either daily or twice daily. The medium change is to minimise the risk of depleting the nutrients and metabolites needed to sustain the growth of the healthy and pRBC populations. Agitation of the cultures consistently produced a higher number of singly-invaded pRBCs, fewer multiply-invaded pRBCs and significantly reduced the tendency of cultures to become asynchronous. This is to say cultures were less likely to become increasingly heterogeneous

[3] An inverse relationship was observed between invasion efficiency and the initial merozoite parasitemia.

with respect to the various stages of the pRBC population over time [14, 26]. Whether there is a causal link between asynchrony and multiply-invaded pRBCs is as yet unknown since agitation of the cultures may be a confounding factor in these experiments.

When looking at any medium one must consider its diffusive capabilities since this will give one an indication of how far certain macromolecules would typically travel in a given period of time. The diffusivity and extent of mixing present in the environment have been shown to arguably be the primary determinants of the number of multiply-invaded pRBCs [14, 26]. All but one study which used static and suspension cultures observed that the proportion of multiply-invaded pRBCs, i.e the ratio of multiply-invaded pRBCs and the total number of invaded cells, was significantly lower in suspension cultures compared to those observed under static conditions [14, 19, 26, 20, 13]. An expected consequence of these findings was that the multiplication rate between cycles was greater in suspension cultures; only a single study reported no significant difference between static and suspension cultures regarding multiply-invaded pRBCs [16]. A possible reason for this result is the extent to which the cultures were mixed compared to other studies. A speed of 125 revolutions per minute (rpm) on an orbital shaker was used while the range of other experiments using an orbital shaker was 8 to 50 rpm [26, 20].

The discussion of multiply-invaded pRBCs is not complete until one considers a measure known as the selectivity index. The selectivity index was first introduced by Simpson and colleagues in a paper where the aim was to determine whether multiple invasions were the result of chance alone or if other factors influencing their numbers were present [33]. In other words, this index is a measure of the deviation of the number of observed multiply-invaded pRBCs from the number expected if invasion were a purely random process. This randomness is defined to be modelled by a Poisson distribution. The reason for choosing the Poisson distribution is because it assumes that events occur independently and at random.[4] An event in this case is the invasion of a merozoite into a singly- or multiply-invaded pRBC. Put simply, a selectivity index greater than one indicates that more multiply-invaded pRBCs are present than would be expected if invasion were truly a random process.

The selectivity index has been shown to correlate negatively with the proportion of RBCs which have been invaded (re-invasion parasitemia) in several studies [14, 30, 33]. Merozoites' preference for certain subsets (age categories) of RBCs has been proposed as the main reason why one would observe a selectivity index greater than one. While the age distribution of the RBC population does contribute other factors such as the hematocrit of a culture and the level of agitation have both been linked to

[4] These two conditions are characteristics of a totally homogeneous culture.

higher selectivity indices as well [14].

1.2.3 Parasitic Factors

The viability of individual merozoites is undoubtedly an important factor in determining the overall invasive potential of the parasite population. A recent study was conducted to determine the viability of merozoites present within late-stage pRBCs by means of a staining technique which measures the difference in the mitochondrial membrane potential of a cell [1]. These data – all obtained from suspension cultures – showed that almost half ($\sim$ 49 %) of the counted parasites within late trophozoites and schizonts were dead or dying. This has significant implications when calculating the invasion efficiency of the merozoite population since many merozoites will not be able to invade under any circumstances.

Perhaps the most important property of the parasite population is the time an average merozoite is able to survive in the extracellular environment after being released from a ruptured schizont. The average half-life $(t_{0.5})$ of a particular merozoite has an obvious effect on its invasive ability and thus the overall percentage of pRBCs after the corresponding replication cycle. The half-life of a population is simply the time it takes for half that population to decay. A half-life of between five and eight minutes has been observed for the merozoite population in experimental cultures kept at 37 °C which is the regular physiological temperature of the host [9, 19]. With the help of a simple exponential decay model, a half-life of five to eight minutes translates into an average lifespan of approximately 7.21 to 11.54 minutes.[5] While the half-life is a vital parameter, perhaps the most intuitive factor is the number of merozoites released by each mature pRBC during schizogony since a greater number of merozoites per schizont means there would be a greater probability of invasion in general.

There have been only a handful of studies which document how many merozoites were released from rupturing schizonts let alone multiply-invaded pRBCs [34, 35, 36, 37, 38]. During the literature review it was not possible to determine whether the amount of merozoites released from multiply-invaded pRBCs equal or exceed that of singly-invaded pRBCs; it may be possible that they release even fewer merozoites than their singly-invaded counterparts. Concerning *P. falciparum*, the vast majority of the literature quotes 16 to be the average number of merozoites released from a single ruptured schizont and this figure is certainly used in many if not all the mathematical models which deal with *P. falciparum* intra-host infection [19, 39, 40].

[5] Average lifespan $= t_{0.5}/\log_e(2)$.

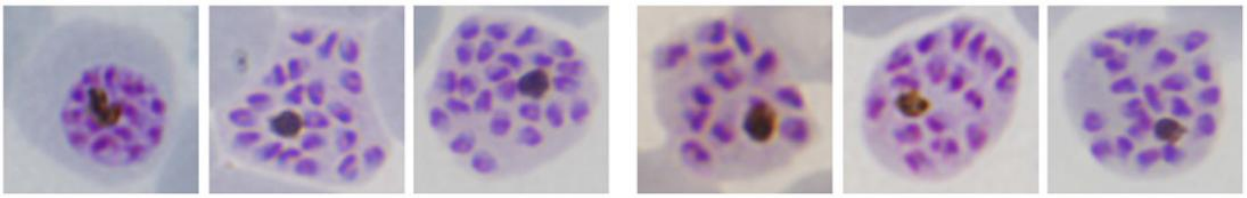

Figure 1.2: Number of merozoites contained in mature schizonts. The photomicrographs show Giemsa-stained blood smears of six mature schizonts containing – from left to right – the mean, minimum and maximum number of merozoites observed in the study. The first three images are of strain D6 while the last three are of strain W2. The merozoites can be clearly distinguished as the purple, spherical objects. The single dark spot in each image is the collection of digested haemoglobin or haemazoin. Image was adapted from [37].

Figure 1.2 shows a representative example of mature schizonts containing the observed mean, minimum and maximum number of merozoites from two different *P. falciparum* strains [37]. Taking the average of the data suggests the distribution of the number of merozoites is approximately Gaussian (normal) with a mean and standard deviation of 17.49 and 2.05 merozoites, respectively.

What should be recognised is that the number of merozoites released from individual schizonts under experimental conditions is not constant. While these observations give one opportunities for various statistical analyses, things become more interesting when one considers multiply-invaded pRBCs and how the distribution of released merozoites might change due to the fact that the RBCs have been invaded by more than one merozoite originally. This issue is perhaps not as trivial as it sounds since one also needs to consider the viability of the newly-released merozoites and how this affects the multiplication rate of subsequent replication cycles.

Now that a review of the available literature on the broad topic of within-host malarial infection which will inform the model construction process has been given, the basic concepts behind deterministic and stochastic mathematical modelling will be established.

1.3 Mathematical Modelling

A mathematical model is a tool researchers can use in order to gain insights into processes and phenomena by describing the variables in the system through one or more mathematical equations. These models serve as mathematical abstractions meant to inform ones understanding of the system being modelled, its variables and parameters, as well as allow one to make predictions about that system [41]. The system being studied in this work is the transition period of malarial infection where daughter merozoites

are released from their parent schizonts and attempt to invade susceptible RBCs in the surrounding environment before these merozoites perish. The objectives were presented in section 1.1 and address what factors affect the invasion success of merozoites and density of multiply-invaded pRBCs, and what effect spatial constraints have on the dynamics of invasion.

Having stated the objectives one must now analyse what is known about the system and its variables, what data is available and what assumptions have to be made in order to achieve the aforementioned objectives. The experimentally observable variables of the system being considered were the merozoites which did not invade any RBCs, those merozoites which *did* successfully invade and reside within the newly-invaded RBCs (pRBCs), the uRBCs which were invaded (pRBCs) and those which were not invaded. Furthermore, interactions between these variables can be observed and the rates at which these interactions occur can be experimentally measured. Because of this one can conclude that a mechanistic mathematical model would be best suited for the problem at hand. A mechanistic model is one whose outputs are the result only of interactions between variables of the system being modelled. These interactions can be represented by specific parameters which quantify the rate at which they proceed. These parameters may be dependent on time, the variables in the system or possibly both. Biologically interpretable parameters are the core of any mechanistic model because they provide information on how the outputs of the system arise through the properties of the system, its variables and their interactions.

In contrast to mechanistic mathematical models there are those which are phenomenological or statistical in nature. These types of models, also known as descriptive or empirical models, aim only to describe the outputs of the system in terms of statistical measures based on the variables of the system without taking into account how these variables interact. This is different to mechanistic models which aim to describe the outputs in terms of biologically meaningful parameters which govern the causal relationships between the variables of the system. It should be noted that a statistical model could have been employed to provide a purely descriptive analysis of the available data. Indeed, it was mentioned in the beginning of this chapter that such descriptive models were found in the literature review which related, for example, the multiplication rate to the number of multiply-invaded pRBCs.

Mechanistic mathematical models can be further categorised into two general varieties: deterministic and stochastic. A deterministic model is one which ignores any random variation which may be present within the system and will always produce the same output for a given set of inputs. By contrast, a stochastic model incorporates the effects of random variation by assuming that each interaction or process is defined by some probability distribution and will thus produce a distribution of possible outcomes. In other words,

for one set of initial conditions or inputs a stochastic model will produce, in theory, a different set of outputs in subsequent simulations. This is not to be confused with a statistical or descriptive model. A stochastic model assumes that the interactions between the variables are statistical in nature whereas a statistical model assumes no knowledge of these interactions.

Deciding whether to use a deterministic or stochastic mathematical model will depend on many factors including the available data but primarily on the size of the populations in the system. The reasoning is rather intuitive when one considers that when a relatively small number of individuals from two or more populations is being modelled the natural randomness in their movements and interactions would be more apparent than if one were modelling a large number of individuals where this randomness is assumed to be negligible. This issue of scale is directly related to how detailed the model should be in order to meet the objectives stated earlier. In order to do this one needs to make one or more assumptions about the system, its variables and their interactions. Having to make assumptions about a system is necessary because the majority of biological systems are simply too complex to allow every variable and interaction to be included in the model.

Some of the variables and parameters in the system and the surrounding environment were reviewed in section 1.2 and included the following: the populations of merozoites, uRBCs and pRBCs, the hematocrit of the solution, the concentrations of metabolites and nutrients, the blood group and hydration state of uRBCs, the intrinsic viability and invasive potential of merozoites, whether the culture was agitated in some manner, which strain of *P. falciparum* one is modelling, the temperature at which the experimental culture was maintained, the age of the uRBCs used in the culture, the time the uRBCs and merozoites were stored prior to the start of the experiments, how many of the mature schizonts failed to rupture and how many merozoites were released by each schizont which did rupture successfully. Most of these variables are measurable or can be externally controlled.

As stated earlier, most biological systems contain too many variables to include in a mathematical model to remain meaningful. This is, however, not a complete disadvantage since a mathematical model should not include every detail surrounding the phenomenon being studied. This is because a mathematical model is meant to meet the pre-decided objectives by including the most important and influential variables of the system which affect the outcomes associated with the objectives. What follows is that one must make certain assumptions about which variables are to be included and, by extension, which interactions are to be omitted. Because these assumptions are based on the supposed relevance of certain variables there will be one or more models which will be inadequate. Again, the construction of these

hypothetical models depend on the objectives and the data which is available. For example, the intrinsic viability of merozoites was only partially quantified by a single study contained in the literature review as discussed in section 1.2.3. However, because only one estimate was found no margin of error could be asserted and no comparisons could be made to different strains or different culture conditions. It was therefore assumed that all merozoites released from mature schizonts were fully and identically viable.

The question of what proportion of schizonts did not rupture successfully was not quantified in any study contained in the literature review. Because of this it was assumed that every schizont ruptured successfully. A third example, one which is ubiquitous throughout mathematical modelling literature is the lumping of processes and parameters into a single, general parameter. A simple example would be the susceptibility of a RBC which lumped its blood group, hydration state, age and its history of invasion into a single constant rate.[6] Chapters 2 and 3 discuss these assumptions and others which had to be made in order to construct the mathematical models used in this work.

1.3.1 Deterministic or Stochastic?

Returning to the question of whether a deterministic or stochastic model should be used to meet the objectives, both can and were in fact used but in different ways. Firstly, the available data for merozoites and uRBCs were given in terms of concentrations (cells/µl). This means that the number of individual cells (both uRBCs and merozoites) was large and the randomness associated with both populations and their interactions can be assumed to be negligible. In addition, suspension cultures – this is to say cultures which were shaken or agitated in some manner – ensure homogeneity with respect to uRBCs and merozoites. Both of these suggest that a deterministic model is better suited for this problem than a stochastic model. Specifically, this model consists of time-dependent ordinary differential equations (ODEs). Quite briefly, an ODE is an equation which describes the rate of change of some quantity (most often in time) as a function of that quantity and other variables which affect that quantity. The term "coupled" refers to the interdependence of several variables such as merozoites and uRBCs.

A general mathematical description of a system of coupled ODEs would be the following;

$$\Omega(\mathbf{p}) := \begin{cases} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\ \mathbf{p},\ t) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x},\ \mathbf{p}),\ \mathbf{x}(t_0) = \mathbf{x}_0(\mathbf{p}) \end{cases} \tag{1.1}$$

[6] The history of invasion of a RBC refers to how many times merozoites have tried to invade that particular RBC as well as whether the RBC had been previously successfully invaded.

where $\dot{\mathbf{x}}$ denotes the first time derivative of $\mathbf{x}$, the variables in the system, i.e. the quantities which are going to change over time. The vector $\mathbf{p}$ denotes the parameters which appear in one or more of the equations in the system. The experimentally observable quantities (output variables) are represented by $\mathbf{y}$ while the initial conditions of the state variables are contained within $\mathbf{x}(t_0)$ which may be a function of the unknown parameters. The distinction between those variables which are observable and those which are not is crucial when analysing the identifiability of the parameters as detailed in section 5.1.

An alternative way of expressing the vectors in equation (1.1) would be to write separate equations for each species where $i \in [1, S]$ and S is the number of species present in the model. For example, the ODE for the i-th state variable could be written as $\dot{x}_i = f_i(\mathbf{x}, \mathbf{p}, t)$. For convenience however, vector notation will be used throughout this work unless explicit separation of terms is necessary for clarification.

As an example, a model describing the dynamics of *P. chabaudi* infection in mice will be briefly discussed [42]. The model split the replication cycle into the transition and developmental stages and tracked the populations of merozoites (M), uRBCs (U), pRBCs (P_1) and doubly-invaded pRBCs (P_2) in both stages over 17 days. This means that several replication cycles were incorporated into the model. The transition period in their model was governed by the following system of ODEs:

$$\frac{dM(t)}{dt} = -\beta M(t)\left(U(t) + P_1(t) + P_2(t) + \mu + I_{m,i}\right) \tag{1.2a}$$

$$\frac{dU(t)}{dt} = -\beta M(t)U(t) \tag{1.2b}$$

$$\frac{dP_1(t)}{dt} = \beta M(t)\left(U(t) - P_1(t)\right) \tag{1.2c}$$

$$\frac{dP_2(t)}{dt} = \beta M(t)P_1(t) \tag{1.2d}$$

Here the infection rate is represented by β, the natural decay rate of merozoites is given by μ and the action of immune effector cells on merozoites in circulation is given by $I_{m,i}$ which was assumed constant during this relatively short phase.[7] The inclusion of this immune constant is justifiable since the entire model (transition and developmental stages) spans a length of time which is greater than the time required for the immune system to become aware of the infection. A biologically realistic assumption is that the population of RBCs remains constant in equation (1.2).[8] This is because the production and removal rates of uRBCs are orders of magnitude faster than the length of the transition period. However, the most challenging assumption to justify in this model is the global nature of the infection rate, i.e an

[7] The m and i subscripts indicate that this particular constant pertains only to the *merozoite* population on the $i-$th day of infection.

[8] This is to say: $U(t) + P_1(t) + P_2(t) = C \in \mathbb{R}$.

uRBC and singly-invaded pRBC are equally as likely to become infected by an invading merozoite.

Equations (1.2a) to (1.2d) was the archetype for the deterministic model constructed in this work where only the transition period is considered where the populations of merozoites, uRBCs and pRBCs were tracked over time. Importantly, because this proposed model has time as its sole independent variable it does not consider or predict any spatial aspects associated with the invasion of merozoites into uRBCs. While this limitation is recognised the objectives do not require any spatial considerations of either species. However, this does not mean that these aspects of the system are irrelevant or cannot be modelled or predicted. To the contrary, if there is interest is spatial dynamics this is precisely where a stochastic mechanistic model can be utilised provided the number of individuals to be modelled is small enough for the random nature of their movements and interactions to be evident.

For the purposes of this work a specific type of stochastic model, namely a random walk model will be introduced. Random walk models are very useful tools for describing the dispersal or redistribution of various populations. Examples include the movement of cells or micro-organisms due to certain chemical gradients and the paths followed by animal populations in the presence of some type of attractor or inhibitor. The model in this work would describe the movement of merozoites from the moment of rupture and be able to track their movements while searching for susceptible RBCs to invade. There are two important points regarding the applicability of a stochastic model for this scenario: it is better suited for static cultures where RBCs are effectively stationary relative to the merozoites released from rupturing schizonts which rely solely on diffusion to move around. Furthermore, the phenomenon of multiply-invaded pRBCs can easily be included into the model structure by simply allowing any RBC in the simulation to be invaded by more than one merozoite, i.e a RBC is equally susceptible to subsequent invasions after the first. Perhaps the biggest advantage of a stochastic model is that one can obtain a visual representation of what the outcome of a typical transition period might look like given that space is an explicit, independent variable unlike the deterministic model. This type of representation can then be compared to experimental data.

One can define a random walk as a mathematical sequence consisting of two or more values (formally known as locations) where the direction of each step is chosen in a random or semi-random manner. The precise nature of this randomness may be subject to the requirements or assumptions of the specific situation. For example, one might ask if each is step weakly or strongly dependent on one or more of the preceding steps or do individuals perhaps have a bias towards a particular direction? Are there obstacles, also called barriers, present in the environment which would change the trajectory of a individual's path

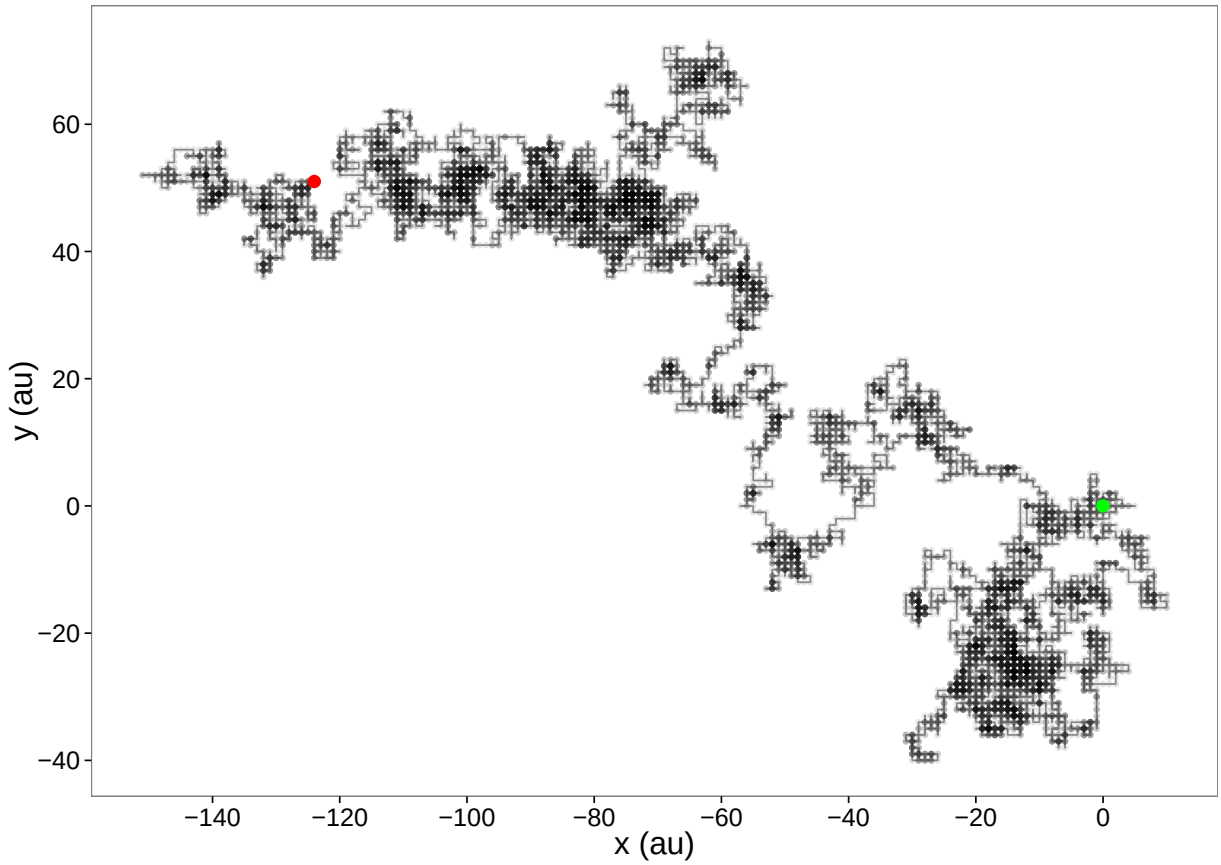

Figure 1.3: Two-dimensional simple random walk. The plot shows a single individual engaged in a simple random walk on a two-dimensional lattice where each step has a length of one and the total number of steps was 10^4. The probabilities of stepping left, right, up or down were equal, i.e 0.25. The beginning and end points of the walk are represented by the green and red points, respectively. The darker the points and steps are the more times they have been traversed through the course of the walk by the individual. Since this is purely an example the axes are in arbitrary units (au).

(a reflective barrier) or cause that individual to stop moving altogether (an absorbing barrier)? Both of these questions among others are discussed in detail in chapter 3 where the stochastic model is constructed.

As a simple example, consider a single individual on a two-dimensional lattice which takes a pre-determined number of steps with the probabilities of stepping left, right, up or down being equal. This would be considered a simple random walk (SRW) since it contains no barriers and is unbiased and uncorrelated with regard to direction and movement. An example of this stochastic process is shown in figure 1.3 where an individual begins at the origin denoted by the green point and stops after 10^4 steps at the position denoted by the red point. One can then calculate various metrics of this particular realisation

including the total distance covered, the absolute displacement of the individual and the angle between the start and end points.

An important property of a simple random walk is that it will converge to the classical diffusion equation as the number of steps increases and the size of each step becomes infinitesimally small. The two-dimensional diffusion equation was the first candidate considered for modelling the dispersal and movement of merozoites but was rejected given that the goal was to model a relatively small number of individual merozoites on a short time scale. Furthermore, the diffusion equation provides a probability gradient of where individuals may end up and not explicitly defined end points of the individuals' paths. The choice of a discrete, random walk model was therefore deemed more suitable.

What must be understood is that the deterministic and stochastic models share most of the same parameters but because of the probabilistic nature of the merozoites' movements and interactions with RBCs in the stochastic model, one will obtain a distribution of possible outcomes instead of only one produced by the deterministic model. The inclusion of mathematical modelling has become an integral part in the biological sciences because of the need to understand and quantify relationships between variables from vast amounts of diverse experimental and clinical data. For malaria research in particular, this practice has provided valuable insights and been fitted to many types of data including control studies done with select rodent species [43, 44, 42].

As stated in the beginning of this chapter, the lack of mechanistic mathematical models describing the transition period of malarial infection, including the phenomenon of multiply-invaded pRBCs, is the primary motivation for the current research. As such, two distinct mathematical models were constructed: one deterministic and one stochastic. The final models are those which allow the inference of certain mechanistic properties of the system and evaluation of the parameters in a biologically meaningful way. Most importantly, when conclusions were drawn from data which differ in regard to the experimental conditions or underlying assumptions of certain mathematical models, care was taken in assessing how valid those conclusions would be. The stochastic, mechanistic mathematical model was a random walk model where merozoites were free to travel from there respective rupture sites and invade susceptible RBCs with the possibility of RBCs being invaded by more than one merozoite. This stochastic model incorporated the parameters present in the deterministic model while also using data relating to the rate of diffusion of merozoites in experimental cultures.

The remainder of this work will be structured as follows: chapters 2 and 3 will utilise the known biology of the *P.falciparum* transition stage presented in the introduction and literature review to construct the

deterministic and stochastic mathematical models, respectively. Chapter 4 introduces the available data used to estimate the free parameters in the deterministic model while also giving the experimentally determined range of values for those parameters found in the literature review. In addition, the conditions in which the data were collected will also be taken into account and related to the assumptions and justifications in the model construction process. The results of the model fitting procedures are the subject of chapter 5. Also included are the outcomes of the random walk model which was not fitted to any additional data but, incorporated the best-fit parameters from the deterministic model. Chapter 6 provides a detailed analysis of the results in chapter 5 while also discussing the limitations of the available data as well as the mathematical models. Major conclusions and possible future research topics are presented in chapter 7. Lastly, the appendix is reserved for mathematical definitions and notations and will be referenced to accordingly in each chapter of the main body.

2. Deterministic Mathematical Model

The framework introduced in this chapter is an elaboration of the ideas and methodologies presented in section 1.3, where the fundamental notions of mathematical modelling were discussed. A foundational, base model will be presented first upon which a model incorporating the possibility of multiple invasions will be discussed.

2.1 Base Model for Erythrocytic Invasion

Consider the following schematic which describes the process of parasitic invasion where each arrow represents a separate process relevant to the rate of invasion. The items in red denote the novelty of the current model and are discussed below.

$$m \quad + \quad r \quad \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \quad c \quad \xrightarrow{k_2} \quad x_r \tag{2.1}$$

The parameters k_1, k_{-1} and k_2 are the rates at which different reactions occur and represent the contact rate, rate of detachment, and the rate of successful invasion, respectively. Since each parameter is assigned to a specific reaction, one may label the three reactions as follows: $v_1 = k_1 \cdot m \cdot r$, $v_2 = k_{-1} \cdot c$ and $v_3 = k_2 \cdot c$. Free merozoites are denoted by m while r and x_r denote uninfected and singly-invaded pRBCs, respectively. The intermediate species (c) is the so-called "merozoite-erythrocyte complex" or MEC. The formation and breakdown rates of the MEC (denoted by k_1 and k_{-1}, respectively) show that the initial contact between a merozoite and RBC is a reversible process; this phenomenon will be elaborated on in section 2.1. The MEC is the most important species when one is looking at the invasion efficiency of free merozoites, the susceptibility of a RBC to subsequent invasions and, in particular, in attempting to form a mechanistic model of parasitic invasion. The reason why this intermediary species has been omitted in all previous studies found in the literature review has been the willingness of researchers to favour simplicity when dealing with questions pertaining to modelling the more general progression of the disease over several days or weeks, thereby justifying the aforementioned omission. This has lead to lumping all three parameters together into a single "invasion rate" leading to a model like that depicted in scheme 2.2.

$$m \quad + \quad r \quad \xrightarrow{k} \quad x_r \tag{2.2}$$

The implications of lumping two or more parameters together to create a new parameter describing the general process being considered was discussed in section 1.3. The parts coloured red in schemes

(2.1) and (2.2) depict the idea of parameter aggregation (lumping) seen in the literature. The explicit consideration of reactions with mechanistic interpretations aims to negate some of the issues associated with model reduction.

The second major factor contributing to the invariable absence of this complex is that experimental data characterising the kinetics surrounding this complex have historically been scarce and largely unreliable. However, recent studies have utilised more advanced experimental techniques in an attempt to uncover the processes which define the transition period [21, 45, 46, 47, 48]. Factors affecting various steps of this critical period have also been determined and quantified to some degree such as the time it takes for a merozoite to first contact a susceptible RBC and how long from this initial contact it takes to penetrate and fully internalise itself inside the host RBC. These advances, along with the data they have generated, helped inspire the inclusion of an intermediate complex as a crucial variable within a mechanistic mathematical model describing the process of parasitic invasion.

The structure of scheme 2.1 was inspired by that describing a typical enzymatic reaction as used in the field of biochemistry where the aforementioned mechanisms may be thought of as distinct reactions resulting in a final product; a pRBC in this context.[1] What is crucial to understand is that scheme 2.1 does not explicitly model the phenomenon of multiple invasions – this is dealt with in more complex models. A brief description of the three reactions will now be given; all units of time are in minutes (min) in accordance with the time-scale used in the model simulations.

2.1.1 Successful Location and Preliminary Attachment

Reaction: $v_1 = k_1 \cdot m \cdot r$.

This is the only reaction which requires a second-order rate constant (k_1) which means it has units of per concentration per minute ($C^{-1}min^{-1}$) where concentration has units of cells per microlitre (cells/µl). The reaction itself (v_1) represents how many contacts are made between merozoites and uRBCs in one minute. By definition the reaction is dependent on the concentrations of merozoites and uRBCs but, also on the degree of homogeneity (amount of mixing) of the environment since more mixing will result into a higher effective contact rate and vice versa. The rate constant k_1 measures the time from schizont rupture to the instant before so-called "tight junction formation" between the merozoite and RBC resulting from correct orientation of the merozoite. Orientation refers to the movement of the

[1] The MEC in this analogy is the enzyme-substrate complex.

attached merozoite along the surface of the uRBC to the point where its apical end (the part of the merozoite responsible for penetrating the RBC) is in contact with uRBC surface. This orientation has been shown to be essential for subsequent invasion and is referred to as apical alignment [21, 49].

Note that this reaction is influenced only by the *quantity* of the RBC population and not the *quality* thereof. The *quantity* refers to the absolute number of RBCs or the hematocrit of the experimental culture; section 1.2.1 describes these in more detail. Therefore, the susceptibility of an uRBC as determined by its age, physical density and the absence or presence of any polymorphisms is not relevant to this reaction. The physical density of individual RBCs, primarily determined by the level of dehydration, has been shown to affect their susceptibility to invasion and cause improper rupturing of mature schizonts [29, 28]. This factor is however, not explicitly considered in the current work.

Physical and spatial constraints may also be important in determining the value of this reaction's parameter (k_1). These constraints are related to the homogeneity (or lack thereof) of the solution or environment. The reason these constraints are important in determining k_1 is rather intuitive since if physical obstacles are preventing a merozoite from successfully reaching an uRBC it would lower the rate at which successful contacts are made, i.e. it would lower v_1.

2.1.2 Detachment / Unsuccessful Invasion

Reaction: $v_2 = k_{-1} \cdot c$.

This reaction effectively reverses what was achieved by the first reaction – it quantifies how many merozoites detach from their associated RBCs every minute. While the reaction depends on the concentration of MECs, its parameter k_{-1} has units of per minute (min^{-1}) and is assumed to depend on the amount of mixing in the environment as well as the quality of the RBC population. Understanding the impact of the former is intuitive since if there is a high level of mixing in the environment the shear forces will facilitate quicker detachment (prevention of the merozoite to properly re-orientate) and result in an unsuccessful invasion attempt. The idea that k_{-1} is dependent on the quality of the RBC population is justified when one recalls that these characteristics, time in circulation, density, state of hydration, etc. all influence the probability of successful invasion. However, these variables of the RBC population are all implicitly included within the parameter k_{-1} and are not taken into account individually.

Independence from the physical and spatial constraints mentioned in section 2.1.1 is also assumed since

one is considering a single complex, the MEC, rendering any obstacles in the environment irrelevant. As explained earlier, this step has traditionally been omitted from mathematical models describing the process of parasitic invasion and has typically been lumped together with the other two reactions.

2.1.3 Successful Invasion

Reaction: $v_3 = k_2 \cdot c$.

The final reaction is the penetration of the merozoite into the host uRBC and subsequent internalisation. Once apical alignment has been achieved (explained in section 2.1.1), the tight junction is formed which binds the merozoite to the uRBC irreversibly; penetration and internalisation then follow. This reaction is defined to begin when the tight junction has formed, however, defining the end-point may be subject to debate. This is because after a merozoite has successfully penetrated and internalised itself, a resting period is observed, after which the host uRBC undergoes a process called echinocytosis in which several, evenly-spaced protrusions appear on its membrane. The uRBC, now called an echinocyte, may remain in this configuration for several minutes before returning to its natural, biconcave shape; of course with the merozoite now internalised.

The issue becomes whether one defines "successful invasion" from the moment of tight junction formation to the first sign the merozoite has fully internalised, or until the end of echinocytosis when it has assumed its regular shape. For the base model, it makes little difference since once a RBC has been successfully invaded it is not subject to further interaction with any other merozoite or RBC. This also means that one of the simplifying assumptions of the base model is that no multiple invasions can occur. The susceptibility or "quality" of the target uRBC is the primary determinant governing this reaction's parameter (k_2); some of these factors were listed in sections 1.2.1, 1.3 and 2.1.2.

To summarise: the base model as well as all subsequent models utilise this parameter in a purely temporal capacity. In other words, the aforementioned RBC characteristics are implicitly included in the estimations of this parameter whose reciprocal quantifies the average time it takes a merozoite to (fully) internalise itself in a host RBC from the time of apical alignment and tight junction formation.

2.1.4 Base Model Assumptions

While reaction in scheme 2.1 does represent a plausible mechanism for the process of invasion and can be classified as a deterministic interpretation thereof, the reaction rates in this base model are assumed to be constant. This will be referred to as the core assumption. The reality is that they are combinations of several underlying factors to be handled by more complex models. Despite this initial limitation, the explicit separation of reactions allows one to analyse the entire process of invasion more accurately which is a hallmark of deterministic mathematical models as discussed in section 1.3. Further underlying assumptions of the base model arising from the general reaction scheme are the following:

(i) a physical interaction between a free merozoite and uninfected RBC is required to initiate the invasion process, (ii) the number of RBCs remains constant during the transition period, (iii) RBCs of all ages are equally susceptible to invasion, (iv) merozoites decay at a rate proportional to their current population, (v) any RBC may only be successfully invaded once, (vi) any MEC may only have one merozoite attached to its surface at any one time, (vii) a pRBC can not be converted back to an MEC, and (viii) each mature schizont releases exactly 16 daughter merozoites upon rupture.

The first assumption is a necessary, but not sufficient condition for successful invasion as indicated by the double arrows in scheme 2.1; this means reactions can proceed in both reactions. Experimental, micrographic evidence also justifies this primary assumption [19, 50]. Assumption (ii) is justified by the fact that the transition period is very short compared to the average lifespan of a RBC and because the immune system is not active during the initial stages of infection. In other words, neither uRBCs nor pRBCs are phagositised or naturally decay in the given time frame which means a closed population of RBCs exists. The assumption that all RBCs are equally likely to be invaded is justified *in vitro* when only RBCs of a certain age range are used for invasion assays. In addition, it has been consistently reported that *P. falciparum* has little preference for RBCs of a specific age to other *Plasmodium* species.

The fourth assumption is based on the fact that merozoites are known to naturally decay in the extracellular environment and this decay is assumed to be exponential which is quantified by the first-order rate constant μ, first introduced in section 1.2.3. Assumption (v) is imposed to ensure the absence of any multiply-invaded pRBCs since the base model is defined to only model primary invasions, while (vi) is merely a simplifying assumption which reduces the number of species and differential equations in the model. The penultimate assumption (ass. (vii)) is based on the fact that once a merozoite has

internalised itself it cannot be removed and as such makes pRBCs the absorbing class of the base model.[2] The last assumption is taken from the literature as discussed in section 1.2.3.

2.1.5 Base Mathematical Model

Now that the reaction scheme for the base model, its underlying assumptions and the relevant parameters have been explained, the actual system of ODEs can be introduced.[3]

$$\frac{dm(t)}{dt} = -k_1 m(t) r(t) + k_{-1} c(t) - \mu m(t) \tag{2.3a}$$
$$\frac{dr(t)}{dt} = -k_1 m(t) r(t) + k_{-1} c(t) \tag{2.3b}$$
$$\frac{dc(t)}{dt} = k_1 m(t) r(t) - (k_{-1} + k_2)\, c(t) \tag{2.3c}$$
$$\frac{dx(t)}{dt} = k_2 c(t) \tag{2.3d}$$

All three reactions described in scheme 2.1 as well as the natural decay rate μ (defined earlier) are present within the ODE system (equation (2.3)). There are two important properties of the above system which should be noted. The first is that it is, for all intents and purposes, a three-dimensional system since all four equations are independent of $x(t)$. This means that the population of pRBCs will not influence any of the species in the model. The analytical solution of $x(t)$ will therefore be an integral of the form

$$x(t) = k_2 \int_0^t c(z) dz \tag{2.4}$$

This independence has implications when considering a steady state solution of the system; it is not difficult to see that only the trivial steady-state exists even when equation (2.3d) is excluded.[4] Obtaining steady state solutions is however, not the goal of the current work. The second consideration – related to the first – is that a conservation relationship exists between uRBCs, pRBCs and MECs since the sum of their respective derivatives is zero. This relationship shows that the sum of all three populations at any given time will always be equal to a constant, the initial population of uRBCs in this case. This is important when parameter identifiability must be done. The next section provides the basis for the multiple invasion mathematical model.

[2] An absorbing class in a mathematical modelling context refers to a species which has no removal terms in its corresponding ODE; this means its growth rate is strictly positive.

[3] Section 1.3 introduced the mathematical notation used in describing a system of coupled ODEs through equation (1.1).

[4] The trivial steady state of a system is where every species has a population of zero.

2.2 Model for Multiple Invasions

This section presents the deterministic mathematical model which describes the phenomenon of multiply-invaded pRBCs along with the appropriate assumptions. The reaction scheme is similar to that of the base model with two defining differences; pRBCs may be invaded more than once and distinct MECs are present.

$$\begin{array}{lcccccc} m & + & r & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & c_1 & \xrightarrow{k_2} & x_r^{(1)} \\ m & + & x_r^{(1)} & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & c_2 & \xrightarrow{k_2} & x_r^{(2)} \\ & & & \vdots & & & \\ m & + & x_r^{(N-1)} & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & c_N & \xrightarrow{k_2} & x_r^{(N)} \end{array} \tag{2.5}$$

The most important consideration about the multiple invasion model, besides the distinct MECs, is that the reaction rates have been kept identical which forms part of the modified assumptions for this new model. A subtle, but very crucial point to recognise about the new model is that while the number of parameters has not increased, it is biologically more plausible given the observable phenomenon of multiply-invaded pRBCs.

2.2.1 Multiply-invaded pRBC Model Assumptions

Most of the assumptions from the base model remain unchanged for this more complex model; only the new or modified assumptions are listed below.

(i) a physical interaction between a free merozoite and an uRBC or pRBC is required to initiate the invasion process, (ii) any RBC may be successfully invaded N times by N distinct merozoites, (iii) each pRBC (of any invasion class) is equally susceptible to additional invasion attempts and, (iv) a unique MEC exists for each invasion class.

The first assumption is simply an extension of that stated in the base model by including the possibility of interactions with pRBCs. The maximum number of merozoites which may invade a single RBC is stated in assumption (ii) and explained further in the appendix. In addition, an "invasion class" denotes the group of pRBCs invaded by a particular number of merozoites; i.e first, second, etc. invasion class. Assumption (iii) is based on the fact that the reaction rates for all invasion classes have been kept the same as mentioned earlier. The most important assumption is the last which says that there is a distinct MEC species between different invasion classes. One could assume a single MEC, but this would not

capture the biologically realistic scenario where a different MEC is required to move from one invasion class to the next. If this simplifying assumption were to be made, multiply-invaded pRBCs could be produced from the same pool of MECs which produce their singly-invaded counterparts which is clearly not biologically accurate and would result in an overestimation of the actual number of multiply-invaded pRBCs. In other words, moving from one invaded class to another now requires transitioning from a unique MEC specific to that particular invasion class in a step-wise manner instead of a global MEC which caters for all invaded classes.

3. Stochastic Mechanistic Model

This chapter describes the construction of the stochastic model whose purpose was to simulate the path followed by one or more merozoites immediately after release from a rupturing schizont in the presence or absence of susceptible RBCs. First and foremost, the biggest simplification is that the model was constructed in two dimensions and not three as in the physical, biological case. This was done because the mathematics required for the construction of the model is more widely used and a greater selection of tools exist for the two-dimensional case than its three-dimensional counterpart. In addition, the results obtained in the two-dimensional case are often transferable into meaningful properties and conclusions for the three-dimensional case.

Secondly, the majority of assumptions of the mechanistic mathematical models listed in sections 2.1.4 and 2.2.1 also apply to the random walk model. Assumptions such as the necessity of a physical interaction for invasion to occur and that a pRBC has a uniform susceptibility regardless of how many merozoites have previously invaded it. However, it must be understood that the random walk model and the mechanistic model are not completely separate or distinct approaches since parameters estimated from the mechanistic model are used to generate the results from the random walk model.

A foundational model whose sole purpose was to govern the possible directions either of the two species (merozoites and RBCs) may travel in was first constructed. Once this model was operational, several factors affecting the dynamics of the interactions between a merozoite and a RBC were added to the model one at a time verifying that the updated model was fully functional before adding the next factor. What this means concretely is that the level of complexity or realism of the model was increased and verified after each iteration. The specific factors which were added – including the assumptions behind them – are discussed in this chapter.

Before this however, the shape and size of merozoites and RBCs should be defined. Merozoites and RBCs present in all simulations were both modelled as circular objects with radii of 0.5 µm and 3.50 µm, respectively. Using these two values along with the assumption of perfect circularity proves to be reasonable when one considers the physical appearance of both species.

3.1 Movement of Merozoites

The two-dimensional nature of the random walk means any location on the plane can be represented by a point (x, y) where the first is the horizontal co-ordinate and the second the vertical co-ordinate. This framework allows one to take a step in one of three directions: horizontal, vertical or a combination of the two. More specifically, movement in any of the three directions may be either forwards or backwards.

All RBCs (invaded or not) were kept stationary for the entire length of any simulation, this was done for two reasons. The first is concerned with the fact that these simulations are only being compared to experimental data obtained from static cultures which, by definition, are free from any external agitation and are mainly reliant on diffusive processes for the physical movement of any macro-molecules present in the environment. In other words, RBCs will displace only minimally while merozoites will move by diffusive processes created by the force with which they are ejected from their parent schizont. The second reason, related to the first, is the assumption that the magnitude of the physical displacement of any particular RBC is negligible compared to that of any given merozoite on the time scale being used which was no more than 15 minutes (a length of time in which virtually all released merozoites had either successfully invaded a RBC or decayed).

That being said, the mechanisms governing the direction in which a particular merozoite was allowed to move from one position to the next was, naturally, a crucial aspect of the random walk model. The initial assumption was that movement occurred on a lattice which meant that a merozoite had only four possible choices for each step: up, down, left or right. The fact that only a finite number of choices exist is a defining feature of working on a lattice regardless of the dimension of the problem. This assumption is inherently unrealistic when one considers the biological phenomenon being studied, or essentially any other for that matter.

A more realistic approach was therefore needed, one which allowed for a continuous choice of direction for each step. One could argue that were the mesh small enough (small enough step size) then working on a lattice would be a fair and justifiable assumption. However, for the reasons discussed in section 1.3.1 such as merozoites having a non-zero, finite step size and because a strong mathematical foundation exists for utilising a more realistic approach, it was decided to abandon the lattice approach and pursue the admittedly more complex but more biologically plausible route.[1] As such, a well-studied circular distribution called the von Mises distribution would be used [51].

[1] It should be noted that while merozoites do not have an intrinsic step size, it is non-zero and finite.

3.1.1 The von Mises Distribution

The von Mises distribution, like all circular distributions, is a continuous probability distributions and as such any angle in the unit circle may be chosen as the direction of movement – this also means that all circular distributions are defined only in two dimensions. The choice of which distribution to use was relatively arbitrary since all that was needed was a method of sampling random angles from a unit circle which all circular distributions are designed to do. However, the von Mises distribution is simpler to deal with analytically than other commonly used circular distributions such as the wrapped normal or wrapped Cauchy distributions [51].

The parameters defining each distribution allow one to increase the probability of choosing an angle in some region of the unit circle thereby limiting the choice of possible directions for a walker's next step. In the case of the von Mises distribution this is controlled by θ_0, the direction in which the distribution is centred and κ, a non-negative real number which provides a measure of how concentrated the distribution is around the centred direction.[2] The probability density function (PDF) of the von Mises distribution is defined as,

$$M(\theta, \theta_0, \kappa) \ := \ \frac{e^{\kappa \cos(\theta - \theta_0)}}{2\pi I_0(\kappa)} \tag{3.1}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order zero.[3] The von Mises distribution is also known as the circular normal distribution since the parameters θ_0 and $1/\kappa$ are analogous to μ and σ^2 of the more familiar normal distribution. Furthermore, as κ increases the von Mises distribution approaches a normal distribution with mean μ and variance σ^2. In other words, one has the limit of

$$\lim_{\kappa \to \infty} M(\theta, \theta_0, \kappa) \ = \mathcal{N}\left(\mu, \sigma^2\right).$$

Figure 3.1 provides an example of the von Mises distribution where 100 angles have been sampled with a mean angle of zero ($\theta_0 = 0$) and concentration parameter of two ($\kappa = 2$). One notices how the distribution of points is roughly centred around zero since κ is not too small. If κ were larger then the sampled points would be even more concentrated around zero.

Having defined the probability density function one can immediately consider the mean sine and cosine of the distribution [51, 52]. These two quantities are the expected values of the sine and cosine of θ for any one realisation of the random walk model, i.e the path of a single walker. An intuitive interpretation of the expected value of a continuous random variable is that it is the average value one would obtain

[2] If $\kappa \approx 0$ the distribution is essentially uniformly dispersed while if κ is large the distribution is very concentrated.

[3] This general function is defined by $I_m(\kappa) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(m\theta)\, e^{\kappa \cos(\theta)} d\theta$.

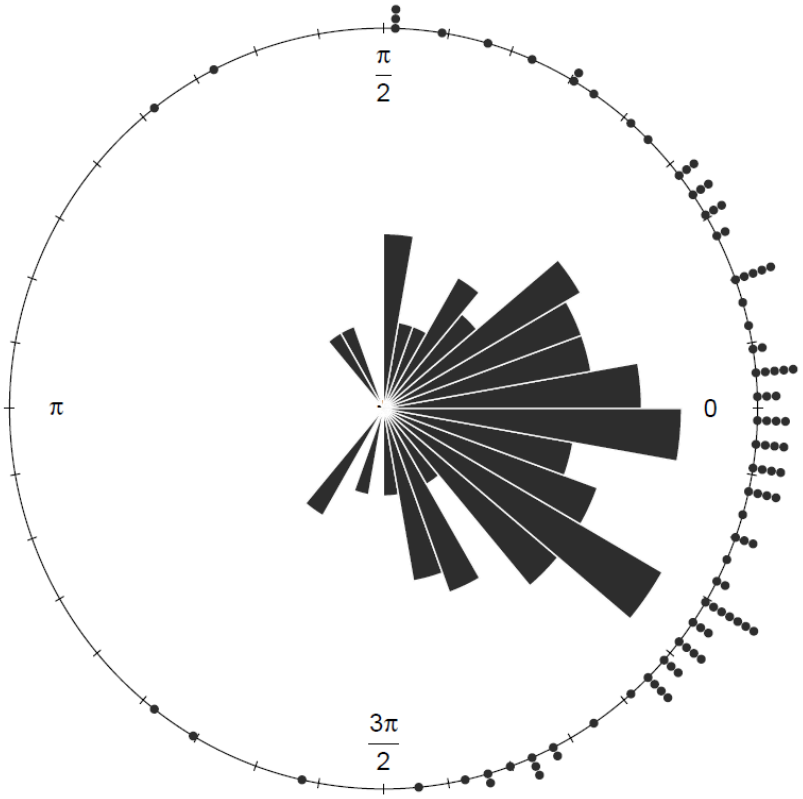

Figure 3.1: A random sample from the von Mises distribution. The mean angle θ_0 was set to zero and the concentration parameter κ to 2. A total of 100 samples were drawn. The unit circle has been divided into 36 bins which means each bin has a width of 10°. Samples falling into the same bin were stacked on top of one another to better show the distribution of points. The length and breadth of each wedge stretching out from the centre correspond to the distribution of points seen on the circle.

after a large number of repetitions of the process it represents. The mean sine is defined as

$$\begin{aligned} s &= E\left(\sin(\theta)\right) && \text{(3.2a)} \\ &= \int_{-\pi}^{\pi} \sin\left(\theta\right) M\left(\theta, \theta_0, \kappa\right) d\theta && \text{(3.2b)} \\ &= \sin(\theta_0)\frac{I_1(\kappa)}{I_0(\kappa)}, && \text{(3.2c)} \end{aligned}$$

and the mean cosine is similarly defined as

$$\begin{aligned} c &= E\left(\cos(\theta)\right) && \text{(3.3a)} \\ &= \int_{-\pi}^{\pi} \cos\left(\theta\right) M\left(\theta, \theta_0, \kappa\right) d\theta && \text{(3.3b)} \\ &= \cos(\theta_0)\frac{I_1(\kappa)}{I_0(\kappa)}. && \text{(3.3c)} \end{aligned}$$

Note that the terms *direction* and *angle* have thus far been interchangeable, however, the former will be used henceforth to avoid confusion with the *turning angle* of a particular step which is defined to be the difference between the subsequent step's direction and its own [52]. In other words, the turning angle quantifies how much and in which direction (clockwise or anticlockwise) the walker had to turn in order

to face its new direction; these turning events have also been termed reorientation events [53, 51]. For a step j the turning angle Φ_j is defined as the following,

$$\Phi_j \;=\; \theta_{j+1} - \theta_j \tag{3.4}$$

where θ_j and θ_{j+1} denote the directions of steps j and $j+1$, respectively. An important question now is whether a random walk will be biased or correlated (or possibly both) since simply drawing a random sample from a circular distribution is not sufficient to do so. While this is a necessary condition, a sufficient condition is deciding if one is sampling the direction or turning angle of a particular step, i.e the interpretation of θ. In a biased random walk (BRW) the direction is being sampled while in a correlated random walk (CRW) one is sampling the turning angle [51]. Deciding which random walk model to choose will always depend on the context in which one is working which in this case is the dispersal and movement of merozoites immediately following rupture.

When looking at a rupturing schizont one notices that a single merozoite is ejected first followed by two or three more merozoites which travel significantly slower. After this the now lysed RBC curls around itself allowing the remaining merozoites to begin moving in any direction they wish relative to the others. It is now tempting to use a biased random walk since each merozoite has a bias regarding which direction it is going to travel in based upon the direction of it's first step. In other words, a merozoite will move in roughly the same direction for the entirety of it's lifespan regardless of any obstacles it may encounter or the direction of any previous steps besides the first.

In contrast, if each merozoite were modelled with a correlated random walk then each step would be dependent on the previous one and as such would alter a merozoite's route accordingly if it were to encounter an obstacle. It was decided that a correlated random walk would be preferable since there would be obstacles present in the environment (in the form of RBCs) which would affect the direction in which merozoites who encounter them will travel. In light of this the von Mises distribution was used to sample the turning angle of each step rather than the direction.

3.1.2 Derivation and Usage of Distance Metrics

In order to fully utilise the chosen distribution one requires two additional properties of any random walk: the mean squared displacement $\left(E\left(R_N^2\right)\right)$ and mean dispersal distance $\left(E\left(R_N\right)\right)$ [51]. The former is the expected value of the squared displacement of the average walker and is thus a measure of how dispersed the population is in space over a given time. The mean dispersal distance gives one an estimate of the

displacement of the average walker in a given time. It is also the expected value of the so-called beeline distance (denoted by R_N) which is simply a straight line from the starting point of the random walk to the end point. For any random walk the mean squared displacement is defined by the following integral

$$E\left(R_N^2\right) = \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} |\mathbf{x}|^2\, p(\mathbf{x},\theta,t)\; d\theta d\mathbf{x} \tag{3.5}$$

where $p(\mathbf{x},\theta,t)$ is the probability density function for the specific random walk and $\mathbf{x} = [x\ y]^T$ is the vector in Cartesian coordinates.[4] Similarly, the mean dispersal distance is obtained by evaluating the integral

$$E\left(R_N\right) = \int_{\mathbb{R}^2}\int_{-\pi}^{\pi} |\mathbf{x}|\, p(\mathbf{x},\theta,t)\; d\theta d\mathbf{x} \tag{3.6}$$

Notice that the mean dispersal distance is not simply the square root of the mean squared displacement. In addition, evaluating the integral in equation (3.6) is made very difficult by the absolute value and resulting square root and as such has proved to be a non-trivial problem to solve in two and three dimensions. However, even though equations (3.5) and (3.6) are generally intractable – particularly when $p(\mathbf{x},\theta,t)$ cannot be directly computed – expressions do exist for both quantities depending on what type of random walk is being considered. Before this can be done, a parameter related to both the mean squared displace and mean dispersal distance needs to be quantified: θ_0, the angle around which the distribution is centred. Because of the absence of any bias in the random walk this angle was set to zero, i.e $\theta_0 = 0$. This may be interpreted as sampling a turning angle from the distribution which has no bias towards either clockwise or anticlockwise directions (right- or left-hand turns) which means that $s = 0$ and $c = I_1(\kappa)/I_0(\kappa)$ from equations (3.2c) and (3.3c). This has several implications since both distance measures are dependent on the mean sine and cosine. For a correlated random walk with a constant step length l it can be shown that the mean squared displacement is given by

$$E\left(R_N^2\right) = l^2\left(N\left(\frac{1+c}{1-c}\right) - \frac{2c(1-c^N)}{(1-c)^2}\right) \tag{3.7}$$

where c and N are the mean cosine and number of steps taken, respectively [54, 51]. In this work a constant step length is assumed denoted by l and is the distance a merozoite covers in a single step. Derivation of the expression in equation (3.7) begins with defining the total displacement of some walker after a path of N steps by

$$R_N = \sum_{m=1}^{N}(x_m, y_m) \tag{3.8}$$

[4] The distance or *norm* of $\mathbf{x}$ is $|\mathbf{x}| = \sqrt{x^2+y^2}$ while the *transpose* of a vector or matrix $\mathbf{x}$ is denoted by $\mathbf{x}^T$.

The mean squared displacement (MSD) will be then the expected value of the square of the total displacement [52]:

$$E\left(R_N^2\right) = E\left(\sum_{m=1}^{N}(x_m, y_m)\right)^2 \tag{3.9a}$$

$$= E\left(\sum_{m=1}^{N}(x_m, y_m)^2\right) + E\left(2\sum_{m>j}^{N-1}\langle(x_m, y_m),(x_j, y_j)\rangle\right) \tag{3.9b}$$

$$= E\left(\sum_{m=1}^{N}\langle Z_m, Z_m\rangle\right) + 2E\left(\sum_{m>j}^{N-1}\langle Z_m, Z_j\rangle\right) \tag{3.9c}$$

where $Z_i = (x_i, y_i)$ and $\langle \mathbf{A}, \mathbf{B}\rangle$ represents the Euclidean dot product between two vectors $\mathbf{A}$ and $\mathbf{B}$ defined as

$$\langle \mathbf{A}, \mathbf{B}\rangle = ||\mathbf{A}||\,||\mathbf{B}||\cos(\theta_{AB}) \tag{3.10}$$

and θ_{AB} is the angle between $\mathbf{A}$ and $\mathbf{B}$. For the special case when $\mathbf{A} = \mathbf{B}$ one obtains $\langle \mathbf{A}, \mathbf{A}\rangle = ||\mathbf{A}||^2 = l^2$. Because the turning angle for each step is chosen independently from any other, the distribution of turning angles may be considered as independent random variables. This is important since one can then exchange the order of the sum and expectation in equation (3.9c) as well as use the fact that the expectation of a product of independent random variables is the product of their expectations.[5] This together with the definition of the dot product in equation (3.10) produces the following

$$E\left(R_N^2\right) = \sum_{m=1}^{N} E\left(l^2\right) + 2\sum_{m>j}^{N-1} E(l_m)E(l_j)E(\cos(\theta_{mj})) \tag{3.11a}$$

$$= Nl^2 + 2l^2\sum_{m>j}^{N-1} E\left(\cos\left(\sum_{k=j}^{m-1}\theta_k\right)\right)$$

At this point the complex form of $\cos(\phi)$ is employed so that the mean cosine appears in the expression; bear in mind that the mean sine was assumed to be zero. The aforementioned complex form is defined with the exponential function where i is the imaginary unit: $\cos(\phi) = (e^{i\phi} + e^{-i\phi})/2$. The fact that $s = 0$ simplifies the calculation significantly and after using the independence of each turning angle and gathering terms one arrives at the formula in equation (3.7).[6] In addition, if N is large then the second term is negligible relative to the first and an approximation for the mean squared displacement can be obtained,

$$E\left(R_N^2\right) \approx Nl^2\left(\frac{1 + c(\kappa)}{1 - c(\kappa)}\right) \tag{3.12}$$

[5] This is to say the covariance between any two of these variables is zero.

[6] For the full derivation see the appendix in [52].

where $c(\kappa)$ is the mean cosine in terms of κ, the concentration parameter. The above exercise of obtaining an explicit expression for the mean squared displacement was necessary for several reasons: (1) the expression in equation (3.12) is given in terms of the step length, number of steps and the mean cosine, all of which are measurable; (2) no such expression for the mean dispersal distance is obtainable in terms of the aforementioned quantities, (3) under certain conditions a simple relationship between the mean squared displacement and the mean dispersal distance exists and (4) the diffusion coefficient which informs one of the dispersal patterns of individuals in biological systems is characterised by the probability distribution of the beeline distance where the expectation of this distribution is the mean displacement distance $E(R_N)$.

In order to use the aforementioned distance metrics one requires knowledge of the step length l of the typical walker, a merozoite in this case. In most scenarios the step length would be determined experimentally by discretising the paths travelled by one more walkers into a number of steps of finite length to obtain a distribution of step lengths. However, the data available were not reliable enough to obtain accurate estimates of these step lengths for individual merozoites. As such, a constant length of 0.4 µm per second was chosen as a realistic estimate since it is smaller than the average radius of a merozoite (0.5 µm). As mentioned, the mean squared displacement is related to the mean dispersal distance by a simple relation provided certain conditions are met [55]. The random walk must be an unbiased, correlated random walk with zero turning angle ($\theta_0 = 0$). Furthermore, the number of steps N must be sufficiently large so that the location coordinates are independently normally distributed with the same variances – both of these assumptions are justified here.

Consider a path beginning at the origin consisting of N steps and which has two components X_N and Y_N which correspond to the Cartesian coordinates of the end point (x_N, y_N). The square of the beeline distance (absolute displacement) R_N is then simply given by $\mathrm{R}_N^2 = X_N^2 + Y_N^2$. Because one is dealing with a correlated random walk the path has no preferred direction or orientation when N is large so that the expectations of the location coordinates are zero, i.e $E(X_N) = E(Y_N) = 0$. In addition, because they share the same variance one has that $E\left(X_N^2\right) = E\left(Y_N^2\right) = \sigma^2$ for some $\sigma > 0$. One can obtain an equation for σ in terms of the expectation of the square of the beeline distance, the mean squared displacement, as follows

$$E\left(R_N^2\right) = E\left(X_N^2\right) + E\left(Y_N^2\right) = 2\sigma^2 \tag{3.13a}$$

$$\therefore \sigma = \sqrt{\frac{E(R_N^2)}{2}} \tag{3.13b}$$

To complete the result one needs to make the following substitutions: Let

$$\overline{X_N} = \frac{X_N}{\sigma} \text{ and } \overline{Y_N} = \frac{Y_N}{\sigma}$$

so that $\left(\overline{X_N}, \overline{Y_N}\right) \sim \mathcal{N}(0,1)$.[7] Since both new variables subscribe to a standard normal distribution one can describe the square root of the sum of their squares as a chi distribution with two degrees of freedom (one for each variable) [55]. Let the square root of their sum be d so that $d \sim \chi_2$ and

$$d^2 = \frac{R_N^2}{\sigma^2} \tag{3.14a}$$

$$\therefore R_N = \sigma \cdot d \tag{3.14b}$$

$$\implies E(R_N) = \sigma E(d). \tag{3.14c}$$

The expectation of a chi-distributed variable with k degrees of freedom is computed in terms of the Γ-function and is given by

$$E(d) = \frac{\sqrt{2} \cdot \Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}.$$

For the case when $k = 2$ and using the definitions of the Γ-function and the error function,[8] the expected value of the chi-distributed variable is

$$E(d) = \sqrt{\frac{\pi}{2}}$$

Using equations (3.13b) and (3.14c) one obtains the mean dispersal distance in terms of the mean squared displacement,

$$E(R_N) = \frac{\sqrt{\pi \cdot E(R_N^2)}}{2}. \tag{3.15}$$

Combined with equation (3.12) one obtains

$$E(R_N) \approx \sqrt{\frac{Nl^2\pi}{4}\left(\frac{1+c(\kappa)}{1-c(\kappa)}\right)}. \tag{3.16}$$

Due to the limitations of the available data, video micrographs showing the rupture of mature schizonts and subsequent invasion of released merozoites, the only measurable quantity was the absolute displacement or beeline distance of each merozoite from the moment of rupture. Twenty merozoites in total were tracked for a period of 181 seconds where the start and end points where noted and the beeline distance recorded. These were then averaged to obtain the mean dispersal distance $E(R_N)$. This in turn was used

[7] The notation $\sim \mathcal{N}(0,1)$ means that the variables are normally distributed with zero mean and a standard deviation of one. This is called the standard normal distribution.

[8] The one-sided error function is defined as the integral $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$ while the Γ-function has the property of $\Gamma(x) = (x-1) \cdot \Gamma(x-1) = (x-1)!$.

Table 3.1: Movement parameters of the unbiased, correlated random walk model.

Steps - N[a]	$\mathbf{E(R_N)}$ **(µm)**	$\mathbf{E(R_N^2)}$ **(µm^2)**	l[b] **(µm)**	κ[a]	$c(\kappa)$[a,c]
181	8.00	81.49	0.40	1.09	0.48

[a] These three parameters are dimensionless.

[b] This length is the distance a merozoite covers in a single step/second.

[c] $c(\kappa)$ represents the mean cosine of the von Mises distribution.

to calculate the mean squared dispersal $E(R_N^2)$ by using equation (3.15). One could then easily obtain the mean cosine (c) and concentration parameter (κ) of the von Mises distribution from equation (3.16). Recall that the mean angle of the distribution was assumed to be zero ($\theta_0 = 0$). Table 3.1 provides the values of the necessary parameters just discussed associated with the movement of merozoites from the moment of rupture.

3.2 Red Blood Cells as Barriers

The physical interaction of a merozoite with a RBC is a necessary but not sufficient condition for invasion as has been stated earlier. In the context of a random walk model, this means that RBCs act as barriers for diffusing merozoites. Recall that a barrier may change the direction in which a walker takes its next step or cause the walker to stop moving altogether for the duration of the simulation. What is interesting in the current model is that RBCs act as both reflective and absorbing barriers depending on whether a merozoite has successfully invaded or not.

At this point an important simplifying assumption must be noted: the possibility of a merozoite interacting with or "bumping into" another merozoite was ignored. In other words, two or more merozoites were allowed to occupy the same space at the same time in the two-dimensional plane. This assumption was deemed reasonable given the objectives of the stochastic model. Importantly, this assumption did not apply to any RBCs present in the simulations since they were kept stationary; they were required to be distinct, non-overlapping circles. The ratio of the sizes merozoites and RBCs in the simulations was 1:7 with a merozoite and RBC having radii of 0.5 µm and 3.50 µm, respectively.

To determine whether a merozoite and a RBC had made contact a measure of the distance between their respective centres was utilised – if this distance was smaller than the sum of their radii then an

interaction had taken place. The formula to calculate this distance is based on elementary trigonometry:

$$D_s = \sqrt{(x_m - x_r)^2 + (y_m - y_r)^2} \tag{3.17}$$

where (x_m, y_m) and (x_r, y_r) are the co-ordinates of the centres belonging to a merozoite and a RBC at any given time, respectively. Furthermore, let each merozoite have a radius of R_m and each RBC a radius of R_r. The condition for an interaction to occur would then be written as: $D_s < R_m + R_r$.

One may now consider what should happen when an interaction does in fact occur. Recall that a merozoite-erythrocyte complex (MEC) first had to be formed before a merozoite could successful invade a RBC; the parameter responsible for this invasion rate was k_2. A consequence of forming a MEC was that a merozoite could detach from its host RBC instead of invading it; this rate of detachment was quantified by the constant k_{-1}. In other words, once a merozoite makes contact with a RBC and forms a MEC, there are only two possible outcomes: successful invasion or detachment. The possibility of the merozoite remaining attached indefinitely was not considered. To be used in the random walk model one must transform the rates for successful invasion and detachment into probabilities for their respective processes. Keep in mind that both of the aforementioned rate constants are first-order rates which means they have the same units. This is crucial when determining the formula for their respective probabilities since one requires their sum; this sum is interpreted as the combined rate of the two processes. This combined rate allows one to define P_{k_2}, the probability of successfully invading a RBC as

$$P_{k_2} = \frac{k_2}{k_{-1} + k_2}. \tag{3.18}$$

One should recognise that because there are exactly two possible outcomes based on a constant probability of success, the fate of a merozoite attached to a RBC may be described by a Bernoulli distribution with the probability given in equation (3.18).[9] A Bernoulli distribution is a special case of the binomial distribution in that it only has one trial. Going one step further one sees that if the merozoite is unsuccessful it will detach and may again try to invade another or the same RBC with the same probability of success as the first attempt. This means that any single merozoite will experience a number of identically, independent dichotomous trials, each of which is Bernoulli distributed. Therefore, the probability of a merozoite successfully invading a merozoite after a finite number of attempts is the sum of that same number of Bernoulli distributions. This is in fact a negative binomial distribution since the order of successes and failures matter unlike the regular binomial distribution. It is logical when considering that a merozoite

[9] In other words, the fate of a merozoite attached to a RBC is a random variable with a Bernoulli distribution.

can only successfully invade once in its lifetime after one or more unsuccessful attempts; it may of course be successful on the first attempt as well.

Because the event of success will always be the final event, the negative binomial distribution gives one the probability of a single success after a fixed number of failures. Here the number of failures would be the number of detachments any particular merozoite experiences. In general, a random variable V is distributed with a negative binomial distribution if $V \sim \mathcal{NB}(r, p)$ where r and p are the number of failures before the first success and the probability of success, respectively. The probability mass function (PMF) of V is given by

$$P(V = k) = \begin{pmatrix} k + r - 1 \\ k \end{pmatrix} p^k (1 - p)^r \tag{3.19}$$

where $k \in \mathbb{N}$ is the number of successes. In other words, equation (3.19) gives the probability that there are k successes after r failures where $k + r$ is the total number of trials. Since $k = 1$ is the only biologically realistic case, the probability that a merozoite successfully invaded a RBC after r unsuccessful attempts would be the following:

$$P(V = 1) = rp(1 - p)^r \tag{3.20}$$

In this context we have that the probability of success $p = P_{k_2}$ from equation (3.18). This calculation can be done for each merozoite present in any simulation of the random walk model. If sufficiently many merozoites successfully invaded during the course of any one simulation then a distribution of probabilities could be obtained and the expected number of detachments an average merozoite will experience determined.

The implementation of the aforementioned Bernoulli distributions for each merozoite was done as follows: once a merozoite came into contact with a RBC a check was performed using equation (3.18) to determine if that merozoite did in fact invade the host RBC successfully. This check involved sampling from a uniform distribution and then comparing that sample to the probability of invasion. If this value was less than or equal to the probability of invasion, invasion was considered successful and the merozoite simply remained attached to the now invaded RBC for the remainder of the simulation period. If it was unsuccessful – this is to say it detached from the host RBC – it moved back to the position it was in before it made contact with the specific RBC. Being in this "new" position it was able to again choose a turning angle from the von Mises distribution for its next step. If this next step resulted in a second interaction with the same or any other RBC, the procedure just described would simply repeat until it successfully invaded or succumbed to natural decay; the topic of the next section.

3.3 Natural Decay of Merozoites

It is well-known that merozoites have a limited lifespan when they are free in the extracellular environment. In the deterministic model merozoites decay at a rate proportional to their current population, i.e. exponentially. A species N is subject to exponential decay if its population is governed by the following equation:

$$N(t) = N_0 e^{-kt} \tag{3.21}$$

where k and N_0 are the decay constant and the initial size of the population, respectively. The key concept to consider in the context of a random walk is that exponential decay process assumes that every individual (merozoite) has the same probability of dying at every time step regardless of how many steps it has already taken – this is because exponential decay is a purely probabilistic process and is characteristic of a uniform probability distribution. One can extend this idea into a probability distribution with a probability mass function (PMF) which is simply a constant p, the probability of death at any given time step. What this means is that the PMF is simply a constant, let this function be $g(t)$ so that one has $g(t) = p$. Finding an expression for p can be done by using the half-life ($t_{0.5}$) of an average merozoite. More specifically, the idea of a half-life is that a randomly chosen merozoite has a 50 % chance of surviving up to that point. Said differently, the cumulative chance of a merozoite dying from the moment of rupture to the half-life is 50 % which gives one the following,

$$\sum_{t=1}^{t_{0.5}} g(t) = \frac{1}{2}, \tag{3.22a}$$

$$p\sum_{t=1}^{t_{0.5}} 1 = \frac{1}{2}, \tag{3.22b}$$

$$\therefore\ p = \frac{1}{2 \cdot t_{0.5}}. \tag{3.22c}$$

Note that a lower limit of $t = 0$ could also have been used but it would not influence the sum and therefore the end result; this is true for all the summations in this section. Equation (3.22c) tells one that the probability of a merozoite dying at any given time step is equal to the reciprocal of twice the half-life. This means that, by definition of a uniform distribution such as this, the maximum time a merozoite could theoretically survive, denoted by T, is precisely twice the half-life. This result is obtained by looking at the cumulative distribution function (CDF) of $g(t)$ with T as the upper limit. The CDF of any discrete probability distribution is the probability that the variable (merozoite lifespan in this case) is smaller than or equal to some value. If this value is set to T then the sum of all the probabilities must

be equal to one as indicated in equation (3.23a):

$$\sum_{t=1}^{T} g(t) = 1, \tag{3.23a}$$

$$\frac{1}{2 \cdot t_{0.5}} \sum_{t=1}^{T} 1 = 1, \tag{3.23b}$$

$$\frac{1}{2 \cdot t_{0.5}} T = 1, \tag{3.23c}$$

$$\therefore \; T = 2 \cdot t_{0.5}. \tag{3.23d}$$

Another useful property of a probability distribution is it's expected value which is also called the first moment of that probability mass function.[10] In this context the first moment would be the life expectancy of the average merozoite which will be noted by $\bar{t}$. To calculate the expected value one multiplies each possible outcome, the number of seconds a merozoite may survive, with the corresponding probability and sum the individual values – equation (3.24) shows this result.

$$\bar{t} = \sum_{t=1}^{T} t \cdot g(t), \tag{3.24a}$$

$$= \frac{1}{2 \cdot t_{0.5}} \sum_{t=1}^{T} t, \tag{3.24b}$$

$$= \frac{1}{2 \cdot t_{0.5}} \frac{T(T+1)}{2}, \tag{3.24c}$$

$$= \frac{1}{2} + t_{0.5}, \tag{3.24d}$$

$$\approx t_{0.5}, \tag{3.24e}$$

The sum in equation (3.24b) has a very well-known result which makes calculating p straightforward.[11] Equation (3.24d) uses the expression for T in terms of the half-life (equation (3.23d)). The approximation in equation (3.24e) is justified since both T and $t_{0.5}$ are at least two orders of magnitude greater than unity making it reasonable to drop the constant term. This expectancy ($\bar{t} = t_{0.5}$) may seem reasonable but, one must consider the biology of the phenomenon being modelled and be able to adequately justify the assumption of a uniform probability of death. What the above analysis implies is that a merozoite which has survived, for example, three minutes is just as likely to die in the next second than a merozoite which has survived for seven minutes.

[10] The concept of the expected value of a random variable was introduced in section 3.1.1 except now one is dealing with a discrete random variable as opposed to a continuous one.

[11] The sum of integers from 1 to n is $\frac{n(n+1)}{2}$.

This is simply not realistic since merozoites are known to be very sensitive to the extracellular conditions of an experimental culture which suggests that merozoites have an increasing probability of death over time. The simplest way to model this is to assume a linear dependency with time. In other words, the chance of death increases with each step a particular merozoite takes.

3.3.1 Time-dependent Death Rate of Merozoites

The idea of a linearly increasing probability of death is rather intuitive. Suppose the initial probability of death is ρ, that is to say a merozoite will die with probability ρ within the first second of being released from a rupturing schizont. Suppose now that the merozoite does not die but survives the first second after release. The chance that it will die between seconds one and two is now twice as great as it was for the previous second, i.e 2ρ. Continuing with this pattern one finds that the probability of a merozoite dying between seconds n and $n+1$ is $(n+1)\rho$. Let the probability of death at each step be denoted by the time-dependent function $f(t)$ defined as

$$f(t) \;=\; \rho \cdot t \tag{3.25}$$

where t is the time in seconds. One can then again use the well-known sum from earlier as well as the half-life as before to find an expression for the initial probability of death (ρ),

$$\sum_{t=1}^{t_{0.5}} f(t) \;=\; \frac{1}{2}, \tag{3.26a}$$

$$\rho \sum_{t=1}^{t_{0.5}} t \;=\; \frac{1}{2}, \tag{3.26b}$$

$$\therefore\ \rho \;=\; \frac{1}{t_{0.5}\,(t_{0.5}+1)}. \tag{3.26c}$$

While the probability mass function in equation (3.25) is arguably more realistic than that associated with an exponential decay model, the two models both give the required 50 % probability of surviving to the half-life even though the methods of calculating the relevant probabilities at each step are different. As before with $g(t)$, calculating the CDF would allow one to find a new estimate for the maximum amount of time (in seconds) a merozoite could theoretically survive. Using the definition of the cumulative distribution function for a discrete random variable (see equation (3.23a)) one obtains a quadratic

equation in T,

$$\sum_{t=1}^{T} f(t) = 1, \tag{3.27a}$$

$$\frac{1}{t_{0.5}\,(t_{0.5}+1)} \sum_{t=1}^{T} t = 1, \tag{3.27b}$$

$$\frac{1}{t_{0.5}\,(t_{0.5}+1)} \frac{T(T+1)}{2} = 1. \tag{3.27c}$$

Solving equation (3.27c) for T, taking only the positive root and using the same approximations as before one finds that

$$T = \sqrt{\frac{1}{4} + 2 \cdot t_{0.5}(t_{0.5}+1)} - \frac{1}{2}. \tag{3.28a}$$

$$T \approx \sqrt{2} \cdot t_{0.5}. \tag{3.28b}$$

Note that $T > t_{0.5}$ as required.[12] One may again calculate the expected value of the distribution ($\bar{t}$) which is obtained with the help of another well-known sum.[13]

$$\bar{t} = \sum_{t=1}^{T} t \cdot f(t), \tag{3.29a}$$

$$= \frac{1}{t_{0.5}\,(t_{0.5}+1)} \sum_{t=1}^{T} t^2, \tag{3.29b}$$

$$= \frac{1}{6} \cdot \frac{T(T+1)(2T+1)}{t_{0.5}\,(t_{0.5}+1)} \tag{3.29c}$$

$$\approx \frac{2}{3}\sqrt{2} \cdot t_{0.5} \tag{3.29d}$$

$$\approx 0.94 \cdot t_{0.5}. \tag{3.29e}$$

Equation (3.29e) provides an interesting result: the expected lifespan of an average merozoite under this model is slightly less than its half-life. Finding this particular value was the final step before this model of natural death could be integrated into the random walk model. The method of implementation was simple: the inverse transform sampling method was used by inverting the appropriate cumulative distribution function, randomly sampling a cumulative probability from zero to one (call this probability λ), and then calculating the merozoite lifespan corresponding to the sampled probability. Recall that the cumulative distribution function is the probability that a random variable will take a value smaller than or equal to some number, say t^* – the unknown lifespan which needs to be calculated. Using $f(t)$ from

[12] It can be shown that the error of this approximation is independent of the half-life and is $\frac{1}{\sqrt{2}} - 1 \approx -0.29$.

[13] The sum of squared integers from 1 to n is $\frac{n(n+1)(2n+1)}{6}$.

equation (3.25) one has that

$$\sum_{t=1}^{t^*} f(t) \;=\; \lambda, \tag{3.30a}$$

$$\Rightarrow\; t^* \;=\; \sqrt{\frac{1}{4} + 2 \cdot \lambda \cdot t_{0.5}(t_{0.5}+1)} - \frac{1}{2}. \tag{3.30b}$$

Note that this expression is very similar to equation (3.28a) except now a factor of λ is present in the second term inside the square root. Since λ is defined to be between zero and one it ensures that t^* will always be smaller than or equal to T which is necessary since T is the maximum lifespan of a randomly chosen merozoite.[14] Therefore, if a simulation requires that 50 merozoites be dispersed then λ would be sampled (with replacement) 50 different times from zero to one. Each of these probabilities would then be used to calculate the corresponding lifespan t^* from equation (3.30b). In other words, each merozoite was assigned a lifespan of pre-determined length before the any simulation was started. The number of merozoites per simulation were varied in order to compare the results with the experimental data.

3.4 Simulation of the Random Walk Model

Constructing the two-dimensional simulation space requires knowledge of the hematocrit since this informs one of how far apart RBCs should be from one another; a lower hematocrit means RBCs would be spaced further apart than if the hematocrit were higher. Briefly, 100 uRBCs were placed on a two-dimensional plane spaced according to the hematocrit used in the experiment from which the data being predicted was obtained. To best illustrate a typical realisation, an example will be presented. Consider a replication cycle in a static environment which has a 3 % initial parasitemia. Three of the 100 uRBCs are replaced with mature schizonts, each of which will release 16 merozoites at the beginning of the simulation. Importantly, choosing which uRBCs to replace with mature schizonts was done randomly for each realisation. When dealing with conditions seen in suspension cultures, instead of replacing the appropriate number of uRBCs with mature schizonts as just described in the static case, the original 100 uRBCs remained where they were and only a number of *merozoites* were placed, in a random manner, in between the original uRBCs. Using the same example as before, a 3 % initial parasitemia meant that 48 merozoites were placed on the two-dimensional plane in between the 100 uRBCs. The simulation then begins with individual merozoites all at different locations instead of groups of 16 merozoites all dispersing from the same location.

[14] Mathematically this is written as: $0 \leq \lambda \leq 1 \implies t^* \leq T$.

To summarise, the stochastic model describing the transition period of *P. falciparum* is an unbiased, correlated random walk with RBCs acting as both reflective and absorbing barriers. Each merozoite was assigned a randomly chosen lifespan from a probability distribution based on a linearly increasing probability of death with each time step. The path of each merozoite was assumed to be governed by the von Mises distribution with mean angle zero where the sampled value for each step was chosen to be the turning angle for the following step; this gives the model its unbiased, correlated properties. If a merozoite made contact with RBC a check was performed against a uniform distribution to determine whether a successful invasion had occurred. In this event the merozoite remained stationary for the rest of the simulation period. If the invasion was unsuccessful, the merozoite detached from the corresponding RBC and was allowed to continuing moving from its position before the interaction except now it is facing the opposite direction, away from the RBC it just tried to invade. Recall that in a correlated random walk the next step a walker takes is dependent on the previous step's direction. This was chosen over a biased random walk where each step's direction is independent of any other since there were barriers, RBCs in this case, present in the environment. If merozoites were biased in their movements then when interacting with a RBC the merozoites would not alter their direction after an unsuccessful invasion.

It was stated in section 1.3.1 that accommodating multiple invasions in the stochastic model was relatively simple since all that was required was for all RBCs to have no memory of their invasion history, i.e how many times they were successfully and unsuccessfully invaded. As a result, any RBC could be invaded by more than one merozoite throughout the course of any simulation. Lastly, if a merozoite failed to invade a RBC before its assigned lifespan was reached, it was defined to have perished and remained stationary at the position of its final step for the remainder of the simulation period.

4. The Available Data

This chapter presents the data available for the fitting and validation of the deterministic and stochastic mathematical models constructed in chapters 2 and 3. As pointed out in the introductory chapter, data pertaining to the phenomenon of multiply-invaded pRBCs was found to be sparse and, at times, incomplete. Because of this, two sets of data were primarily used in the course of model fitting and validation while other, smaller datasets were used for less stringent validation. The consequences and implications of the data and methods used in each study presented here are explored in the discussion. Importantly, figures of the experimental and clinical data in this chapter are presented in such a way which included all of the available data of that particular study in a clear and consistent manner.[1] In addition, studies which obtained estimations for all or a subset of the parameters present in the mathematical models constructed in chapter 2 are also included.

4.1 Baum et al. (2003)

The research conducted by Baum and colleagues in their paper, "*Erythrocyte invasion phenotypes of Plasmodium falciparum in the Gambia.*", was published in 2003 and was primarily concerned with determining which RBC receptors are used most often during invasion of different *P. falciparum* field isolates, while the degree of inhibition of invasion was the secondary objective [17]. The researchers were looking at interactions between RBC receptors and merozoite ligands. Single-cycle, "dilution factor" invasion assays were carried out for each patient isolate to determine the re-invasion parasitemia and multiplication rate. Briefly, a suspension of diluted, *asynchronous, unlabelled* donor pRBCs was added to an equal volume of *labelled*, O^+ uninfected target RBCs.[2] This mixture was diluted with complete RPMI 1640 to obtain an hematocrit of 1 % and a range of starting parasitemias. After one replication cycle (48 to 52 hours), the number of *labelled* pRBCs were counted and used to determine several indicators of disease severity.

Multiply-invaded pRBCs The number of multiply-invaded pRBCs were counted in the peripheral blood smears of each isolate and used to calculate the *in vivo* selectivity index. Recall that the selectivity index is a measure used to determine whether there were more multiply-invaded pRBCs than would be

[1] In other words, the format of the figures presented here is independent of the source of the data.

[2] All isolates were diluted by a factor of two before being added to suspension of uRBCs.

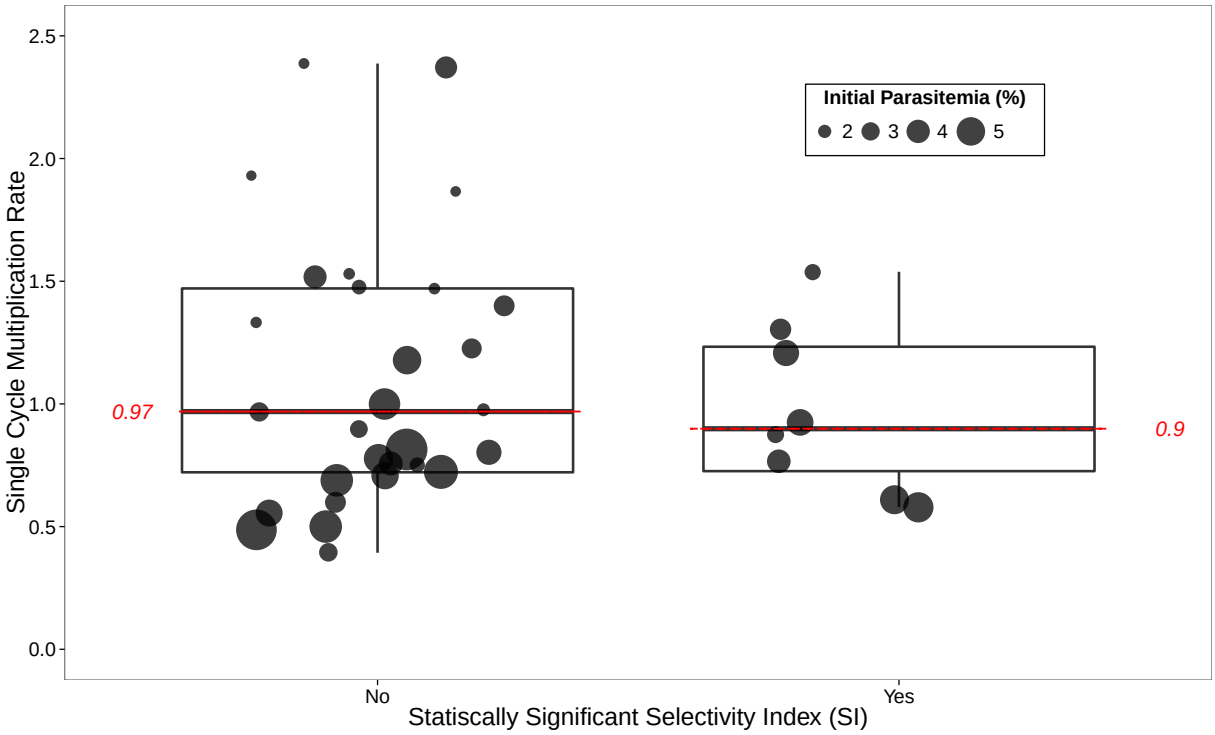

Figure 4.1: Scatterplots overlay box-and-whisker plots of the multiplication rates (dimensionless) stratified by the statistical significance of the isolates' selectivity indices [17]. Each point represents a single patient isolate ($n = 38$). The medians of the multiplication rate for those isolates with and without a statistically significant selectivity indices were 0.90 and 0.97, respectively and are denoted by the red, dashed lines ($p > 0.4$). The size of each point is the relative initial parasitemia associated with that particular experiment. The initial parasitemia was found to be negatively correlated with the multiplication in both groups (-0.59 and -0.66) while only the first group (statistically insignificant selectivity index) had a significant correlation; $p < 0.01$ compared to $p > 0.04$. Significance tests were performed with unpaired student t-tests.

expected by chance alone. Unfortunately, only a subset of the calculations involving multiply-invaded pRBCs were accessible due to the inclusion criteria being defined as those isolates which showed a statistically significant selectivity index. The number of multiply-invaded pRBCs were not however, counted for the re-invasion parasitemia, i.e no data for multiply-invaded pRBCs were recorded for the *in vitro* cultures. This omission is clear because pRBCs were counted at the mature schizont stage and not the ring or early trophozoite stages when multiply-invaded pRBCs can be visually confirmed. As a result, it was not possible to fit the more complex mathematical models to any multiple invasion data from this particular study. Most importantly, the experiments were performed under static culture conditions which means that no mixing or agitation was introduced at any stage during the experiments.

The following calculations were made to determine the initial conditions which will be used in the

mechanistic mathematical models. The concentrations of parasitised and uninfected RBCs – expressed as cells/µl and denoted by square brackets – were calculated from the protocol utilised in this study. Briefly, 10 µl (approximately 10^7 cells) of both the target cell and parasitised donor cell suspensions were added to 180 ml of RPMI 1640 complete medium for a total volume of 200 ml, a total concentration of 10^5 cells/µl and a final hematocrit of 1 %. The parasitised donor cell suspension contained both uRBCs and pRBCs since these came directly from patients. Let p_i denote the proportion of those RBCs which were infected for the i-th patient. For example, if a patient presented with a 5 % parasitemia then p_i for this patient would be 0.05. The target cell suspension was the population of O^+ uninfected RBCs as discussed earlier. Therefore, the concentration of pRBCs was

$$[\text{pRBC}_i] = \frac{p_i \times 10^7}{200} \tag{4.1a}$$

$$= p_i \cdot 5 \times 10^4 \tag{4.1b}$$

Similarly, the uRBC concentration was calculated as

$$[\text{uRBC}_i] = \frac{(1 - p_i) \times 10^7 + 10^7}{200} \tag{4.2a}$$

$$= 10^5 - p_i \cdot 5 \times 10^4 \tag{4.2b}$$

$$= 10^5 - [\text{pRBC}_i] \tag{4.2c}$$

Adding equations (4.1b) and (4.2b) gives one a total of 10^5 cells/µl which is the approximate concentration of RBCs in a suspension at 1 % hematocrit [17]. For the purposes of fitting the mathematical model one requires the initial concentration of merozoites, not the concentration of pRBCs at the start of the experiment. Multiplying the total number of pRBCs by 16 gives one the required input:

$$\left[m_0^i\right] = 16 \cdot [\text{pRBC}_i] \tag{4.3}$$

4.2 Clough et al. (1998)

The objective of this particular study, "*The role of rosetting in the multiplication of Plasmodium falciparum: rosette formation neither enhances nor targets parasite invasion into uninfected red cells.*" published in 1998 by Clough et al. [16], was to investigate and quantify the effects of rosette formation on the number of multiply-invaded pRBCs and the multiplicative ability of *P. falciparum* merozoites under static conditions as well as those observed in the microvasculature (suspension culture conditions) in an *in vitro* setting. A rosette is simply an aggregation formed when a mature pRBC spontaneously

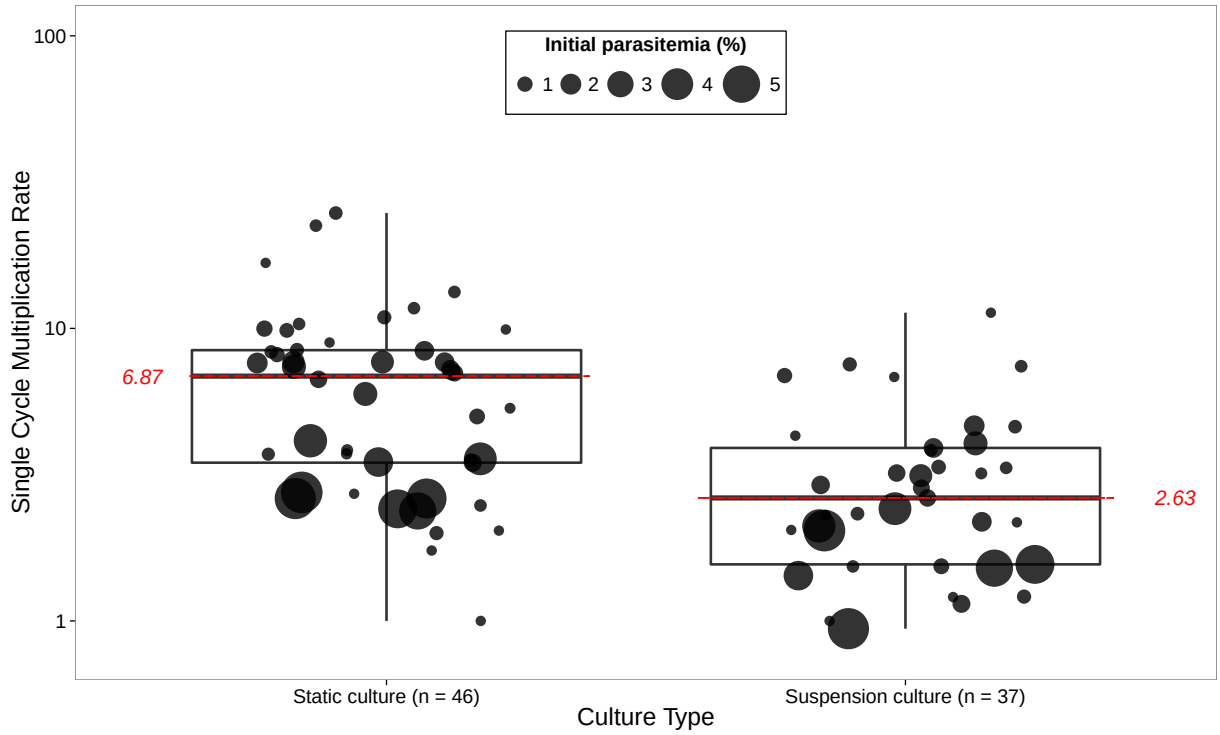

Figure 4.2: Scatterplots overlayed with box-and-whisker plots for the multiplication rates (dimensionless) stratified by culture from Clough and colleagues [16]. Cultures were conducted in parallel, one in static conditions and the other with agitation (125 rpm in an orbital shaking incubator) over a range of initial parasitemias. The y-axis was log-transformed since the data span two orders of magnitude. The medians of the multiplication rate for static and suspension cultures were 6.87 and 2.63, respectively. These were statistically significant ($p < 0.01$) and are denoted by the red, dashed lines. Each point denotes a single experiment while the size of each point is the relative initial parasitemia associated with that particular experiment. In both culture types, the initial parasitemia was found to be negatively correlated with the multiplication rate (-1.00 ($p < 0.03$) and -0.45 ($p < 0.04$) for static and suspension cultures, respectively.). Significance tests were performed with unpaired student t-tests.

binds to two or more uRBCs and has been suggested to be an important factor which contributes to severe disease [56, 16, 7, 20]. Rosetting and non-rosetting parasite lines were selected from the same, laboratory-adapted clone and cultured in parallel – one in static and one in suspension – to compare multiplication rates. The suspension cultures were mixed at 125 rpm in an orbital shaking incubator, sufficiently high to mix healthy and invaded RBCs but, not enough to disrupt rosettes. Briefly, single-cycle invasion assays were performed where *synchronised*, mature pRBC were added to uninfected A^+ target RBCs; the range of starting parasitemias was between 0.1 % and 5 %. The resulting mixture was diluted with complete RMPI 1640 to a final volume of 2 ml with an hematocrit of 2 % to 5 %.[3] After 15 to 20 hours, the number of pRBCs were counted (including those which had been multiply-invaded) and used

[3] The specific, final hematocrit was not given.

to determine the re-invasion parasitemia, the multiplication rate and the percentage of multiply-invaded pRBCs in both rosetting and non-rosetting pRBCs.

The experimental data are summarised in figure 4.2 where the multiplication rate after a single cycle for the two sets of cultures is shown. Note that differences between rosetting and non-rosetting parasite lines was not taken into account since cytoadherence is not explicitly being considered in the current work. That being said, a surprising observation was the higher multiplication rates within static cultures compared to those in suspension. This can be seen by the red, dashed lines indicating the mean PEMR of each culture. This seems to be the opposite of what one would expect given the arguments and data from several other studies discussed in section 1.2.1. A simple explanation to reconcile these differences are the speeds at which the cultures were agitated; this will be elaborated on in chapter 6. A second observation is the inverse relationship between the multiplication rate and the initial parasitemia, i.e the larger points are more often than not below the smaller points in figure 4.2. This is explained in part by the absolute numbers of pRBCs and partly by physical and spatial constraints.

Multiply-invaded pRBCs The number of multiply-invaded pRBCs could be determined for this study given that the cells were counted at the ring or early trophozoite stage when multiple rings inside a single pRBC could be clearly seen. This is in contrast to the methods in section 4.1 where multiply-invaded pRBCs were not able to be distinguished from singly-invaded pRBCs. Unfortunately, a detailed breakdown of the percentage of multiply-invaded pRBCs for each separate experiment was not given. Instead, the percentage of pRBCs which had been invaded by one, two or more than two merozoites was determined for static and suspension cultures in each of eight experiments (four per culture). Figure 4.3 illustrates that suspension cultures produced fewer multiply-invaded pRBCs; an observation consistent with that in other studies [20, 26]. This observation, together with the lower multiplication rate indicates that a significantly lower proportion of the available merozoites successfully invaded compared to those in static cultures. This can once again be attributed to the speed at which the cultures were shaken.

4.3 Kinetic parameters

Before discussing the kinetic parameters an important issue regarding the current work should be reiterated: while estimates were available for most of the rate constants present in the base model (discussed below in detail), there was no direct time-series data available to fit the mechanistic mathematical models. A

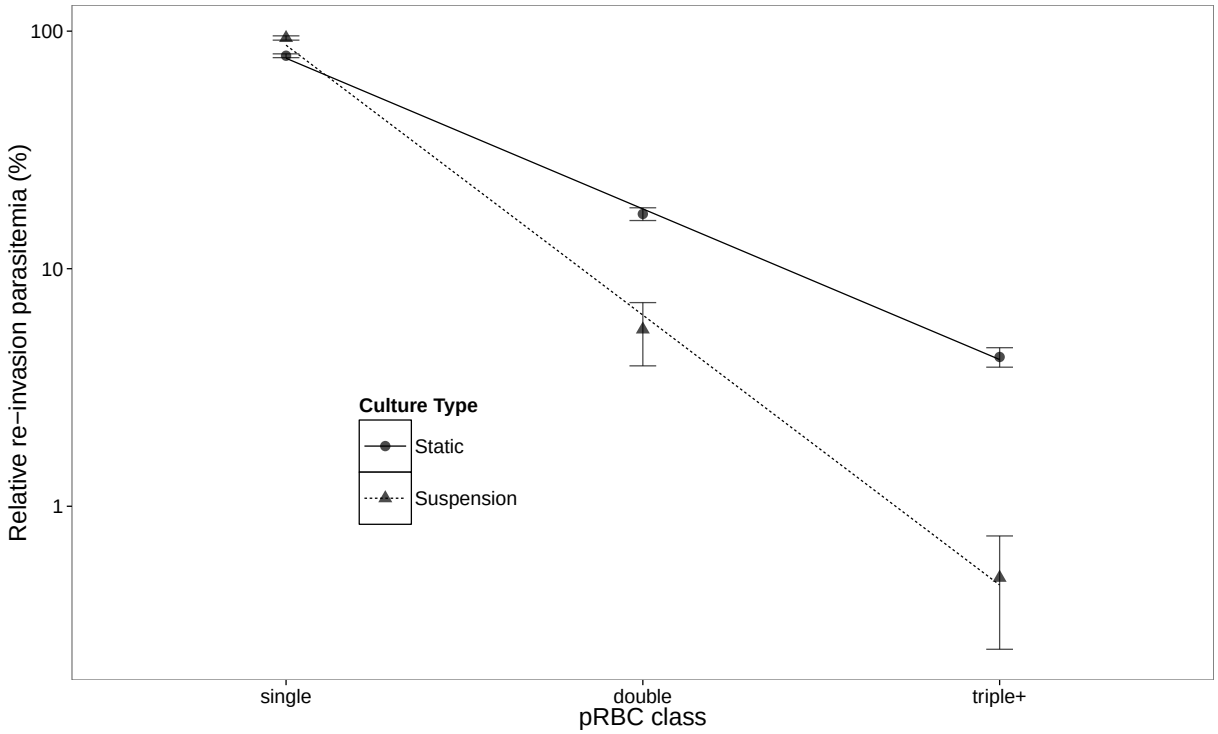

Figure 4.3: Multiple invasion data from Clough and colleagues. The relative parasitemia of each invasion class for static and suspension cultures is shown by points and triangles, respectively. The third invasion class (triple+) denotes the pooled number of pRBCs which were invaded by three or more merozoites. The percentages for single, double and triple+ pRBCs in static cultures was 78.80 %, 17.00 % and 4.25 %, respectively. For suspension cultures these were 93.60 %, 5.55 % and 0.50 %, respectively. Error bars denote the standard error (SE) associated with four experiments each for static and suspension cultures. Linear regression lines for static and suspension cultures are given by the solid and dashed lines whose gradients are -0.63 and -1.14, respectively.

range of realistic values for the parameters is needed to proceed with the model construction process described in section 1.3. Importantly, confirming the units of the required free parameters is essential since this informs one of which estimates may be directly compared and which may not.

The estimates obtained from the literature for three of the four parameters present in the base model are shown in table 4.1. The primary goals of the studies cited in table 4.1 revolved around dissecting one or more distinct aspects or periods of the transition period at a microscopic level and investigating which factors have the most influence over some particular stage of the invasion process. Researchers specifically studying the kinetics of schizont rupture and merozoite dispersal have attempted to quantify parameters such as the initial velocity with which merozoites are ejected and how long it takes the average merozoite to bind and successfully invade a susceptible erythrocyte [58, 21, 46, 57]. Often the experimental design is such that invasion rates between physiologically normal RBCs and those

Table 4.1: Kinetic parameter estimates obtained from the literature.

Effective parameter	Units	Initial estimate (range)	Description	Reference
$1/k_1^*$	sec	121 (21 to 258) 14 (8 to 34) 48.36 (9 to 112)	Time from egress to initial contact and re-orientation	[57] [45] [46]
$1/k_2$	sec	528 (345 to 765) 244.3 (136.03 to 748)	Time from re-orientation to successful invasion[a]	[57] [46]
$1/\mu$	min	7.21 (1.19 to 21.64)[b] 15.90 (7.21 to 28.85)[b]	Merozoite lifespan	[19] [9]

[a] The end-point of successful invasion was defined to be when the pRBC had returned to its regular shape after echinocytosis, ; see section 2.1.3 for the full discussion on this parameter.

[b] These ranges were obtained by varying the temperature of the experimental culture.

which have some type of polymorphism (whether naturally acquired or experimentally induced) may be compared to determine the impact of that specific polymorphism on invasion rate or invasion efficiency [45, 47, 48]. Other studies have compared invasion rates between treated and untreated RBCs to find out which anti-malarial compound is the most effective at limiting invasion and thus parasite propagation in general [47].

Regardless of the experimental design or objectives of the aforementioned studies, none of them developed any type of mathematical model with which to utilise the experimentally determined parameters. Because of this all of the parameters are strictly functions of time; in other words, concentrations of the relevant species were not explicitly taken into account. Therefore, no estimates for the second-order rate parameter k_1 could be found. However, estimates for the analogous first-order rate constant, k_1^* were available.[4] While this new constant will not be explicitly used in the deterministic model it is included so that it may be compared to the results obtained from the random walk model since one is able to determine the time it took the merozoites in the simulations to first contact a RBC. This serves as a form of validation since neither this parameter nor k_1 are explicitly present in the random walk model.

All of the data presented in this chapter was used to fit the base deterministic mathematical model

[4] This new parameter k_1^* has units of sec^{-1} while k_1 has units of $C^{-1}\text{sec}^{-1}$.

constructed in chapter 2 to obtain estimates for all four parameters. The chapter which follows gives the results of the model fitting procedure as well as the results from the stochastic, random walk model.

5. Results

This chapter is comprised of three sections: an analysis pertaining to the identifiability of the parameters in the base mechanistic model, estimates of the fitted parameters and their corresponding correlation coefficients for both the base and multiple invasion mechanistic models, and a series of plots showing the available data, fitted model results as well as results from the random walk model. All model fitting procedures, simulations and plots were done using the *R* statistical software program (version 3.1.3), Revolution R Enterprise for Windows (version 7.4) and related third-party packages [59].[1]

5.1 Parameter Identifiability Analysis

This section will briefly go through the theory of parameter identifiability of mathematical models as well as give an analysis of the parameters appearing in the base model. Firstly, recall that parameter identifiability has two components: structural and practical. Structural parameter identifiability is a purely theoretical exercise since it assumes perfect experimental data in the sense of having enough data of the observable variable(s). Practical parameter identifiability relies on the available experimental data and can determine whether the unknown parameters can be accurately estimated. The central idea behind structural parameter identifiability is to determine which, if any, of the parameters in the model can be uniquely identified. This is to say, can one or more of the parameters be shown to be unique in the parameter space $\mathbf{P}$ of the model structure $\Omega(\mathbf{p})$ presented in equation (5.1)?

$$\Omega(\mathbf{p}) := \begin{cases} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\ \mathbf{p},\ t) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x},\ \mathbf{p},\ t),\ \mathbf{x}(t_0) = \mathbf{x}_0(\mathbf{p}) \end{cases} \tag{5.1}$$

Specifically, if a parameter p_j is uniquely, structurally identifiable for almost any $\mathbf{p}^* \in \mathbf{P}$, then

$$\Omega(\mathbf{p}) = \Omega(\mathbf{p}^*) \tag{5.2a}$$

$$\implies p_j = p_j^*. \tag{5.2b}$$

If this can be shown for every $j \in [1, S]$ then the entire model structure is said to be uniquely identifiable. There are several methods available for assessing the identifiability of the parameters in a mathematical model particularly those which are non-linear in the state variables. The goal of any chosen method is to obtain a series of coefficients, usually non-linear with respect to the unknown parameters and species in

[1] *FME*, *RevoScaleR* and *ggplot2* were the primary third-party packages.

the model. The task is then to form the so-called Jacobian matrix which has as many rows as there are non-zero coefficients and as many columns as there are unknown parameters. This matrix, denoted by J, is comprised of the partial derivatives of the coefficients with respect to the unknown parameters evaluated at the initial conditions.

$$J := \begin{pmatrix} \frac{\partial g_1}{\partial p_1} & \frac{\partial g_1}{\partial p_2} & \cdots & \frac{\partial g_1}{\partial p_K} \\ \frac{\partial g_2}{\partial p_1} & \frac{\partial g_2}{\partial p_2} & \cdots & \frac{\partial g_2}{\partial p_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_S}{\partial p_1} & \frac{\partial g_S}{\partial p_2} & \cdots & \frac{\partial g_S}{\partial p_K} \end{pmatrix} \tag{5.3}$$

where $\mathbf{g} = [g_1, g_2, \ldots, g_S]$ and $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ are the vectors of coefficients and free parameters, respectively. The conditions for local identifiability come directly from elementary linear algebra. If $\text{rank}(J) = k$ then J is invertible and all the unknown parameters are at least locally identifiable. Equivalently, if the determinant of the Jacobian is non-zero $(\det(J) \neq 0)$ one has the same result. Clearly, for these conditions to hold there must be at least K coefficients since there are K parameters be identified.

In this work a generating series approach was used. Briefly, a generating function is used to produce a series of coefficients which in this case are the observable outputs $(\mathbf{h}(\mathbf{x}, \mathbf{p}, t))$ and their successive Lie derivatives along the vector field $\mathbf{f}(\mathbf{x}, \mathbf{p}, t)$. The Lie derivative of $\mathbf{h}$ along the vector field $\mathbf{f}$ is defined as

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{f}} \mathbf{h}(\mathbf{x}, \mathbf{p}, t) = \sum_{i=1}^{S} \mathbf{f}_i(\mathbf{x}, \mathbf{p}, t) \frac{\partial \mathbf{h}(\mathbf{x}, \mathbf{p}, t)}{\partial x_i}$$

where $\mathbf{f}_i$ is the i-th component of $\mathbf{f}$. Dropping the independent variables one obtains

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{f}} \mathbf{h}_j = \sum_{i=1}^{S} \mathbf{f}_i \frac{\partial \mathbf{h}_j}{\partial x_i} \tag{5.4}$$

where $\mathbf{h}_j$ is the j-th component of $\mathbf{h}$, the vector of observable species. The second Lie derivative of $\mathbf{h}$ along $\mathbf{f}$ would simply be

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{f}} \boldsymbol{\mathcal{L}}_{\boldsymbol{f}} \mathbf{h}_j = \sum_{i=1}^{S} \mathbf{f}_i \frac{\partial \left(\boldsymbol{\mathcal{L}}_{\boldsymbol{f}} \mathbf{h}_j \right)}{\partial x_i} \tag{5.5}$$

and so on. The required coefficients are the expressions in equations (5.4) and (5.5) evaluated at the initial conditions $\mathbf{x}_0(\mathbf{p})$ and are obtained from the vector of observable species in the model, $\mathbf{h}(\mathbf{x}, \mathbf{p}, t)$. Which observable species should be included may depend on any conservation relationships present in the original mechanistic model. In the base and multiple invasion models the sum of uRBCs, MECs and pRBCs remained constant over time which means at most two of these three species should be used

when obtaining the coefficients since one of them can always be expressed in terms of the other two. Fortunately this was not an issue as the observable species which were chosen were the merozoite and uRBC populations – this meant that $\mathbf{h} = [m\ r]^T$. The general initial conditions for the state variables were the following: $m(0) = m_0$, $r(0) = r_0$, $c(0) = x(0) = 0$.[2] With these conditions the required coefficients are as follows

$$\boldsymbol{\mathcal{L}_f}\mathbf{h}_1|_0 = -m_0\left(\mu + k_1 r_0\right) \tag{5.6a}$$

$$\boldsymbol{\mathcal{L}_f}\mathbf{h}_2|_0 = -k_1 m_0 r_0 \tag{5.6b}$$

$$\boldsymbol{\mathcal{L}_f}\boldsymbol{\mathcal{L}_f}\mathbf{h}_1|_0 = \mu m_0\left(\mu + k_1 r_0 + k_1 m_0\right) + 2k_1^2 m_0^2 r_0 + k_{-1} k_1 m_0 r_0 \tag{5.6c}$$

$$\boldsymbol{\mathcal{L}_f}\boldsymbol{\mathcal{L}_f}\mathbf{h}_2|_0 = k_1 r_0 m_0\left(\mu + k_1 r_0 + k_{-1}\right) + k_1^2 m_0^2 - k_{-1} k_2 \tag{5.6d}$$

Equations (5.6a) to (5.6d) should be recognised as g_1 to g_4 from the Jacobian matrix in equation (5.3). Deriving these with the respect to each free parameter in the base mechanistic model gives one the Jacobian specific to this problem, J_s. The columns correspond to k_{-1}, k_1, k_2 and μ, respectively.

$$J_s = \begin{pmatrix} 0 & -m_0 r_0 & 0 & -m_0 \\ 0 & -m_0 r_0 & 0 & 0 \\ m_0 r_0 k_1 & m_0\left((m_0 + r_0)\mu + r_0 k_{-1} + 4 m_0 r_0 k_1\right) & 0 & 2m_0\mu + m_0(m_0 + r_0)k_1 \\ m_0 r_0 k_1 - k_2 & m_0 r_0 \mu + m_0 r_0 k_{-1} + 2m_0\left(m_0 + r_0^2\right)k_1 & -k_{-1} & m_0 r_0 k_1 \end{pmatrix}$$

It can easily be verified by the cofactor method that the determinant of J_s is non-zero:

$$\det(J_s) = -m_0^3 r_0^2 k_{-1} k_1 < 0. \tag{5.7}$$

Equation (5.7) tells one that all four parameters are at least locally identifiable. Global identifiability requires that $\Omega(\mathbf{p}) = \Omega(\mathbf{p}^*) \implies p_j = p_j^*$ for every $j \in [1, 4]$ since there are four parameters to estimate. To show this, consider first equation (5.6b) with the aforementioned condition:

$$-k_1 m_0 r_0 = -k_1^* m_0 r_0 \tag{5.8a}$$

$$\implies k_1 = k_1^*. \tag{5.8b}$$

Using this result with equation (5.6a) one has the following,

$$-m_0\left(\mu + k_1 r_0\right) = -m_0\left(\mu^* + k_1 r_0\right) \tag{5.9a}$$

$$\mu + k_1 r_0 = \mu^* + k_1 r_0 \tag{5.9b}$$

$$\implies \mu = \mu^*. \tag{5.9c}$$

[2] Of course, $\{m_0, r_0\} > 0$.

One can continue this process with equations (5.6c) and (5.6d) to confirm that the global identifiability condition is met.

The implication of this is simple: given sufficient, continuous, noise-free data one would be able to obtain accurate estimates for each of the free parameters and be confident of those estimates. This is the point where practical parameter identifiability and correlation coefficients come in since they deal explicitly with the aforementioned issue of sufficient, continuous, noise-free data. Correlation coefficients provide a measure of how correlated two parameters are by revealing linear or pseudo-linear relationships between them. Let ξ_{ij} be a correlation coefficient where i and j are the indices of two distinct parameters. The maximum absolute value of ξ_{ij} is one which means that $|\xi_{ij}| \in [0, 1]$. The higher the absolute value of ξ_{ij} the more linear the relationship between the two corresponding parameters is said to be. Quite simply, if the quantity or quality of the data are inadequate one or more of the parameters will be strongly correlated and the confidence intervals of the estimates of the parameters will therefore be too large for one to trust those estimates. The section which follows explores this issue in detail.

5.2 Parameter Estimates

The base model was fitted to the data from sections 4.1 and 4.2 and the best-fit parameter estimates were then used in the random walk model. Specifically, k_{-1}, k_1, k_2 and μ were estimated with the base model and then only the estimates of k_{-1}, k_2 and μ were used as input for the random walk model. The parameter k_1 is not explicitly used in the random walk model but it's first-order analogue k_1^* could be obtained from the random walk simulations and was therefore subject to validation from the available data. The criterion for choosing the best set of parameter estimates was based on which set minimised the sum of squares between the observed data and the predicted values generated by the corresponding parameter set.

The base model was first fitted to the data produced by Baum and colleagues where experiments were performed exclusively under static conditions with a 1 % hematocrit. Clough et al. did not specify the exact hematocrit used in the experimental procedures but, a range of 2 % to 5 % was quoted. Therefore, the base model was fitted to two separate scenarios, one for each extreme of the given range. Care must be taken not to directly compare the estimates for Baum et al. with those obtained for the static cultures from Clough et al. given that the hematocrit used in each was different. Several dozen parameter-fitting procedures were run with a set of different initial estimates in order to establish a local minimum that

Table 5.1: Parameter correlations for the base mechanistic model.

Baum et al.				**Clough et al.**[b]							
Static (1% Hct)				**Static (2%/5% Hct)**				**Suspension (2%/5% Hct)**			
$\mathbf{k_1}$				$\mathbf{k_1}$	0.41	0.08	-1	$\mathbf{k_1}$	1	-0.50	-1
0.19	$\mathbf{k_2}$			0.26	$\mathbf{k_2}$	-0.24	-0.42	1	$\mathbf{k_2}$	-0.51	-1
−0.24	0.35	$\boldsymbol{\mu}$		−0.37	−0.08	$\boldsymbol{\mu}$	-0.08	−0.44	−0.44	$\boldsymbol{\mu}$	0.50
−0.14	0.71	0.65	$\mathbf{k_{-1}}$	−1	−0.27	0.37	$\mathbf{k_{-1}}$	−1	−1	0.44	$\mathbf{k_{-1}}$

[a] Hct = hematocrit.

[b] The entries in red correspond to the values obtained from the parameters fitted when a 5 % hematocrit was used while the entries in black used a 2 % hematocrit.

best approximates the global minimum for the model-fitting exercises. Unfortunately an absolute global minimum could not be guaranteed but, the parameter estimates in table 5.2 were accepted when the sum of squared residuals (SSR) remained constant for three consecutive simulations for each set of initial estimates.

Before looking at the parameter estimates for the base and multiple invasion mechanistic models one must first examine the issue of parameter identifiability more closely to determine whether those estimates are accurate and can be trusted to a reasonable degree of accuracy.

5.2.1 Base Mathematical Model

Table 5.1 shows the correlation coefficients for each of the four parameters obtained for each fitting procedure. Data for both static and suspension cultures where available from the Clough et al. study paper whereas only static cultures were evaluated in the Baum et al. study. Furthermore, the entries in red in table 5.1 correspond to the values obtained from the parameters fitted when a 5 % hematocrit was used while the entries in black correspond to a 2 % hematocrit. Baum et al. only used a 1 % hematocrit for their experimental cultures. For example, the values 0.26 and 0.41 (the latter in red) are the correlation coefficients for k_1 and k_2 obtained from the static culture data of Clough et al. Recall from section 1.3 that two parameters are highly correlated if the absolute value of their correlation

Table 5.2: Base model: estimated parameters

Parameter	Units[a]	Baum et al.[b]	Clough et al.			
			Hematocrit: 2 %		Hematocrit: 5 %	
		Static	Static	Suspension	Static	Suspension
k_{-1}		3.84	3.02	3.76	5.12	2.00
μ	min^{-1}	1.33×10^{-1}	1.39×10^{-1}	1.38×10^{-1}	1.37×10^{-1}	1.32×10^{-1}
k_2		8.04×10^{-2}	8.11×10^{-2}	8.04×10^{-2}	7.85×10^{-2}	8.96×10^{-2}
k_1	$\text{C}^{-1}\text{min}^{-1}$	3.98×10^{-6}	1.10×10^{-5}	5.38×10^{-6}	7.60×10^{-6}	9.66×10^{-7}

[a] "C" represents units of concentration (cells/µl) making this parameter a second-order rate constant.

[b] A 1 % hematocrit was used for the experimental cultures in this study.

coefficient is close to one. Conversely, the nearer to zero it is the less correlated the two parameters are said to be.

A cursory examination of table 5.1 quickly reveals an issue with k_{-1} and k_1: these parameters are completely, negatively correlated for the data from Clough et al. which means that one could replace either parameter with a negative multiple of the other. However, the same coefficient for the Baum et al. data is relatively small (-0.14). A second observation to take note of is the fact that the suspension cultures show significantly higher correlations than their static counterparts; this may be partly due to fewer data points being available. More specifically, k_{-1}, k_1 and k_2 are all perfectly correlated which means that any two of these parameters may be written as multiples of the third.[3] A consequence of this is that the parameter estimates obtained from the suspension culture data cannot be relied upon to accurately quantify the biological processes they represent. This ultimately forces one to consider only the estimates from the static culture experiments.

The least correlated parameter in general is arguably the death rate μ when looking at the coefficients in tab: ParamCorrelations. One notices in table 5.2 which shows the best-fit parameters for the base model that the estimate for μ is essentially the same which is a promising result since it seems reasonable to assume that this rate should be independent of the hematocrit of the solution and the type of RBCs which were used. The average value for this parameter, calculated only from the estimates obtained from static cultures, is $0.136\,\text{min}^{-1}$ which is 7 minutes and 21 seconds.[4] Focusing on k_2 – the rate of

[3] What this means is that one then effectively has only two parameters for the base model, one of k_{-1}, k_1 or k_2, and μ.

[4] A range of 5 to 11 minutes was used for the estimation procedures. The full range given in table 4.1 was not used

successful invasion – one sees again that there are not large differences between the estimates and the average value translates into a merozoite taking roughly 12 and a half minutes to fully successfully invade a susceptible RBC. Interestingly this was very close to the upper limit of the estimations obtained from the literature for this specific parameter. Recall that successful invasion was defined from the moment the apical end of the merozoite irreversibly binds with the host RBC (tight junction formation) to the end of echinocytosis.

The only parameter for which no estimates were available was k_{-1}, the rate of detachment. As a result an arbitrary range of $0.01\,\mathrm{s}^{-1}$ to $60\,\mathrm{s}^{-1}$ was used. Taking the reciprocal of the values in table 5.1 from static cultures gives one an average value of 15 seconds. This means that 15 seconds is the lifespan of the average MEC. Lastly, the effective contact rate between merozoites and uRBCs (k_1) is the only second-order rate constant which makes obtaining a simple, direct interpretation difficult; however, the literal interpretation is the number of contacts an average RBC or merozoite concentration unit experiences per minute. One can see from table 5.2 that the three relevant estimates for this parameter span an order of magnitude which is to be expected since it does rely on the absolute numbers of merozoites and RBCs and these differ considerably between the two studies. Because of this it makes little sense to calculate an average value.

5.2.2 Multiple Invasion Mathematical Model

The multiple invasion model was only fitted to the data from Clough et al. since this was the only study which had sufficient data to justify this particular model. Recall that the multiple invasion model defined separate variables for each invasion class, i.e $x_r^{(1)}$ denoted singly-invaded pRBCs, $x_r^{(2)}$ doubly-invaded and so forth. The data on multiply-invaded pRBCs were given as relative percentages of the re-invasion parasitemia and were averaged across four experiments each for static and suspension cultures as discussed in section 4.2 and figure 4.3. The percentages for singly-, doubly- and triply-invaded pRBCs were as follows: 78.80 %, 17.00 % and 4.25 % for static cultures and 93.60 %, 5.55 % and 0.50 % for suspension cultures. As an example, isolate 15 from the static culture experiments had a re-invasion parasitemia of 6.80 % which means the contributions of singly-, doubly- and triply-invaded pRBCs to the total re-invasion parasitemia was 5.36 %, 1.16 % and 0.29 %, respectively.[5] To implement this for fitting of the multiple invasion model, the re-invasion parasitemia of each isolate (data point) was split into

because those ranges correspond to temperatures other than 37 °C.

[5] $6.80 \times 0.79 = 5.36$, $6.80 \times 0.17 = 1.16$ and $6.80 \times 0.04 = 0.49$.

Table 5.3: Parameter correlations for the multiple invasion mechanistic model.

Static (2%/5% Hct)[b]				**Suspension (2%/5% Hct)**[b]			
$\mathbf{k_1}$	0.41	0.08	-1	$\mathbf{k_1}$	0.19	-0.53	-1
0.26	$\mathbf{k_2}$	-0.24	-0.42	1	$\mathbf{k_2}$	-0.06	-0.14
−0.37	−0.08	$\boldsymbol{\mu}$	-0.08	−0.44	−0.44	$\boldsymbol{\mu}$	0.53
−1	−0.27	0.37	$\mathbf{k_{-1}}$	−1	−1	0.44	$\mathbf{k_{-1}}$

[a] The entries in red correspond to the values obtained from the parameters fitted when a 5 % hematocrit (Hct) was used while the entries in black used a 2 % hematocrit.

three separate parasitemias, one for each invasion class as just described.

The correlation coefficients for the parameters in the multiple invasion model are shown in table 5.3. Bear in mind these are the same four parameters which appear in the base model. Again, the entries in red correspond to the values obtained when fitted with a 5 % hematocrit while the entries in black correspond to a 2 % hematocrit. The best-fit parameters for the multiple invasion model are shown in table 5.4 and one should immediately notice that the estimates are generally of the same order of magnitude with the exception of k_{-1} and k_1 for both static and suspension cultures using a 5 % hematocrit. However, table 5.3 shows that these two parameters are completely correlated which means one cannot interpret each parameter separately and rely on their respective estimates.[6] At best only their ratio can be obtained and compared with the same ratio given by the estimates of the base model. Consider first only the static cultures. Since k_{-1} and k_1 are completely negatively correlated in both the base and multiple invasion models one may write $k_{-1} = -\beta k_1$ where $\beta > 0$ and has units of concentration. The values of β (one each for the base and multiple invasion models) would then be of the same order of magnitude for their respective static cultures. The same can be done for the suspension cultures.

5.3 Mathematical Model Outcomes

The fitted model for the data from both studies are shown in figures 5.1 and 5.2 which may be divided into their respective top and bottom panels. The panels in the top row are scatterplots showing the

[6] $\xi_{12} = -1$ where the indices 1 and 2 refer to k_{-1} and k_1, respectively.

Table 5.4: Multiple invasion model: estimated parameters for Clough et al.

Parameter	Units[a]	Hematocrit: 2 %		Hematocrit: 5 %	
		Static	Suspension	Static	Suspension
k_{-1}		1.89	1.44	2.70×10^{1}	1.66×10^{1}
μ	min^{-1}	1.39×10^{-1}	1.33×10^{-1}	1.38×10^{-1}	1.37×10^{-1}
k_2		7.85×10^{-2}	8.05×10^{-2}	7.85×10^{-2}	7.86×10^{-2}
k_1	$\text{C}^{-1}\text{min}^{-1}$	7.37×10^{-5}	2.05×10^{-6}	4.20×10^{-5}	9.57×10^{-6}

[a] "C" represents units of concentration (cells/µl) making this parameter a second-order rate constant.

experimental data (filled circles), fitted meachanistic model results (open circles) as well as the results from the random walk model (smaller, coloured circles). The independent variable in the scatterplots is the initial parasitemia while the dependent variable is the re-invasion (final) parasitemia. The colour gradients for the random walk model results denote the multiplication rate of each point keeping in mind that each point is a separate simulation for a particular integer-valued initial parasitemia. The gradients range from light green for low multiplication rates to bright red for high multiplication rates. The bottom panels display histograms of only the results from the random walk model for each initial parasitemia subset (integer value). Both the independent and dependent variables are the same as in the scatterplots. The histograms show the distribution of the re-invasion parasitemia more clearly and were plotted such that the initial parasitemia from the top and bottom graphs line up vertically. The dashed, red lines and text denote the position and value of the means of each distribution, respectively.

Figure 5.1 shows that the base mechanistic model captures the roughly linear relationship between the initial and final parasitemias. The dashed line is a linear approximation to the data which has been forced through the origin. The reason it has been forced through the origin is to preserve the biological realism of the system since one cannot have a re-invasion parasitemia greater than zero if there were no pRBCs to begin with, i.e 0 % initial parasitemia. The results from the random walk model are consistently higher, on average, for each initial parasitemia subset. However, the stochastic nature of the random walk process did seem to accommodate the full range of the the data. Figure 5.2 shows the experimental data from Clough et al. obtained from the laboratory-adapted *P. falciparum* strains and the best fitted model results along with the random walk model results. The final fitted values were identical when

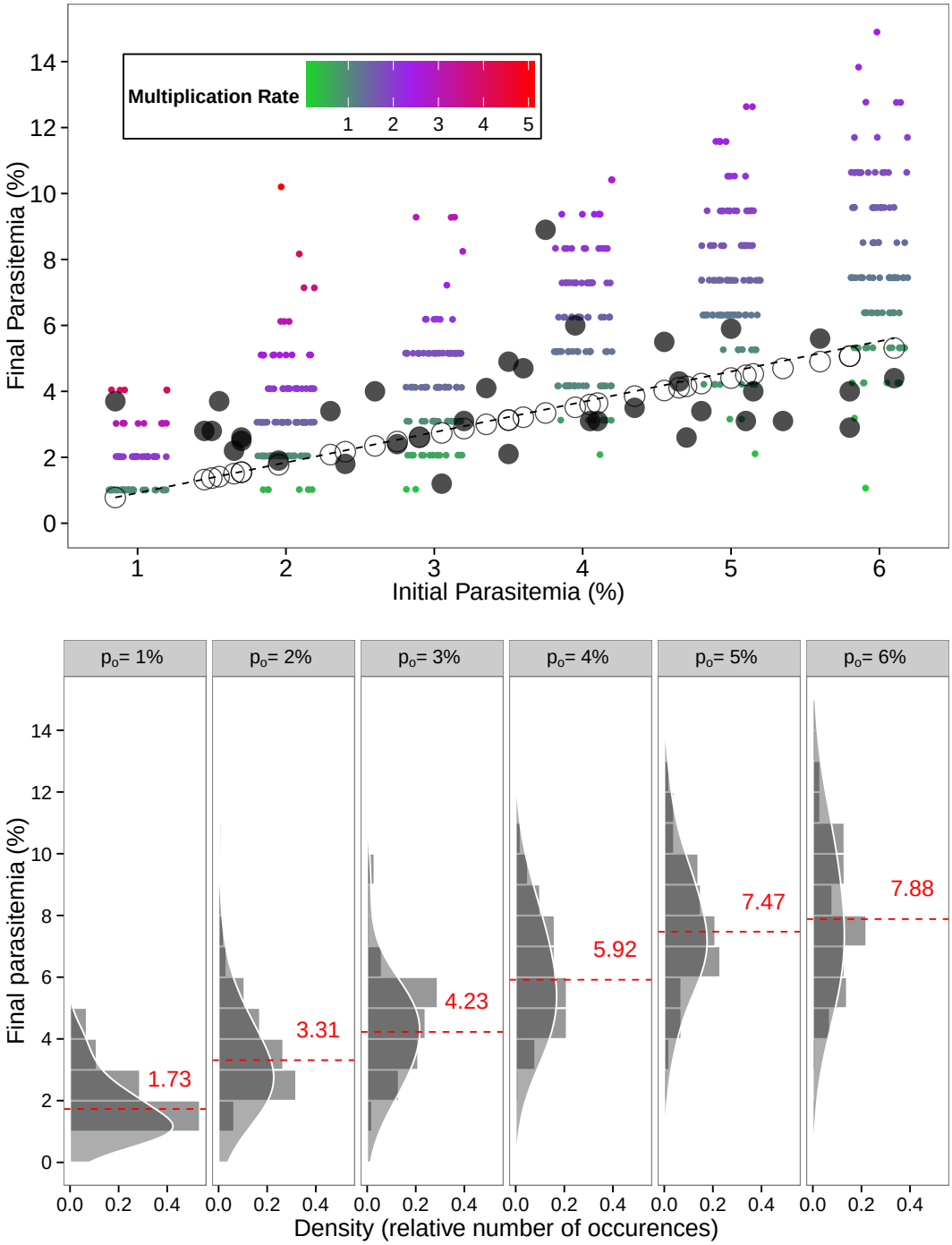

Figure 5.1: Experimental, fitted and predicted values of the re-invasion parasitemia from the base mechanistic model and random walk model for data from Baum and colleagues [17]. The data were obtained from static cultures and are represented by large, closed circles (●) where each of these represent a single patient isolate ($n = 38$). The fitted mechanistic model values are shown as large, open circles (◯). The smaller, coloured points are the results from the random walk model and the colour of each point represents the multiplication rate of a single simulation. The histograms in the bottom panel show the distribution of the re-invasion parasitemia for each integer-valued initial parasitemia (see the main text).

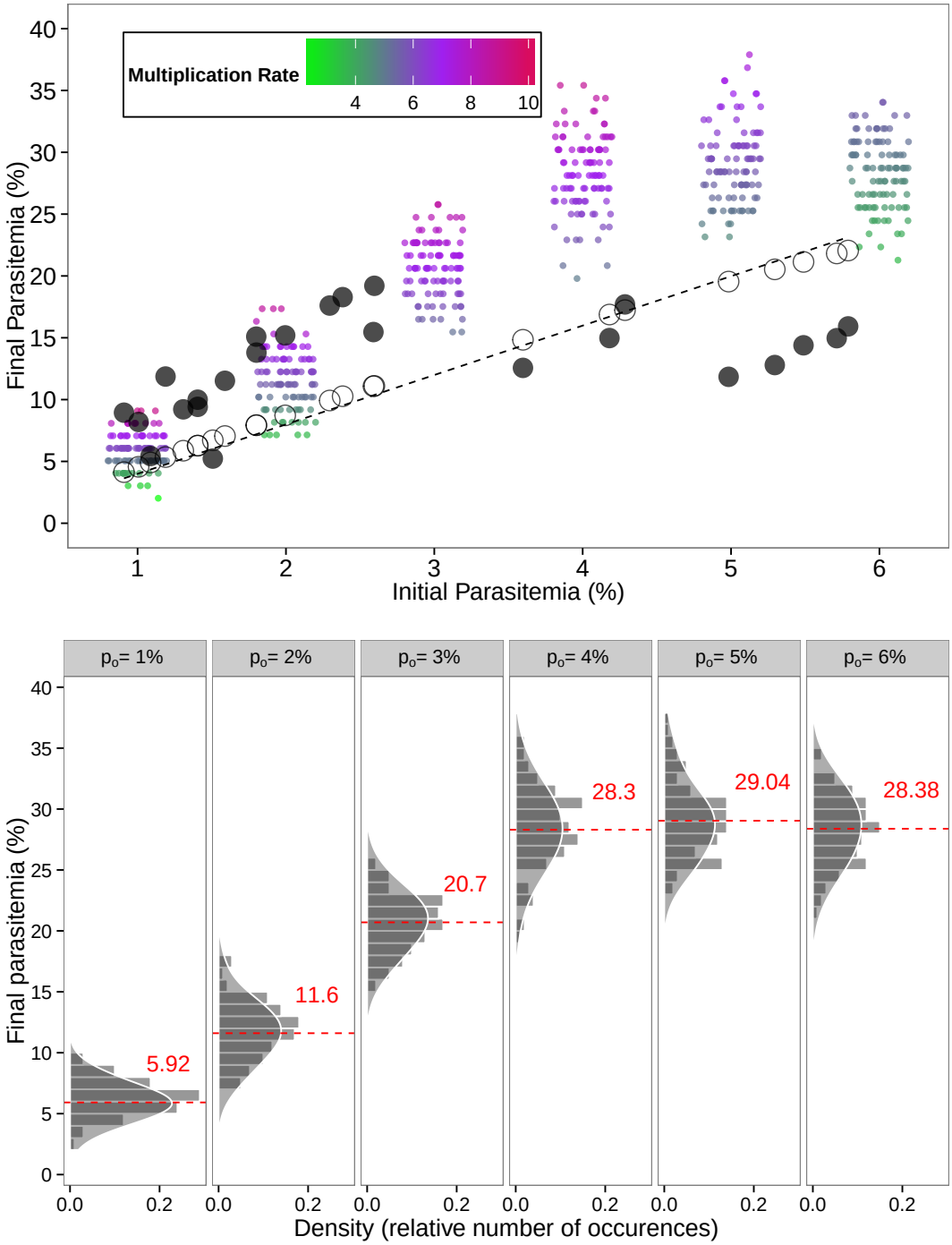

Figure 5.2: Experimental, fitted and predicted values of the re-invasion parasitemia from the base mechanistic model and random walk model for the static culture data from Clough and colleagues [16]. The data were obtained from static cultures and are represented by large, closed circles (●) while the fitted mechanistic model values are shown as large, open circles (◯). The smaller, coloured points are the results from the random walk model. Each data point represents a single patient isolate ($n = 25$). The colour of each point represents the multiplication rate of that particular simulation. The histograms in the bottom panel show the distribution of the re-invasion parasitemia for each integer-valued initial parasitemia (see the main text).

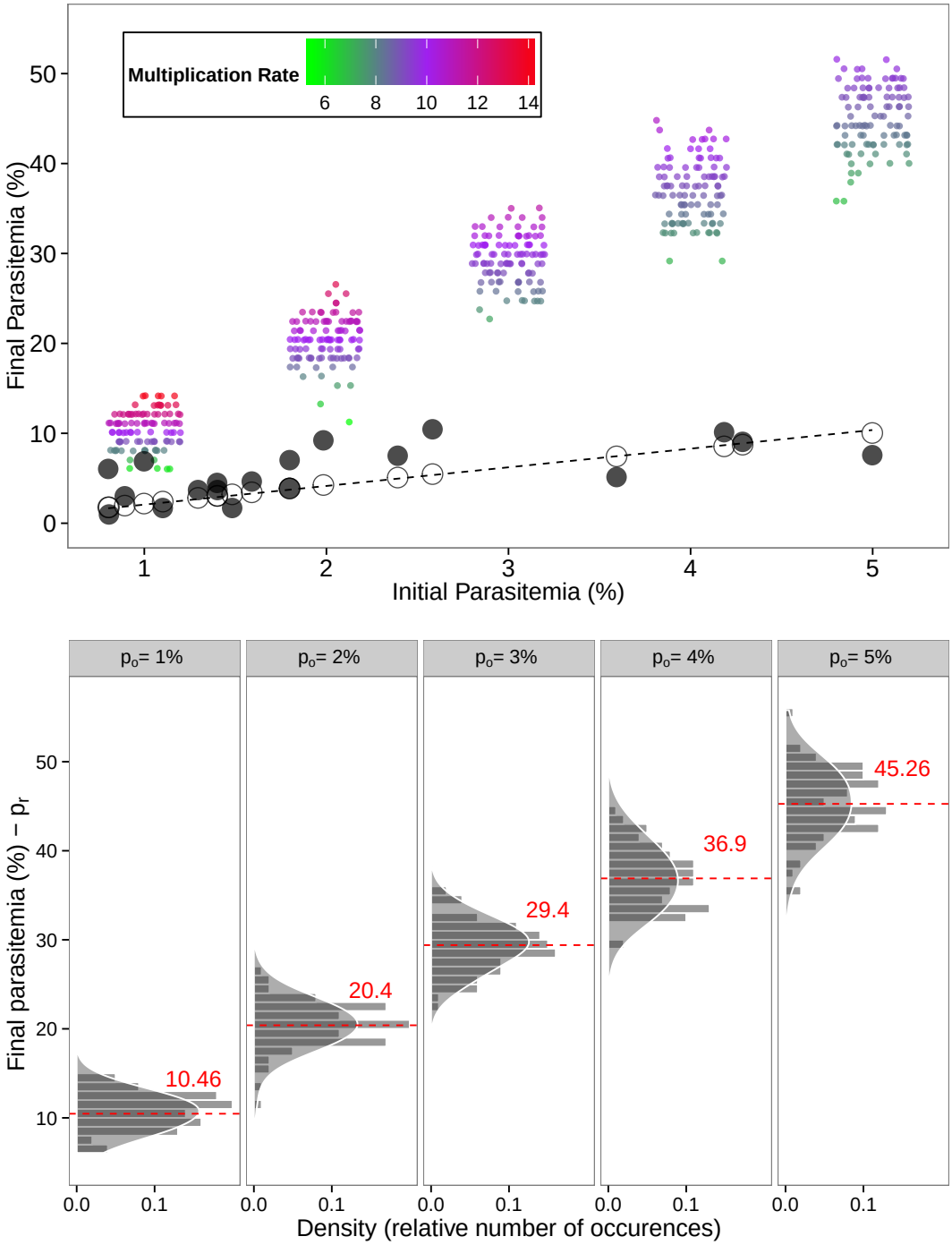

Figure 5.3: Experimental, fitted and predicted values of the re-invasion parasitemia from the base mechanistic model and random walk model for the suspension culture data from Clough and colleagues [16]. The data are represented by large, closed circles (●) while the fitted mechanistic model values are shown as large, open circles (◯). The smaller, coloured points denote the outcome of the random walk model. Each data point represents a single patient isolate ($n = 22$). The colour of each point represents the multiplication rate of that particular simulation. The histograms in the bottom panel show the distribution of the re-invasion parasitemia for each integer-valued initial parasitemia (see the main text)

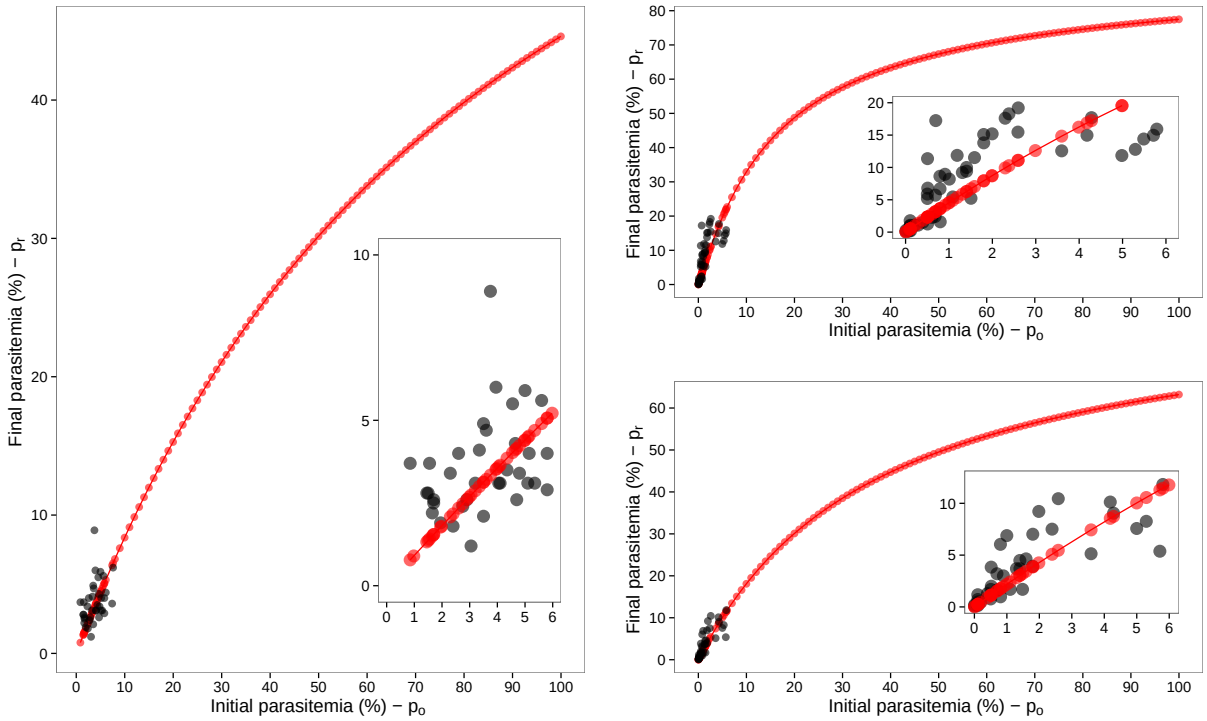

Figure 5.4: Extrapolated base mathematical model for data from both studies. The graph on the left displays the results from Baum et al. while the top right and bottom right graphs show the results from static and suspension cultures from Clough et al, respectively. The main figures show the entire extrapolation while the insets shows a magnified area where only the range of values for which experimental data were available. The data and extrapolated model values are denoted by black and red points, respectively. The red line linking the model points is included for clarity. Each data point represents a single patient isolate.

using either a 2 % or 5 % hematocrit; therefore, only a single fitted curve is shown.[7] The most important observation is the fact that the model cannot capture the saturation effect which begins between 3 % to 4 % initial parasitemia.

This saturation effect is proposed to be due metabolic limitations, spatial effects or possibly a combination of these and other, unknown factors. The model does however, capture the initial, linear relationship between the initial and re-invasion parasitemia; a well-known and expected observation. The results from the random walk model are curious since it does a better job at describing the aforementioned initial, linear relationship with the same parameters as the base mechanistic model. Most notable is the saturation effect seen at around a 4 % initial parasitemia. However, it is still well above the saturation point seen in the data.

Regarding the multiple invasion model, the visual fit was exactly the same as the base model and

[7] Of course, the parameter estimates were not identical, this can be seen in table 5.2.

therefore graphs of those fits were not included. For completion, the fitted mechanistic model as well as the random walk model results for the suspension culture data are displayed in figure 5.3. It was earlier noted that the parameter estimations from these data were unreliable, however, it is worth noting that the mechanistic model does appear to describe the data more accurately in the case of the static cultures. This is an important point since this model is better suited theoretically to deal with suspension cultures where homogeneity is assumed. A final observation is that the random walk model significantly over-estimates the re-invasion parasitemia and, by extension, the multiplication rate. This is thought to be due to the fact that merozoites were individually placed on the two-dimensional grid for the random walk simulation for suspension cultures as explained in section 3.4.

For a broader perspective, the entire theoretical range of starting parasitemias was used as input in the base mathematical model with the best-fit parameters in table 5.2 to illustrate how the model behaves for all possible initial values. In other words, the base mathematical model was extrapolated to include areas where there was no experimental data available; figure 5.4 shows this extrapolation.

Lastly, a more explicit look at the stochastic nature of the random walk model should be highlighted. Figure 5.5 shows the result of a typical realisation with 100 RBCs, five of which are mature schizonts (dark red circles) which means the initial parasitemia was 5 %. The merozoites are shown as small, green circles and the number inside specific RBCs denotes the number of merozoites which successfully invaded those particular RBCs. In this example there were a total of 80 merozoites which were tracked since each of the five schizonts produced 16 merozoites. Counting up the values inside the pRBCs one sees that 15 merozoites invaded 13 different RBCs. This means that the re-invasion parasitemia was $100\times13/(100 - 5) = 13.68$ %. The invasion and pseudo-invasion efficiencies were $100\times15/80 = 18.50$ % and $100\times13/80 = 16.25$ %, respectively.

Perhaps the most significant part of the random walk model is the time it takes the average merozoite to reach an uRBc for the first time. This is of course determined by the length of each step the merozoite takes as well as the degree to which the steps are correlated. The parameters governing these two variables were discussed in sections 3.1.1 and 3.1.2 but the practical results can be seen in figure 5.5 and described by a particular probability distribution. A continuous random variable X which conforms to a Γ-distribution with shape parameter α and rate parameter β is denoted by $X \sim \Gamma(\alpha, \beta)$.[8]

[8] The probability density function is $\Gamma(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ where $\Gamma(\alpha)$ is the Γ-function evaluated at α.

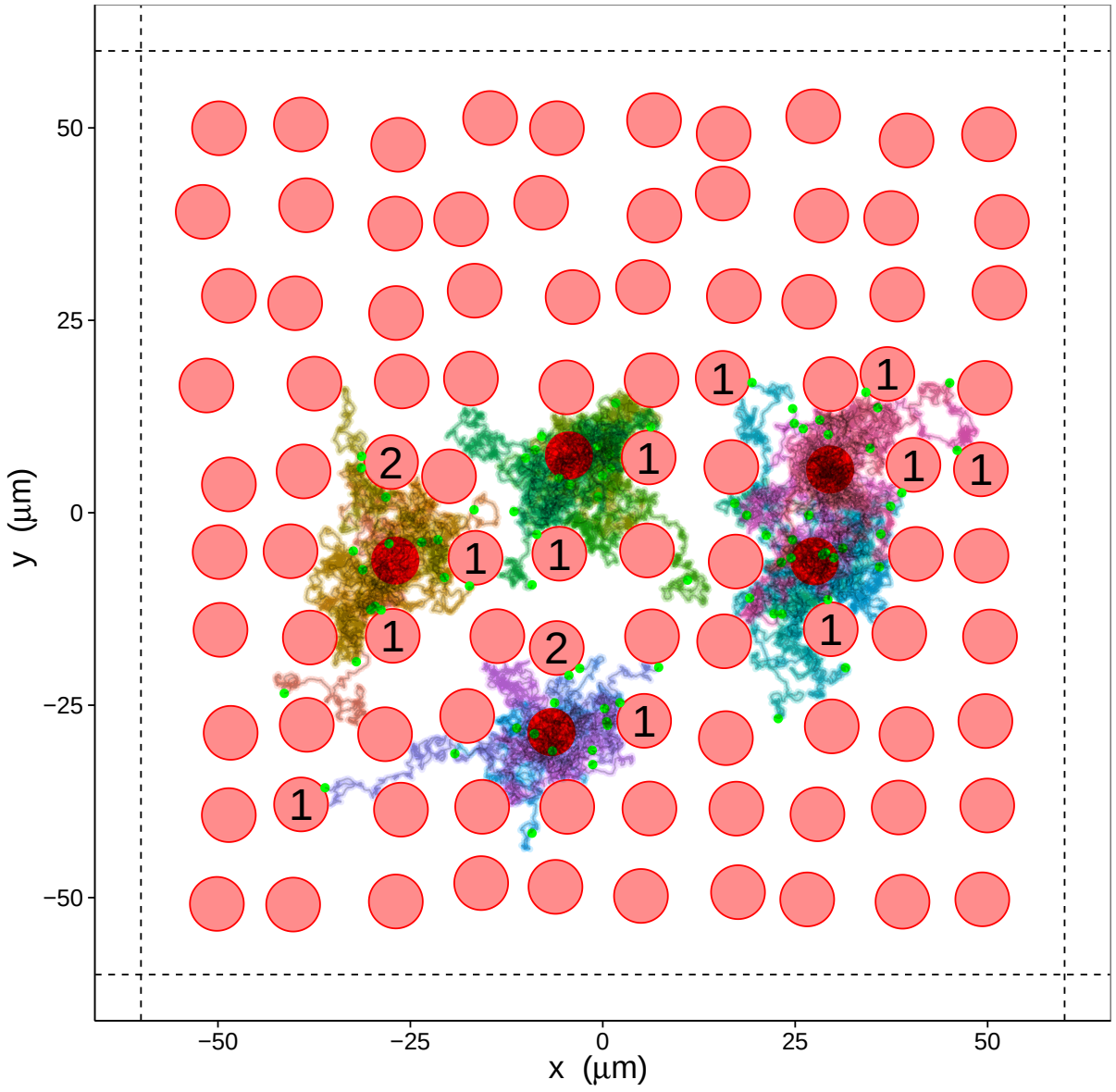

Figure 5.5: Example of the random walk model for static culture conditions. The initial and re-invasion parasitemias for this simulation were 5 % and 13.68 %, respectively. This indicates a multiplication rate of 2.74 for a single cycle. The pale, red circles denote the RBCs present in the simulation while the darker red circles indicate the five mature schizonts who each released 16 merozoites (smaller, green circles). The paths of individual merozoites are shown as a colour gradient, each colour belonging to a different merozoite. The values inside specific pRBCs denote the number of merozoites which successfully invaded those particular RBCs – these merozoites can be seen as being attached to the pRBCs. The selectivity index was 2.14 and the invasion and pseudo-invasion efficiencies were 18.50 % and 16.25 %, respectively. Both axes are given in µm and the black, dashed lines indicate the boundaries of the simulation ($|x| = |y| = 60$).

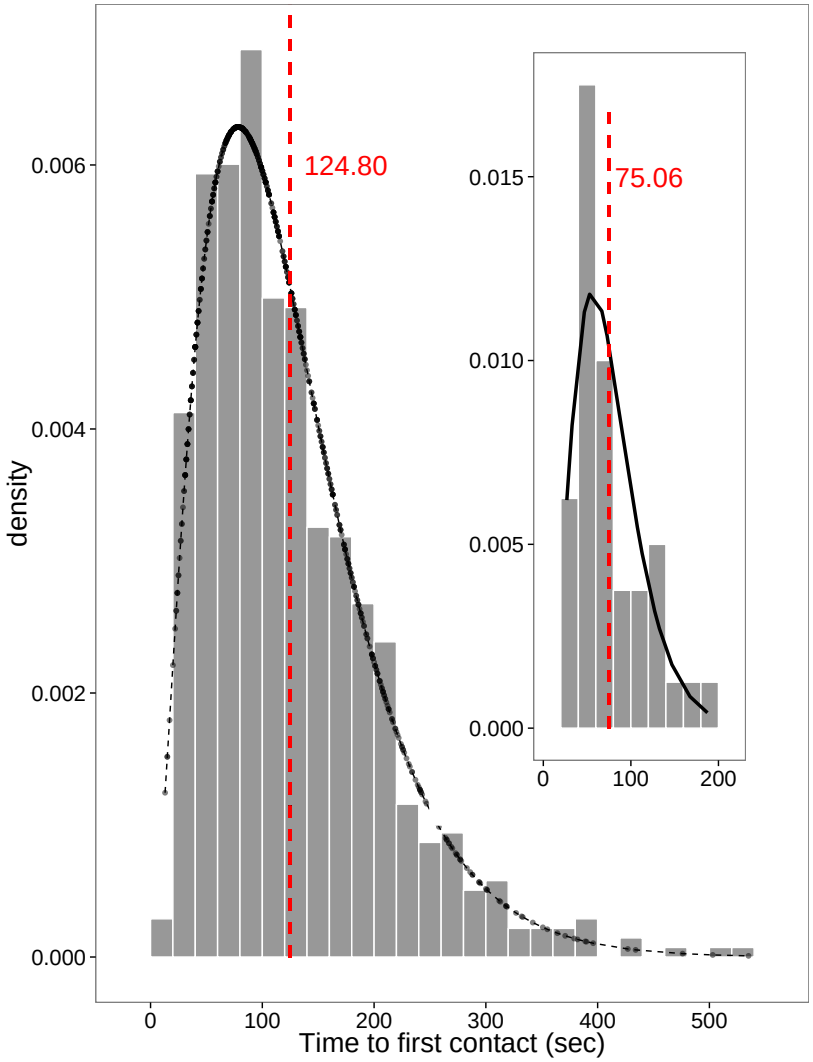

Figure 5.6: Stochastic model results of first contact times under static conditions and the available data. The main figure shows the distribution of times a merozoite took to first contact a RBC. The inset shows the data from the only study which provided a distribution of first contact times, the axes are the same as in the main figure [21]. A total of 800 merozoites were simulated by the model but only 691 could be included in the distribution since the rest decayed before reaching a RBC. Each bar has a width of 20 seconds and each point represents a single merozoite. Therefore all points falling into the 1 to 20 second range fall into the first bin, those in the 21 to 40 second range in the second bin and so forth. The curves which overlay the model and data are fitted Γ-distributions. The red, dashed lines indicate the expected time of first contact (75.06 and 124.80 seconds for the data and model, respectively).

The Γ-distribution estimates the time between Poisson distributed events. In other words, $\Gamma(\alpha, \beta)$ gives the probability distribution of how long it takes for α events to occur given that the events occur randomly and independently with the mean time between events given by β. Figure 5.6 shows the distribution of first contact times of merozoites and RBCs; these times represent k_1^* in table 4.1. In other words, the histogram is a distribution of the time it took a merozoite to first come into contact with a RBC under static conditions using the value of κ with which all other plots in this chapter was produced.[9] Very importantly, these data were generated from a different simulation that produced figure 5.5 since many more merozoites and RBCs had to be simulated to obtain a distribution which could be reliably estimated.[10] Apart from the number of merozoites and RBCs all other parameters were unchanged.

[9] This value was 1.09 given in table 3.1.

[10] An hematocrit of 1 % was used and the initial parasitemia was set to 5 %.

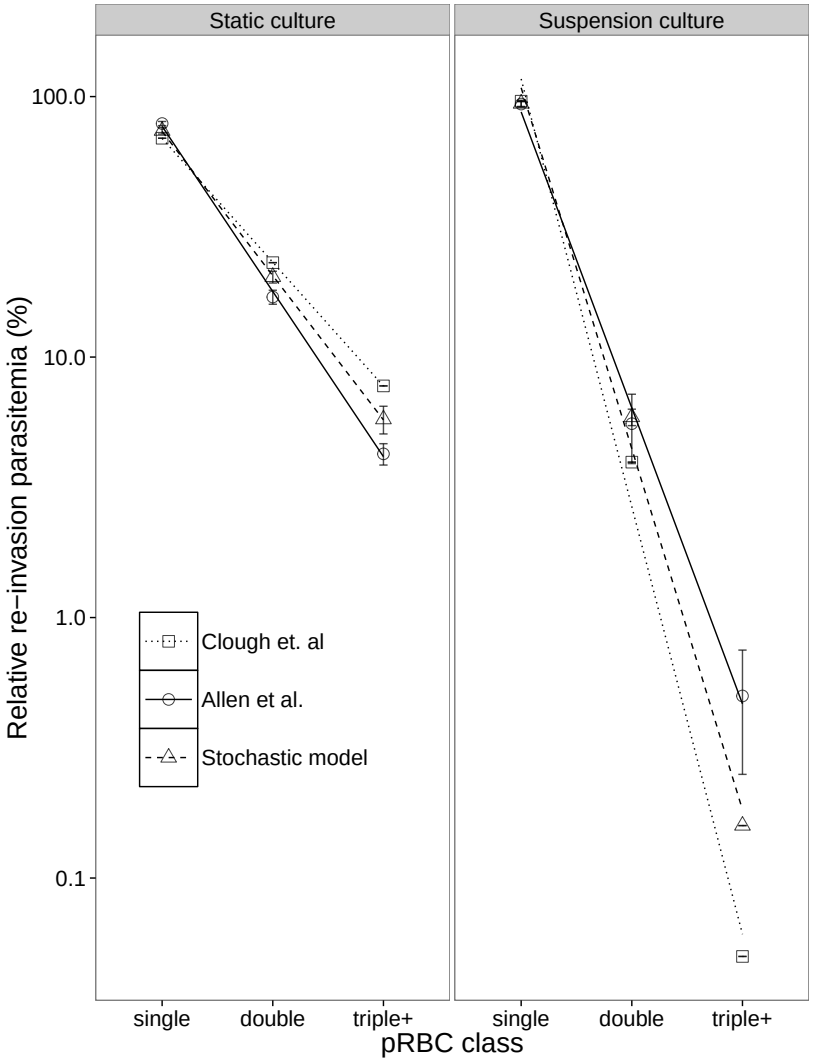

Figure 5.7: Prediction of the proportion of multiply-invaded pRBCs from the stochastic model. This figure is similar to figure 4.3 with the addition of data from another study ([14]) and the predictions from the stochastic model. The left and right panels are for static and suspension cultures, respectively. The percentages for single, double and triple+ pRBCs from the stochastic model were 73.93 %, 20.30 % and 5.77 % for static cultures and 93.96 %, 5.88 % and 0.16 % for suspension cultures. The third invasion class (triple+) denotes the pooled number of pRBCs which were invaded by three or more merozoites. Linear regression lines are given for the data from both studies as well as the stochastic model. Error bars denote the standard error (SE) associated with four experiments each for static and suspension cultures and four simulations of the stochastic model with a 2 % initial parasitemia.

The shape (α) and rate (β) parameters were estimated to be 2.68 and 0.021 while the associated standard errors were 0.136 and 0.001 for the stochastic model, respectively. These same parameters for the data were 3.979 and 0.053 and the standard errors were 0.854 and 0.012. The fact the the standard errors are larger for the data than the models estimates indicate that there were fewer data points than simulated merozoites. The expected values are given by α/β which means that the expected time a merozoite takes to first contact a RBC is about 75 and 125 seconds for the data and stochastic model, respectively. It is important to notice that the results in figure 5.6 are independent of k_{-1} and k_2, the rates of detachment and successful invasion, respectively. This is advantageous since it lends respectability to the predictions of the random walk model.

A final illustration of the predictions the stochastic model can make pertains to the number of multiply-

invaded pRBCs. Figure 5.7 shows these predictions for static and suspension cultures along with data from two different articles in the literature review. The first is from Clough et al. where an hematocrit of 2 % to 5 % was used and the other from Allen et al. where the hematocrit was ≈5 % [16, 14]. It should also be noted that the y-axis in figure 5.7 as relative and not absolute re-invasion parasitemia. This was done because neither study presented their multiply-invaded pRBC data in absolute terms but only relative to the total number of pRBCs. The predictions from the stochastic model seem to lie in between the two aforementioned studies which may be seen as encouraging at least in a broad sense given the range and uncertainty of the hematocrits which were used and because these are relative and not absolute values.

This concludes the results obtained from the deterministic and stochastic models constructed for the current work. A more detailed discussion of these results follow in the next chapter.

6. Discussion

All of the steps of constructing a mathematical model have been presented in the preceding four chapters, particularly chapter 2 which defined the basic model of erythrocytic invasion during the transition period of malarial infection. This chapter aims to elaborate on the results obtained in chapter 5, discuss the implications thereof, the limitations of the models presented in this work and provide possible experimental scenarios which may aid in obtaining the data needed to better estimate the free parameters and validate the aforementioned models.

First and foremost it must be noted that because the mechanistic mathematical models used in this work require concentrations of cells (cells/µl) as their input, the models are ideally suited for homogeneous environments, i.e suspension cultures where the different populations of cells are assumed to be well-mixed before and during the transition phase. That being said, the models were fitted to data obtained from static cultures mainly due to the limited amount of data which were available; more on this topic will be discussed shortly.

Before an analysis of the performance of the mathematical model is presented, a closer examination of the data used for the estimation procedures should be done in order to clarify exactly why additional and more relevant data is needed to obtain satisfactory results in terms of parameter estimation. Firstly, the data available was unfortunately not ideal for the models constructed in this work. This is primarily because the models produce time-series estimates, i.e they track how the populations of merozoites, MECs, pRBCs and uRBCs change over some length of time – a period of 20 minutes was used for all simulations. All of the data used to estimate the parameters in the base model are non-time-series data, i.e they do not track how the relevant populations change over time. Instead what is given is the end result of the first *in vitro* replication cycle for laboratory or clinical isolates. This "end result" is the number of pRBCs, usually only singly-invaded, but sometimes multiply-invaded pRBCs were also counted; chapter 4 describes the data in more detail.

The second point to raise about the available data concerns the studies which determined (experimentally) various parameter values related to the kinetics and dynamics of RBC invasion. While the estimates extracted from these studies were, of course, vital to the overall model construction process no time-series data were given in any of these studies either. The probable reasons why no data of this type was given was discussed in section 4.3. To reiterate, the time-scale of the transition period is very short compared to the rest of the replication cycle and this has made it very difficult to obtain accurate measurements of

how many merozoites, pRBCs, uRBCs, and particularly how many MECs are present in the experimental culture at any specific time. It was also the case that the majority of the studies looking at the dynamics of invasion were not concerned with obtaining any time-series data, only the timings of individual invasion events.

Thus far global problems with the available data and relevant parameters have been pointed out. The discussion will now turn to the specific data obtained from the two different studies which were used in the model construction process. First consider figure 4.1 where the individual multiplication are displayed as points and summarised as box-and-whisker plots. Certainly the most important characteristic of these data is simply the surprisingly low values – the majority of which were centred around unity. This means that one ended up with the same number of pRBCs as were initially present after a single replication cycle, on average. The first reason for this is suspected to be the fact that only labelled pRBCs were counted when determining the re-invasion parasitemia. Recall that labelled and unlabelled uRBCs were present in the experimental culture. The reason this was done was to ensure that the possible effects different blood groups might have on the invasive potential of the invading merozoites was not a confounding variable. This is very significant when interpreting both the multiplication rate and re-invasion parasitemia since merozoites released by the initial population of pRBCs may very well have invaded uRBCs which originated from the *unlabelled* donor cell suspension. Because these were not counted as part of the re-invasion pRBC population, one must acknowledge that the published data is, to some degree, an underestimation of the true invasive and multiplicative abilities of the merozoite populations of each isolate.

In addition, these were field isolates and as such were not specifically adapted for laboratory purposes. This is the second reason proposed to explain the atypically low multiplication rate. The parameter estimates obtained from fitting the model to these data will be underestimates of the "true" values. The asynchronous nature of the cultures is expected to have had a negligible effect since the newly-invaded pRBCs were counted after 48 to 52 hours, the full length of a typical replication cycle. Put another way, all the pRBCs will have ruptured and the daughter merozoites released from these ruptured cells will have had enough time to invade the available, susceptible RBC population.

Moving to the data from Clough et al. it must be noted that it has a single peculiar issue which is that the data do not reflect the vast majority of research which consistently demonstrates that suspension cultures attain a higher re-invasion parasitemia, i.e a greater multiplication rate than their static counterparts for the same initial parasitemia. The researchers offer no concrete explanation for this result; rosetting was

also eliminated as a possible contributing factor since the researchers showed it was statically insignificant when comparing static and suspension cultures. Furthermore, the suspension cultures produced fewer multiply-invaded pRBCs than static cultures, shown in figure 4.3, which – while in agreement with other studies on this matter – makes the above result even more puzzling. Taking both of these outcomes together one sees that not only were fewer RBCs invaded in suspension culture experiments, but also that fewer individual merozoites successfully invaded. This ultimately translates into a significantly lower invasion efficiency for merozoites invading in suspension cultures compared to static cultures. One possible explanation for these seemingly anomalous results may be attributed to the speed at which the suspension cultures were agitated; this was mentioned in section 4.2. The argument is that the speed of mixing was too high for many merozoites to attach themselves to a susceptible RBC and form a tight junction which would lead to successful invasion. This is to say the rate of detachment was very high. That being said, another direct consequence of a high level of agitation is that the contact rate between merozoites and RBCs is also significantly increased, on average. One might assume this increase in contact rate would effectively negate the high rate of detachment but, this is does not seem to be the case here.

Both the base and multiple invasion mechanistic mathematical models were capable of capturing the initial linear relationship between the starting and RIP in both static and suspension cultures. This is a consistent, observable phenomenon and is rather intuitive since if there are more pRBCs at the start of a replication cycle there should be more pRBCs at the end of that particular cycle – up to a point. An interesting result from this exercise is the observation in figure 5.1 that the base model coincides almost exactly with a linear estimate of the data denoted by the dashed line forced through the origin. This result speaks directly to the utility of a mechanistic mathematical model: quantifying a linear relationship by statistical methods between two variables is, of course, a useful exercise. However, obtaining the same relationship with a mechanistic model gives one not only more but, relevant and useful information about that relationship through interpretation of the parameters in the particular model. To help illustrate this idea, consider again figure 5.1. The linear regression line has a gradient of 0.87, this is understood to be the average multiplication rate. This is the only insight one can gain from this simple, descriptive model whereas the base model provides details on the separate steps of the invasion process, not only the outcomes. A closer look at the fitted parameters governing those steps in the base model is now necessary – see table 5.2.

Each of the three first-order parameters, k_{-1}, μ and k_2, are within the same order of magnitude which

may ultimately be seen as a mostly positive result since the latter two processes are both assumed to be unaffected by neither the culture type being used nor the hematocrit of the solution. More specifically, the average lifespan of a randomly chosen merozoite (μ) was estimated at around seven and a half minutes while the rate of successful invasion (k_2) was estimated at approximately 12 and half minutes. The estimates for the rate of detachment, k_{-1}, are more problematic precisely because they are of the same order of magnitude. Intuitively one would expect the detachment rate to be higher in suspension cultures than in static cultures simply because the former are constantly agitated and thus collisions between merozoites and RBCs would be far more numerous. This translates directly into merozoites having more failed attempts at invading susceptible RBCs than they would experience in static cultures. The average value for this parameter was $3.84\,\text{min}^{-1}$ which is approximately 15 and a half seconds which is interpreted as the lifespan of the average merozoite-erythrocyte complex.

The fact that these estimates are all within the same order of magnitude speaks directly to three separate issues: the first is that of practical identifiability caused by having an insufficient number of data points, the second is that the mechanistic model itself is perhaps not detailed enough to distinguish between static and suspension cultures and the third is because there is very little experimental data pertaining to this specific process. This last point is likely the most crucial since if estimates were available for this parameter (for both culture types) one could then use the appropriate estimates with the corresponding experimental parasitemia (point) data and fit the deterministic model based on those observed ranges of values. Having this type of experimentally determined data would also address the second issue of model specificity regarding the culture being used. Unfortunately these type of data have proven to be very difficult and cumbersome to obtain and as such there are very few estimates available for this as well as other processes involved in the invasion process.

There are a few modifications which could be made to the modelling framework defined in chapter 2. For example, one simplifying assumption was that a RBC may only have one merozoite attached to its surface at any given time, i.e any merozoite-erythrocyte complex (MEC) consists of only one RBC and a single merozoite. However, videomicrographic evidence suggests that this is not always the case since it has been observed that two or more merozoites can in fact bind to the same RBC and attempt to re-orientate and penetrate that RBC one after the other. In light of this, one may decide to incorporate this phenomenon into the model structure to obtain a more realistic description of the invasion process. This possibility was considered during the model construction process but, was eventually excluded for two reasons. The first is because the aforementioned simplifying assumption (a single merozoite limit for

each MEC) was sufficient to capture the desired level of detail describing the transition period of malarial infection. Ignoring this simplifying assumption and allowing multiple merozoites to simultaneously bind to a single RBC would have added unnecessary complexity to the model structure in the context of its intended purpose; this is one of the hallmarks of the model construction process discussed in section 1.3. The second reason – related to the first – why this simplifying assumption was upheld is because there are no data available pertaining to those MECs with more than one attached merozoite except for the fact that they can be observed.

One of the defining aspects of the base model is that the singly-invaded pRBC species is the absorbing class of the system. Recall that an absorbing class in a mathematical model is one whose population can only increase or remain constant since there are no removal terms of any kind present in the ODE governing how that particular species changes over time. This was justified given that the time-scale of the transition period is short relative to the rest of the replication cycle. This assumption may however, contribute to the inability of the base model to capture the saturation effect seen in both the static and suspension cultures; figures 5.2 to 5.4 show this definitively for the static data, less so for the suspension data. Therefore, one or more mechanisms could be introduced which may cause the population of singly-invaded pRBCs to start decreasing after some initial parasitemia. In other words, the model structure has to be modified in some way in order to capture the aforementioned saturation effect; this also means that additional assumptions have to be made and adequately justified.

Possibly the simplest modification one could make is use the fact that mature schizonts have been observed to rupture normally but produce both viable and non-viable progeny [1]; this result was discussed in section 1.2.3. In addition, the event of egress abortion whereby a schizont fails to rupture was detailed in section 1.2.1 and is also worth considering. In the deterministic model these two phenomena would be relevant at the beginning of the replication cycle whereby one could first multiply the initial number of schizonts by some constant depending on what proportion of schizonts are thought to not be able to rupture (egress abortion). One could then multiply the number of released merozoites by another constant reflecting the proportion of those merozoites thought to be non-viable. Both of the aforementioned constants would be constrained between zero and one since these are understood to represent proportions of the relevant species (schizonts or merozoites). The only data available suggest that roughly half of the merozoites released by some schizonts are in fact viable. The prevalence of egress abortion has, however, not been definitively measured [29, 1, 28]. Furthermore, while the inclusion of these constants do add some biological realism to the model, they do not change the qualitative behaviour of the model

and do therefore not succeed in producing the saturation effect seen in both culture types.

Another path one could take is to consider the following scenario: what if the population of singly-invaded pRBCs was *not* an absorbing class and in fact did transition to another class in some manner or otherwise became non-viable soon after they are invaded? What is clear is that this would certainly not be the natural death rate or removal due to eventual rupture since only a single replication cycle is being considered in the model. However, a conceivable interpretation might be one related to the egress abortion phenomenon mentioned earlier. If one were to accept this view then the new term in the ODE for pRBCs would simply be a removal term of the form $-k_x x(t)$ where k_x is a first-order rate constant. The reciprocal of this new rate constant would be the time it takes for a singly-invaded pRBC to become non-viable in the sense of not rupturing at all. However, while the experimental evidence is present [29, 28], this new rate constant is still an ill-defined proposition given the time-scale and the fact that only the subsequent replication cycle would be affected since it is that cycle which would suffer the consequences of having fewer pRBCs. This, combined with the fact that the inclusion of this new parameter did not improve the fit of the model means that it was not included in the final deterministic model. Nonetheless, while this particular rate constant and its implementation were not successful in this work it may well be relevant when combining the appropriate experimental data and modelling several replication cycles over a number of weeks.

From the discussion above one is forced to retain the singly-invaded pRBC population as the absorbing class in the system. One should then rather look to modify the parameter responsible for the contact rate between merozoites and uRBCs, namely k_1. This modification considers the physical and spatial effects present during the transition period of the replication cycle. The general framework for this phenomenon of macromolecular crowding was originally developed for intracellular enzyme kinetics, however, the idea can be extrapolated to an environment where parasite-RBC interactions take place. One is looking to formulate a function dependent on time which may or may not include the densities of the species thought to be relevant to the particular process being modelled.[1] Several hypotheses could be made about what type of expression would be suitable to include which would capture the effects of spatial limitations in an appropriate manner. The most general mathematical formulation for the desired rate function would be the following:

$$k_1(t) \;=\; f\left(\mathbf{x},\; \mathbf{p},\; t\right). \tag{6.1}$$

Equation (6.1) is not to be confused with the vector function definition of an ODE given in section 1.3.

[1] The time dependence of this rate function can be either implicit or explicit.

Typically not all the species in the model will be relevant to the purpose of the rate function, but it is convention to write the function in the most general terms. The fractal-like kinetic model implements a power function in which time is explicitly taken into account. A new constant called the fractal parameter (h) is also introduced [60].

$$k_1(t) \;:=\; k_0 \cdot t^{-h}, \qquad 0 \leqslant h \leqslant 1, \quad t \geqslant 1 \tag{6.2}$$

The fractal parameter can be thought of as a measure of the dimensionality of the system. For example: in a completely homogeneous culture one requires that $h = 0$ which reduces the rate function into a time-independent rate constant, k_0. This is precisely the form of the current base mechanistic model. Conversely, the less homogeneous a solution is observed or defined to be, the closer to unity the fractal parameter will be. More generally, this is a function which decreases with time which means as time progresses the lower the rate at which contacts would be made between merozoites and RBCs. An obvious drawback of this particular kinetic model is the singularity produced at $t = 0$. This is clearly not biologically realistic since it implies that $k_1(t)$ will be infinitely large as $t \to 0$. This limitation prompted other researchers to construct a modified expression; one of these alternatives is the so-called Zipf-Mandelbrot model of reaction dynamics,

$$k_1(t) \;:=\; \frac{k_0}{(\tau + t)^h}, \qquad \tau > 0, \quad t \geqslant 1 \tag{6.3}$$

This form immediately seems more plausible since the singularity at the start of the reaction is no longer present. In addition, the parameter responsible for solving this problem (τ), has a physical interpretation: it is the time required for the reaction of interest to first be affected by any spatial constraints. In other words, a period of τ seconds/minutes must pass before macromolecular crowding will affect the reaction. This interpretation adds some respectability to equation (6.3) by incorporating a sense of biological realism in that spatial restrictions are not relevant the moment the reaction begins. That being said, the Zipf-Mandelbrot model is not strictly mechanistic, but it does have an underlying physical basis which may make it suitable for inclusion in certain mathematical models.

In the current context, this kinetic model is only plausible if one assumes that the observed saturation in both static and suspension cultures is due to spatial effects limiting the ability of merozoites to invade susceptible RBCs. Again, this is proposed to be captured by equations (6.2) and (6.3) which are decreasing functions of time and serve as a proxy for the spatial limitations experienced by the merozoites and RBCs. The evidence supporting this theory involves the ratio of the concentrations of merozoites and RBCs [19]. More specifically, a version of the phenomenon of substrate inhibition is arguably the best suited solution when attempting to account for spatial constraints.

The theory is that there exists a certain substrate concentration at which the specific reaction – merozoite invasion into RBCs in this case – proceeds at a maximal rate. Beyond this optimal substrate concentration the rate begins to decline down to a minimal rate. Again, this theory and its accompanying equations are borrowed from biochemistry where the kinetics of specific biochemical pathways are frequently modelled using these equations. Furthermore, it is always assumed that these reactions involve only a single substrate but, in the current context both merozoites and RBCs act as two different substrates. This forces one to reconsider the applicability of the classical biochemical equations for substrate inhibition kinetics. Mulchandani & Luong provide a very useful overview of various substrate inhibition equations in the context of microbial growth inhibition [61]. Several of these equations were tested with the "substrate" defined as the ratio of the concentrations of merozoites and RBCs ($m(t)/r_0$). The model which gave the best results had the following form:

$$k_1(t) = \frac{k_0 S(t)}{K_s + S(t)} \left(1 - \frac{S(t)}{S_m}\right)^n \tag{6.4}$$

where k_0 is the maximal rate, K_s the half saturation substrate concentration, S_m the maximum substrate concentration which supported invasion and n is the so-called shape factor which indicates the severity of the effects of spatial constraints on the reaction rate. The substrate (ratio of merozoites to RBCs) is denoted by $S(t)$. Figure 6.1 shows the results when equation (6.4) is incorporated into the base model. One immediately notices that this model fairs significantly better than the base or multiple invasion models based purely on the criteria of minimising the sum of squares between the data and model. This is because now the contact rate between merozoites and RBCs is a function of their ratio at any given time. This ratio is assumed to act as a proxy for the spatial effects which the base and multiple invasion models do not account for. However, while providing a significantly better fit there are several problems with this model, most of which are consequences of the parameters not being identifiable.

The new model has seven parameters in total and none were able to be uniquely identified. In addition, because of the limited amount of data all seven parameters were strongly correlated with one another, i.e the correlation coefficients were greater than 0.50 for all possible parameter combinations. This meant that the estimates could not be trusted to a reasonable degree of accuracy. One can look at specific parameters to see this explicitly. First, the estimate for the shape factor n was 5.77. This value in isolation is meaningless if one does not have another value for this parameter. In other words, the estimate here is not relative to any other estimate and thus no accurate interpretation can be made. Secondly, the values for k_{-1} and k_2 are virtually the same which means that when a merozoite comes into contact with a RBC it has an equal probability of detaching or successfully invading that RBC.

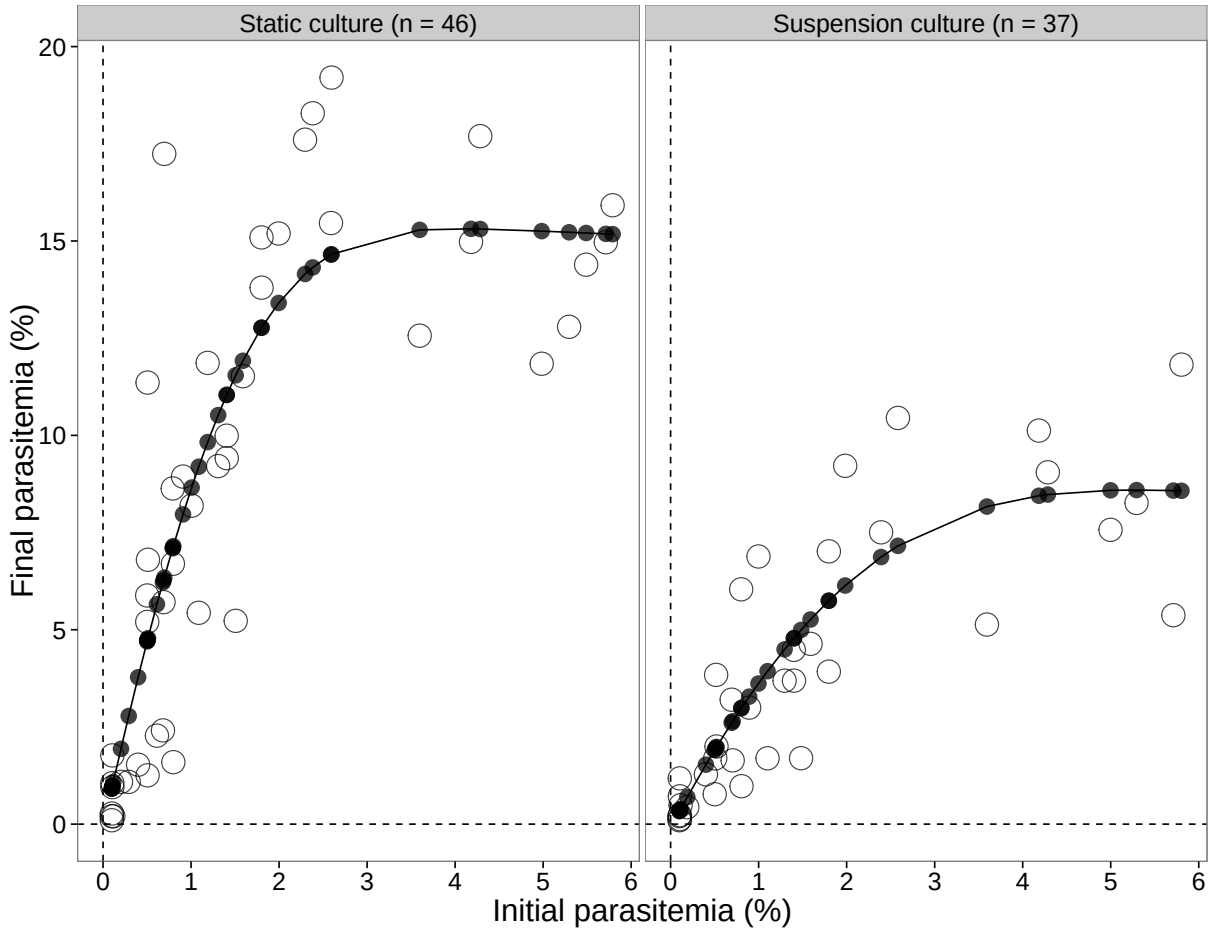

Figure 6.1: Model fit with substrate inhibition kinetics for static and suspension cultures from Clough et al. [16]. The data are shown as open circles while the fitted model is denoted by the filled points and connecting curve. The fitted parameters were as follows: $k_0 = 1.35 \times 10^{-6}$, $k_2 = 0.32$, $\mu = 6.78 \times 10^{-2}$, $k_{-1} = 0.34$, $K_s = 1.45 \times 10^{-3}$, $S_m = 1.06$ and $n = 5.77$.

This is simply not biologically realistic as seen from all of the data gathered during the literature review (including videographic data).

Turning attention now to the multiple invasion deterministic mathematical model, it was stated in section 5.3 that this model represented the data no better than the base model for either culture. Specifically, it was unable to capture the saturation effect which is the key characteristic of both cultures, particularly static cultures. The critical question now is whether to accept or reject the multiple invasion model over the base model based on the aforementioned results as well its relative complexity. The argument against accepting this model stems from a fundamental notion of a mathematical model in that it is meant to be able to describe the data in as simple a form as possible while still being biologically plausible. If one were to subscribe to this idea then the base model would undoubtedly be the more

attractive choice since it performs just as well as the more complex multiple invasion model and it does not model as many species. In addition, it contains the three essential processes which govern the invasion process. The counter argument, advocating for the acceptance of the multiple invasion model, is based on the fact that it allows for the possibility of a susceptible RBC being invaded by more than one merozoite – making it a more realistic abstraction of the physical system – while utilising the same number of free parameters and describing the data just as well as the simpler base model. A further justification for choosing the multiple invasion model is that multiply-invaded pRBCs were present in the population of pRBCs which made up the RIP data, thus making this model better suited and more relevant to describe said data especially since the number of parameters has remained the same compared to the base model. The data pertaining to multiple invasions – stratified by culture – were depicted in figure 4.3.

In light of this it is tempting to state or acknowledge that the multiple invasion model is somehow "correct" in explaining the dynamics of RBC invasion by free merozoites in an *in vitro* environment. Of course, this is simply an example of a more general fallacy in which any model, mathematical or otherwise, is deemed to be correct outside the consideration of other, related models. This relates back to one of the fundamental principles of mathematical modelling: no model is ever correct, it is only better than one or more related models which describe the same process, each in a different level of detail. Section 1.3 discussed the most important ideas surrounding construction of a mathematical model, one of which was the understanding that a mathematical model describes or embodies how much one does not know about the system being studied; this is the case in this example. To summarise: while this model is more biologically plausible and performs just as well as the base model, numerous issues such as parameter identifiability, sufficient data and distinct parameters are still present. As such, several modifications can be made to the construction to improve it's utility.

One of the modifications which could be made is to introduce two or more distinct rate parameters to account for the different invasion classes; specifically the parameter responsible for successful location and preliminary attachment k_1. Section 2.1.1 defined this parameter as the contact rate between a free merozoite and an uRBC. Currently, the multiple invasion model has a single rate constant describing this process for each invasion class which might not be the case since the contact rate between a merozoite and an uRBC would be greater than that between a merozoite and a pRBC. This is simply because uRBCs are more numerous than pRBCs, particularly in an homogeneous environment. In addition to adding a level of specificity and biological realism, this modification forces one to also consider the rate

of successful invasion which may differ in a similar manner as just described between uRBCs and pRBCs. In other words, whether it is easier or more difficult for a merozoite to invade a RBC which had previously been invaded is not captured in either the multiple invasion or base model; this rate was quantified by the first-order rate constant k_2.

Figure 5.4 effectively summarises the results of the mechanistic mathematical model and most importantly depicts the eventual implications of the fitted parameter values. The inset of figure 5.4 shows only the range of values for which experimental data were available, it is the same data and fitted model values seen in figure 5.2. While the fitted model is certainly not ideal as discussed earlier, the extrapolated plot shows an interesting prediction. The base model (shown in red) is a monotonically increasing function and attains a maximum re-invasion parasitemia of around 78 %. Because both merozoites and susceptible RBCs are required to produce a pRBC and contribute to the re-invasion parasitemia, if one or both of these species are in short supply it would limit the number of new, successful invasions, i.e the number of new pRBCs. At low initial parasitemias the number of uRBCs is much greater than the total number of released daughter merozoites at the start of the transition period.

Interestingly, an initial parasitemia of a little under 6 % creates a scenario where one would have an equal number of merozoites and uRBCs to begin the transition period assuming each mature schizont produces 16 daughter merozoites. These initial conditions are the theoretical minimum which can achieve a re-invasion parasitemia of 100 % since each uRBC would then be infected with exactly one merozoite. This has never been observed experimentally and is certainly not the case here. On the same note, researchers never utilise the full, possible range of initial parasitemias for various reasons, the primary one being that culture mediums simply cannot sustain the growth of a population of cells with such initial conditions. Studying the main graph in figure 5.4 more closely one sees an initial rise to ≈ 10 % with a subsequent decrease in the rate at which new pRBCs are being formed.[2] The main point here is that the prediction of the base mathematical model is very similar, qualitatively, to the data but not quantitatively as far as static cultures are concerned.

Finally, more can be said about the predictions made by the stochastic model presented in figures 5.6 and 5.7. The time to first contact between a merozoite and a RBC (regardless of whether that RBC had previously been invaded or not) was shown to subscribe to a Γ-distribution which agrees qualitatively with the only data available for this particular parameter [21]. However, while the model is in qualitative

[2] Mathematically this means that the derivatives of the model (denoted by f), have the following properties: $f' > 0$ and $f'' < 0$ for all initial parasitemias.

agreement it does not agree quantitatively. Specifically, it invariably overestimated the time it takes a merozoite to first contact a RBC. This could be due to many factors including the hematocrit used in the study from which the data were collected and certain limitations of the stochastic model such as having RBCs remain stationary for the entire simulation period or an inaccurate estimation of the step length and turning angle of the average merozoite. Unfortunately the data available in the literature review was not sufficient to obtain reliable estimates for these parameters.

7. Conclusion

The primary aim of the current work was to construct two mechanistic mathematical models – one deterministic and one stochastic – describing the dynamics of the transition period of *P. falciparum* malarial infection, a period characterised by the invasion of susceptible RBCs by free, extracellular merozoites. The second aim was to investigate the phenomenon of multiply-invaded pRBCs and which factors are responsible for the relative density of these particular types of pRBCs which have been experimentally observed. Two different mechanistic models were constructed (one including and the other excluding the possibility of multiple invasions); these models were then fitted to experimental data obtained from static as well as suspension cultures. Neither model was able to explain the observed trend, specifically the saturation effect which has been attributed to low levels of mixing within the experimental culture which would limit the number of interactions between merozoites and uRBCs, competition exclusion between merozoites as well as the depletion of nutrients and metabolites required for the development of all species involved. The single random walk model was then used to predict the same available data using the parameters estimated with the base mechanistic model. Unlike both mechanistic models, the random walk model was able to capture the observed saturation effect albeit at a higher re-invasion parasitemia. This was proposed to be due to spatial effects being explicitly taken into account in the the stochastic model – a visual representation of a typical simulation of the model was given in figure 5.5 where the physical limitations experienced by invading merozoites are evident.

The parameters estimated by the two mechanistic models were shown to be uniquely, structurally identifiable under certain conditions but not practically identifiable which does bring into question the accuracy of these estimates. Several modifications to the base deterministic model were proposed in chapter 6 of which a subset were able to adequately capture the saturation effect mentioned earlier. However, while these modifications do have a mechanistic basis, all of which are derived from models describing the dynamics of enzymatic reactions, identifying the parameters present in these models was unsuccessful even though circumstantial data do support their inclusion. Furthermore, additional parameters which required estimation were present in all the modifications which were tested. Their presence only exacerbated the issue of identifiability which rendered the justification of their inclusion in the deterministic models ultimately futile.

That being said, the theoretical and practical foundations of the models constructed herein are well-grounded and supported by experimental observations. Specifically, the merozoite-erythrocyte complex

(MEC) present in every model is an observable species and an integral part of the invasion process but has thus far not been included in any mathematical models of within-host malarial infection. The reasons for this were given in chapter 2 and are a consequence of practical limitations as well as the desire of researchers to investigate the progression of the disease over several days or weeks and not the inner workings of the transition period which occurs on a significantly shorter time-scale. Certainly the biggest obstacle in the current work was the insufficient amount of data with which to validate either of the mechanistic models or indeed allow use of the random walk model in a more effective manner in terms of estimating the distribution of step lengths or the intrinsic viability of the average merozoite.

The most significant conclusion gained from this research was that while the mathematical models were constructed in a theoretically sound manner in which all interactions were based on observable phenomena, the data required to robustly validate these models and acquire accurate estimates for the biologically realistic parameters was not available. From a broader perspective one may argue that this is a case where practical limitations currently overshadow the theoretical significance or insights which have been gained but as data becomes available this limitation can be lifted.

Bibliography

[1] E. M. Pasini, D. van den Ierssel, H. J. Vial, and C. H. M. Kocken, "A novel live-dead staining methodology to study malaria parasite viability.," *Malaria journal*, vol. 12, p. 190, Jan. 2013.

[2] World Health Organisation, "World Malaria Report 2013," 2013.

[3] A. G. Craig, M. F. M. Khairul, and P. R. Patil, "Cytoadherence and severe malaria.," *The Malaysian journal of medical sciences : MJMS*, vol. 19, pp. 5–18, Apr. 2012.

[4] L. Bannister and G. Mitchell, "The ins, outs and roundabouts of malaria," *Trends in Parasitology*, vol. 19, pp. 209–213, May 2003.

[5] M. Ho and N. J. White, "Molecular mechanisms of cytoadherence in malaria.," *The American journal of physiology*, vol. 276, pp. C1231–42, June 1999.

[6] D. Cromer, S. E. Best, C. Engwerda, A. Haque, and M. Davenport, "Where have all the parasites gone? Modelling early malaria parasite sequestration dynamics.," *PloS one*, vol. 8, p. e55961, Jan. 2013.

[7] O. K. Doumbo, M. a. Thera, A. K. Koné, A. Raza, L. J. Tempest, K. E. Lyke, C. V. Plowe, and J. A. Rowe, "High levels of Plasmodium falciparum rosetting in all clinical forms of severe malaria in African children.," *The American journal of tropical medicine and hygiene*, vol. 81, pp. 987–93, Dec. 2009.

[8] M. N. Wykes and J. Horne-Debets, "Dendritic cells: the Trojan horse of malaria?," *International journal for parasitology*, vol. 42, pp. 583–7, May 2012.

[9] M. J. Boyle, D. W. Wilson, and J. G. Beeson, "New approaches to studying Plasmodium falciparum merozoite invasion and insights into invasion biology.," *International journal for parasitology*, vol. 43, pp. 1–10, Jan. 2013.

[10] A. Radfar, D. Méndez, C. Moneriz, M. Linares, P. Marín-García, A. Puyet, A. Diez, and J. M. Bautista, "Synchronous culture of Plasmodium falciparum at high parasitemia levels.," *Nature protocols*, vol. 4, pp. 1899–915, Jan. 2009.

[11] A. K. Bei, T. M. DeSimone, A. S. Badiane, A. D. Ahouidi, T. Dieye, D. Ndiaye, O. Sarr, O. Ndir, S. Mboup, and M. T. Duraisingh, "A flow cytometry-based assay for measuring invasion of red blood

cells by plasmodium falciparum," *American Journal of Hematology*, vol. 85, no. 4, pp. 234–237, 2010.

[12] J. W. Zolg, A. J. Macleod, I. H. Dickson, and J. G. Scaife, "Yields MODIFICATIONS OF THE IN VITRO FALCIPARUM : PLASMODIUM YIELDS IMPROVING PARASITIC CULTURE," *Journal of Parasitology*, vol. 68, no. 6, pp. 1072–1080, 1982.

[13] A. Saul, "Plasmodium falciparum : Automated Assay of Erythrocyte Using Flow Cytofluorometry Invasion," *Experimental parasitology*, vol. 71, no. 54, pp. 64–71, 1982.

[14] R. J. W. Allen and K. Kirk, "Plasmodium falciparum culture: the benefits of shaking.," *Molecular and biochemical parasitology*, vol. 169, pp. 63–5, Jan. 2010.

[15] M. B. Hoshen, R. Heinrich, W. D. Stein, and H. Ginsburg, "Mathematical modelling of the within host dynamics of Plasmodium falciparum," *Parasitology*, vol. 121, no. 03, pp. 227–235, 2000.

[16] B. Clough, F. a. Atilola, and G. Pasvoi, "The role of rosetting in the multiplication of Plasmodium falciparum: rosette formation neither enhances nor targets parasite invasion into uninfected red cells.," *British journal of haematology*, vol. 100, pp. 99–104, Jan. 1998.

[17] J. Baum, M. Pinder, and D. J. Conway, "Erythrocyte invasion phenotypes of plasmodium falciparum in the gambia," *Infection and Immunity*, vol. 71, no. 4, pp. 1856–1863, 2003.

[18] A. K. Bei, C. Brugnara, and M. T. Duraisingh, "In vitro genetic analysis of an erythrocyte determinant of malaria infection.," *The Journal of infectious diseases*, vol. 202, pp. 1722–7, Dec. 2010.

[19] M. J. Boyle, D. W. Wilson, J. S. Richards, D. T. Riglar, K. K. A. Tetteh, D. J. Conway, S. A. Ralph, J. Baum, and J. G. Beeson, "Isolation of viable plasmodium falciparum merozoites to define erythrocyte invasion events and advance vaccine and drug development," *Proceedings of the National Academy of Sciences*, vol. 107, no. 32, pp. 14378–14383, 2010.

[20] U. Ribacke, K. Moll, L. Albrecht, H. Ahmed Ismail, J. Normark, E. Flaberg, L. Szekely, K. Hultenby, K. E. M. Persson, T. G. Egwang, and M. Wahlgren, "Improved in vitro culture of Plasmodium falciparum permits establishment of clinical isolates with preserved multiplication, invasion and rosetting phenotypes.," *PloS one*, vol. 8, p. e69781, Jan. 2013.

[21] A. J. Crick, T. Tiffert, S. M. Shah, J. Kotar, V. L. Lew, and P. Cicuta, "An automated live imaging platform for studying merozoite egress-invasion in malaria cultures.," *Biophysical journal*, vol. 104, pp. 997–1005, Mar. 2013.

[22] A. D. Pillai, R. Addo, P. Sharma, W. Nguitragool, P. Srinivasan, and S. A. Desai, "Malaria parasites tolerate a broad range of ionic environments and do not require host cation remodelling," *Molecular microbiology*, vol. 88, no. February, pp. 20–34, 2013.

[23] S. Glushakova, D. Yin, N. Gartner, and J. Zimmerberg, "Quantification of malaria parasite release from infected erythrocytes: inhibition by protein-free media.," *Malaria journal*, vol. 6, p. 61, Jan. 2007.

[24] L. Mata-Cantero, M. J. Lafuente, L. Sanz, and M. S. Rodriguez, "Magnetic isolation of Plasmodium falciparum schizonts iRBCs to generate a high parasitaemia and synchronized in vitro culture.," *Malaria journal*, vol. 13, p. 112, Jan. 2014.

[25] S. C. Bhakdi, A. Ottinger, S. Somsri, P. Sratongno, P. Pannadaporn, P. Chimma, P. Malasit, K. Pattanapanyasat, and H. P. H. Neumann, "Optimized high gradient magnetic separation for isolation of Plasmodium-infected red blood cells," *Malaria journal*, vol. 9, no. 38, pp. 1–9, 2010.

[26] J. P. Dalton, C. G. Demanga, S. J. Reiling, J. Wunderlich, J. W. L. Eng, and P. Rohrbach, "Large-scale growth of the Plasmodium falciparum malaria parasite in a wave bioreactor," *International journal for . . .*, no. February, pp. 5–10, 2012.

[27] K. Chotivanich, R. Udomsangpetch, K. Pattanapanyasat, W. Chierakul, S. Looareesuwan, and N. White, "Hemoglobin E : a balanced polymorphism protective against high parasitemias and thus severe P falciparum malaria Hemoglobin E : a balanced polymorphism protective against high parasitemias and thus severe P falciparum malaria," *Blood*, pp. 1172–1176, 2002.

[28] T. Tiffert, V. L. Lew, H. Ginsburg, M. Krugliak, L. Croisille, and N. Mohandas, "The hydration state of human red blood cells and their susceptibility to invasion by Plasmodium falciparum.," *Blood*, vol. 105, pp. 4853–60, June 2005.

[29] S. Glushakova, G. Humphrey, E. Leikina, A. Balaban, J. Miller, and J. Zimmerberg, "New stages in the program of malaria parasite egress imaged in normal and sickle erythrocytes.," *Current biology : CB*, vol. 20, pp. 1117–21, June 2010.

[30] P. M. Lantos, a. D. Ahouidi, a. K. Bei, C. V. Jennings, O. Sarr, O. Ndir, D. F. Wirth, S. Mboup, and M. T. Duraisingh, "Erythrocyte invasion profiles are associated with a common invasion ligand polymorphism in Senegalese isolates of Plasmodium falciparum.," *Parasitology*, vol. 136, pp. 1–9, Jan. 2009.

[31] N. Gomez-Escobar, A. Amambua-Ngwa, M. Walther, J. Okebe, A. Ebonyi, and D. J. Conway, "Erythrocyte invasion and merozoite ligand gene expression in severe and mild Plasmodium falciparum malaria.," *The Journal of infectious diseases*, vol. 201, pp. 444–52, Feb. 2010.

[32] G. A. Butcher, "Factors affecting the in vitro culture of Plasmodium falciparum and Plasmodium knowlesi.," *Bulletin of the World Health Organization*, vol. 57 Suppl 1, pp. 17–26, Jan. 1979.

[33] J. A. Simpson, N. J. White, K. Silamut, K. Chotivanich, and S. Pulrittayakamee, "Red cell selectivity in malaria : a study of multiple-infected erythrocytes," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 93, pp. 165–168, 1999.

[34] D. Dorin-Semblat and A. Sicard, "Disruption of the PfPK7 gene impairs schizogony and sporogony in the human malaria parasite Plasmodium falciparum," *Eukaryotic . . .*, vol. 7, pp. 279–85, Feb. 2008.

[35] L. Mancio-Silva, J. J. Lopez-Rubio, A. Claes, and A. Scherf, "Sir2a regulates rDNA transcription and multiplication rate in the human malaria parasite Plasmodium falciparum.," *Nature communications*, vol. 4, p. 1530, Jan. 2013.

[36] H. B. Reilly, H. Wang, J. a. Steuter, A. M. Marx, and M. T. Ferdig, "Quantitative dissection of clone-specific growth rates in cultured malaria parasites.," *International journal for parasitology*, vol. 37, pp. 1599–1607, Dec. 2007.

[37] M. S. Tucker, T. Mutka, K. Sparks, J. Patel, and D. E. Kyle, "Phenotypic and genotypic analysis of in vitro-selected artemisinin-resistant progeny of Plasmodium falciparum.," *Antimicrobial agents and chemotherapy*, vol. 56, pp. 302–14, Jan. 2012.

[38] L. Wentzinger, S. Bopp, H. Tenor, J. Klar, R. Brun, H. P. Beck, and T. Seebeck, "Cyclic nucleotide-specific phosphodiesterases of Plasmodium falciparum: PfPDEalpha, a non-essential cGMP-specific PDE that is an integral membrane protein.," *International journal for parasitology*, vol. 38, pp. 1625–37, Dec. 2008.

[39] M. Eichner, H. H. Diebner, L. Molineaux, W. E. Collins, G. M. Jeffery, and K. Dietz, "Genesis, sequestration and survival of Plasmodium falciparum gametocytes: parameter estimates from fitting a model to malariatherapy data.," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 95, no. 5, pp. 497–501, 2001.

[40] B. K. Mutai and J. N. Waitumbi, "Apoptosis stalks Plasmodium falciparum maintained in continuous culture condition.," *Malaria journal*, vol. 9 Suppl 3, p. S6, Jan. 2010.

[41] N. Mideo, T. Day, and A. F. Read, "Modelling malaria pathogenesis.," *Cellular microbiology*, vol. 10, pp. 1947–55, Oct. 2008.

[42] M. R. Miller, L. Rberg, A. F. Read, and N. J. Savill, "Quantitative analysis of immune response and erythropoiesis during rodent malarial infection," *PLoS Comput Biol*, vol. 6, p. e1000946, 09 2010.

[43] S. Huijben, W. a. Nelson, A. R. Wargo, D. G. Sim, D. R. Drew, and A. F. Read, "Chemotherapy, within-host ecology and the fitness of drug-resistant malaria parasites.," *Evolution; international journal of organic evolution*, vol. 64, pp. 2952–68, Oct. 2010.

[44] N. Mideo, V. C. Barclay, B. H. K. Chan, N. J. Savill, A. F. Read, and T. Day, "Understanding and predicting strain-specific patterns of pathogenesis in the rodent malaria Plasmodium chabaudi.," *The American naturalist*, vol. 172, pp. 214–38, Nov. 2008.

[45] X. Gao, K. Gunalan, S. S. L. Yap, and P. R. Preiser, "Triggers of key calcium signals during erythrocyte invasion by Plasmodium falciparum.," *Nature communications*, vol. 4, p. 2862, Jan. 2013.

[46] P. R. Gilson and B. S. Crabb, "Morphology and kinetics of the three distinct phases of red blood cell invasion by plasmodium falciparum merozoites," *International Journal for Parasitology*, vol. 39, no. 1, pp. 91 – 96, 2009.

[47] D. T. Riglar, D. Richard, D. W. Wilson, M. J. Boyle, C. Dekiwadia, L. Turnbull, F. Angrisano, D. S. Marapana, K. L. Rogers, C. B. Whitchurch, J. G. Beeson, A. F. Cowman, S. A. Ralph, and J. Baum, "Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte.," *Cell host & microbe*, vol. 9, pp. 9–20, Jan. 2011.

[48] P. Srinivasan, W. L. Beatty, A. Diouf, R. Herrera, X. Ambroggio, J. K. Moch, J. S. Tyler, D. L. Narum, S. K. Pierce, J. C. Boothroyd, J. D. Haynes, and L. H. Miller, "Binding of Plasmodium

merozoite proteins RON2 and AMA1 triggers commitment to invasion.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 13275–80, Aug. 2011.

[49] V. L. Lew and T. Tiffert, "Is invasion efficiency in malaria controlled by pre-invasion events?," *Trends in parasitology*, vol. 23, pp. 481–4, Oct. 2007.

[50] K. L. Harvey, P. R. Gilson, and B. S. Crabb, "A model for the progression of receptor-ligand interactions during erythrocyte invasion by Plasmodium falciparum.," *International journal for parasitology*, vol. 42, pp. 567–73, May 2012.

[51] E. a. Codling, M. J. Plank, and S. Benhamou, "Random walk models in biology.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 5, no. 25, pp. 813–834, 2008.

[52] P. M. Kareiva and N. Shigesada, "Analysing insect movement as a correlated random walk," *Oecologia*, vol. 56, pp. 234–238, 1983.

[53] F. Bartumeus, M. G. E. D. Luz, G. M. Viswanathan, and J. Catalan, "Animal Search Strategies : A Quantitative Random-Walk Analysis Published by : Ecological Society of America ANIMAL SEARCH STRATEGIES : A QUANTITATIVE RANDOM-WALK ANALYSIS," vol. 86, no. 11, pp. 3078–3087, 2008.

[54] S. Benhamou, "How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension?," *Journal of Theoretical Biology*, vol. 229, no. 2, pp. 209–220, 2004.

[55] P. Bovet and S. Benhamou, "Spatial analysis of animals' movements using a correlated random walk model," *Journal of Theoretical Biology*, vol. 131, no. 4, pp. 419–433, 1988.

[56] Y. Adams and J. A. Rowe, "The Effect of Anti-Rosetting Agents against Malaria Parasites under Physiological Flow Conditions.," *PloS one*, vol. 8, p. e73999, Jan. 2013.

[57] K. Yahata, M. Treeck, R. Culleton, T.-w. Gilberger, and O. Kaneko, "Time-Lapse Imaging of Red Blood Cell Invasion by the Rodent Malaria Parasite Plasmodium yoelii," vol. 7, no. 12, pp. 1–7, 2012.

[58] M. Abkarian, G. Massiera, L. Berry, M. Roques, and C. Braun-Breton, "A novel mechanism for egress of malarial parasites from red blood cells.," *Blood*, vol. 117, pp. 4118–4124, Apr. 2011.

[59] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[60] S. Schnell and T. E. Turner, "Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws.," *Prog. Biophys. Mol. Biol.*, vol. 85, no. 2-3, pp. 235–260, 2004.

[61] a. Mulchandani and J. Luong, "Microbial inhibition kinetics revisited," *Enzyme and Microbial Technology*, vol. 11, pp. 66–73, 1989.

Appendix A.

This appendix describes various so-called parasite virulence factors (PVFs) and their notation used throughout this thesis. Let the following denote a pRBC containing n ring-stage parasites at some time t in the τ-th replication cycle,

$$x_{r,\tau}^{(n)}(t) \;:=\; \left\{ x_{r,\tau}^{(n)}(t) \in \mathbb{N} \cup \{0\} \;\middle|\; n \in [1, N] \;;\;\; \tau \in \mathbb{N} \right\}. \tag{A.1}$$

where N is the maximum number of parasites which are able to invade a single RBC. This is simply a practical assumption since only a finite number of parasites are be able to invade a single RBC. The parasite (merozoite) and uninfected RBC (uRBC) populations are represented by $m(t)$ and $r(t)$, respectively. The units for all three populations is that of concentration, i.e cells per microlitre (cells/µl).

Using parasitic measures defined to be between zero and one is often preferred or even necessary when manipulating the PVFs described below and determining relationships between them. As such, all PVFs defined as percentages will have a parallel definition which is restricted between zero and unity and which will be denoted with a bar above the index.

A.1 Initial parasitemia

In an experimental setting, the following definition may be used:

initial parasitemia (IP) *The percentage of total RBC which are parasitised before rupture,*

$$p_0 \;:=\; \frac{x_s}{r_0} \cdot 100\% \tag{A.2a}$$

$$:=\; \frac{X}{r_0} \cdot 100\% \tag{A.2b}$$

Equation (A.2a) refers to the case where all the pRBCs are at the mature schizont stage as is the case when one or more synchronisation procedures have been performed; equation (A.2b) represents the scenario where no synchronisation has been performed and pRBCs at several distinct developmental stages may exist (an asynchronous pRBC population). One should note that the initial number of pRBCs of a particular replication cycle will be the same as the final amount of the previous replication cycle; in other words, the following equality exists,

$$x_{s,\tau+1} \;=\; \sum_{n=1}^{N} x_{r,\tau}^{(n)} \tag{A.3}$$

The summation in equation (A.3) represents the different classes of pRBCs, i.e those which have been invaded by one parasite $(n = 1)$, by two parasites $(n = 2)$, etc. This applies to all replication cycles with the exception of the first since there are no invaded cells to consider prior to the start of the experiment. In addition, this scenario only makes sense when some form of synchronisation has been performed prior to the first cycle.

A.2 Re-invasion parasitemia

This quantity – also called the final parasitemia – is the one most often reported in studies looking at the multiplication ability of the merozoite population:

re-invasion parasitemia (RIP) *The percentage of total RBC which are parasitised after a single replication cycle,*

$$p_r := \frac{\sum_{n=1}^{N} x_{r,\tau}^{(n)}}{r_0 - \sum_{i=1}^{\tau-1} \sum_{n=1}^{N} x_{r,\tau-i}^{(n)}} \cdot 100\% \tag{A.4}$$

The n-summations in equation (A.4) are the same as that explained in appendix A.1, the τ-summation in the denominator represents all the replication cycles prior to the current one, this is seen by the upper limit $(\tau - 1)$. In other words, equation (A.4) is quantifying the percentage of pRBCs after a particular replication cycle relative to the number of remaining uRBCs. What is important to note is that this definition assumes that no additional healthy RBCs are added at any point during the experiment; this can however, be easily integrated into the definition if required.

Regarding clinical or field data (patient isolates), it does not make sense to distinguish between initial and final parasitemia since the blood obtained from an infected individual is sampled without knowledge of the developmental stages of the population of pRBCs; this is due to the asynchrony of infection within the host. Because of this, parasitemia determined from patient isolates cannot be labelled as "initial" or "final" but, simply as the peripheral parasitemia.

A.3 Multiple invasion measures

The following two quantities explicitly take the number of multiply-invaded pRBCs into account. The first is relative to the total number of RBCs and the second to the total number of pRBCs,

multiple invasion parasitemia (MIP) *The percentage of multiply-invaded pRBCs out of the total number of RBCs after re-invasion,*

$$p_r^M := \frac{\sum_{n=2}^{N} x_{r,\tau}^{(n)}}{r_0 - \sum_{i=1}^{\tau-1}\sum_{n=1}^{N} x_{r,\tau-i}^{(n)}} \cdot 100\% \tag{A.5}$$

multiple invasion proportion (MiP) *The percentage of multiply-invaded pRBCs out of the total number of pRBCs after re-invasion,*

$$p^M := \frac{\sum_{n=2}^{N} x_{r,\tau}^{(n)}}{\sum_{n=1}^{N} x_{r,\tau}^{(n)}} \cdot 100\% \tag{A.6}$$

Note the summations in the numerators of both definitions begin at $(n = 2)$ since only multiply-invaded pRBCs are being considered.

A.4 Merozoite parasitemia

A final definition regarding parasitemia pertains to individual parasites and not simply the pRBCs themselves,

merozoite parasitemia (MP) *The number of merozoites involved in all successful invasions observed after re-invasion divided by the total number of RBCs,*

$$p_m := \frac{\sum_{n=1}^{N} n x_{r,\tau}^{(n)}}{r_0 - \sum_{i=1}^{\tau-1}\sum_{n=1}^{N} x_{r,\tau-i}^{(n)}} \cdot 100\% \tag{A.7}$$

This definition, besides the multiple invasion proportion and parasitemia defined earlier, is perhaps the most useful measure when considering multiply-invaded pRBCs since individual invasion events are taken into account. Two obvious yet important observations are the following: $p_m \geqslant p_r$ and $p_m = p_r \iff p^M = 0$.

A.5 Invasion rate and invasion efficiency

The next three definitions deal with how well and how quickly free parasites are able to invade susceptible RBCs. Instead of simply considering the total number of RBCs which have been invaded one could form a more precise definition by determining the number of successful parasite invasions; this definition has a clear, direct link to the number of multiply-invaded pRBCs similar to the merozoite parasitemia (MP) defined earlier in equation (A.7).

invasion efficiency *The percentage of merozoites involved in all successful invasions observed after re-invasion out of the total number of merozoites present immediately after schizogony of the previous population of pRBCs,*

$$\Gamma := \frac{\sum_{n=1}^{N} n x_{r,\tau+1}^{(n)}}{m_0^{\tau}} \cdot 100\% \tag{A.8a}$$

$$= \frac{\sum_{n=1}^{N} n x_{r,\tau+1}^{(n)}}{\epsilon \cdot \sum_{n=1}^{N} x_{r,\tau}^{(n)}} \cdot 100\% \tag{A.8b}$$

The parameter ϵ in equation (A.8b) is the average number of merozoites released by a single rupturing schizont; it is often necessary to use an average since one can rarely determine the exact number of merozoites at the beginning of the initial cycle as is assumed in equation (A.8a).

A second definition of invasion efficiency simply uses the number of pRBCs regardless of how many times a particular RBC has been invaded; this is termed the pseudo-invasion efficiency:

pseudo-invasion efficiency *The percentage of pRBCs observed after re-invasion relative to the total number of merozoites present immediately after schizogony of the previous population of pRBCs,*

$$\Gamma^* := \frac{\sum_{n=1}^{N} x_{r,\tau+1}^{(n)}}{m_0^{\tau}} \cdot 100\% \tag{A.9a}$$

$$= \frac{\sum_{n=1}^{N} x_{r,\tau+1}^{(n)}}{\gamma \cdot \sum_{n=1}^{N} x_{r,\tau}^{(n)}} \cdot 100\% \tag{A.9b}$$

By definition, $\Gamma^* \leqslant \Gamma$. The most intuitive interpretation of this secondary definition is as the *lower limit* of invasion efficiency since individual merozoites are not explicitly considered. The third definition below considers how quickly invasions occur and may be thought of as the rate of change of the number of merozoites which successfully invade,

invasion rate *The number of new invasions observed during a pre-defined period of time,*

$$\Gamma_r := \frac{\sum_{n=1}^{N} n x_{r,\tau}^{(n)}}{\Delta t} \tag{A.10}$$

where Δt is the chosen length of time. The invasion rate may also be interpreted as the incidence of invasion. This is arguably the most difficult PVF to measure since, depending on the chosen time-scale, newly-invaded pRBCs may not be distinguishable from ones invaded prior to the start of the current time period. Methods to allow invasion to occur only at specific times do exist however.

All of the definitions so far have been defined as percentages, with the exception of the invasion rate which has units of cells/s. The final three PVFs are dimensionless quantities which define the fold increase in the number of pRBCs after one replication cycle and the difference between the expected and observed number of multiply-invaded pRBCs.

A.6 Multiplication Rates

These two indices are arguably the most important indices when dealing with *in vitro* experimental cultures and the propagation potential of the merozoite population in general. The first defines the fold increase in the number of pRBCs after one replication cycle,

parasitised erythrocyte multiplication rate (PEMR) *The re-invasion parasitemia divided by the initial parasitemia,*

$$\text{PEMR} := \frac{p_r^\tau}{p_0^\tau} \tag{A.11a}$$

$$= \frac{\sum_{n=1}^{N} x_{r,\tau+1}^{(n)}}{\sum_{n=1}^{N} x_{r,\tau}^{(n)}} \tag{A.11b}$$

merozoite multiplication rate (MMR) *The merozoite parasitemia of a particular cycle divided by the merozoite parasitemia of the previous cycle,*

$$\text{MMR} := \frac{p_m^{\tau+1}}{p_m^{\tau}} \tag{A.12a}$$

$$= \frac{\sum_{n=1}^{N} n x_{r,\tau+1}^{(n)}}{\sum_{n=1}^{N} n x_{r,\tau}^{(n)}} \tag{A.12b}$$

A.7 Selectivity Index

This index determines whether there would be more or less multiply-invaded pRBCs than would be expected by chance alone or if other factors might be influencing their relative density. The term "selectivity index" refers to the fact that certain uRBCs are more susceptible to invasion than others by virtue of their age, i.e. the amount of time they have spent in circulation.

selectivity index *The number of multiply-invaded pRBCs observed after re-invasion divided by the number expected from a Poisson distribution.*

The reason for choosing the Poisson distribution is because of its two primary assumptions: events should occur independently and at random.[1] In addition, this distribution is suited for data where the chances of a particular event occurring are low as they are in this context. Put simply, an SI > 1 indicates more multiply-invaded pRBCs than would be expected if invasion were truly a random process with respect to spatial and susceptibility considerations. The expected number of RBCs invaded by n merozoites is defined as the total number of pRBCs multiplied by the probability of being invaded by n merozoites divided by the probability of being invaded at all;

$$E\left(X = n\right) := \sum_{n=1}^{N} x_r^{(n)} \cdot \frac{P\left(X = n\right)}{P\left(X \geqslant 1\right)} \tag{A.13}$$

recall the notation used to denote pRBCs in appendix A.1. The SI can thus be defined as

$$\text{SI} := \frac{\sum_{n=2}^{N} x_r^{(n)}}{E(X \geqslant 2)} \tag{A.14}$$

[1] These two conditions are hallmarks of a completely homogeneous culture.

The mean for the distribution is calculated from the following equation: $\lambda = -\log(1-\overline{p_r})$; this translates into the parasitemia being approximated as the probability of a RBC being invaded by any number of merozoites: $P(X \geq 1) = 1 - e^{-\lambda}$. With this information the SI can be written as shown below.

$$\text{SI} = \frac{\sum_{n=2}^{N} x_r^{(n)}}{\sum_{n=1}^{N} x_r^{(n)}} \cdot \frac{\overline{p_r}}{\overline{p_r} + \log\left(1-\bar{p_r}\right)^{(1-\overline{p_r})}} \tag{A.15}$$

Besides using equation (A.15) to calculate the precise value of the SI in various scenarios, studying this PVF in an analytical manner may prove useful when establishing relationships with regard to other PVFs. The simplest way is to find an approximation and determine in what range that approximation is valid; one such approximation is the following:

$$\text{SI} \approx \frac{\sum_{n=2}^{N} x_{r,\tau}^{(n)}}{\sum_{n=1}^{N} x_{r,\tau}^{(n)}} \cdot \frac{2}{\overline{p_r}} \tag{A.16a}$$

$$= 2 \cdot \frac{\overline{p_r^M}}{(\overline{p_r})^2} \tag{A.16b}$$

The approximation given in equation (A.16b) is obtained by computing the Taylor series of the expression in $\overline{p_r}$ of equation (A.15) evaluated at zero and discarding all but the first term. The validity of this approximation is limited to when the parasitemia is below 10%, i.e when $\overline{p_r} < 0.10$.