Mathematical models for exploring insecticide resistance in vector mosquitoes.

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy

by

Susana do Carmo Pinto Barbosa

November 2012

Contents

Abstract iii

Publications iv

Contributors Statement v

Acknowledgements vi

List of Figures vii

List of Tables ix

Abbreviations x

1 Introduction 1

1.1 Vector borne diseases 1
1.2 Vector control 2
1.3 Insecticide resistance 5
1.4 Mathematical modelling of insecticide resistance 7
1.5 Thesis outline 13

2 Challenges in estimating insecticide selection pressures from mosquito field data 15

2.1 Introduction 15
2.2 Model and methods 17
2.3 Results 21
2.4 Discussion 30

3 The importance of modelling the spread of insecticide resistance in a heterogeneous environment: the example of adding synergists to bed nets 36

3.1 Introduction 36
3.2 Methods 38

3.3 Results 46
3.4 Discussion 54

4 Modelling insecticide impact and resistance in a malaria vector mosquito population 58
4.1 Introduction 58
4.2 Description of the model 61
4.3 Population reproductive rate 69
4.4 Malaria transmission 70
4.5 Analysis 71
4.5.1 Fixed points 72
4.5.2 Sensitivity analysis 76
4.6 Scenarios of insecticide deployment 78
4.6.1 Single interventions 79
4.6.2 Combined interventions 85
4.7 Conclusion 89

5 Final remarks and conclusions 92

Bibliography 96

A Comment 112

B Co-authorship paper 114

Abstract

The emergence and spread of insecticide resistance compromises the control of mosquito borne diseases that are responsible for millions of deaths every year in tropical and subtropical areas. Mathematical modelling is a valuable tool that can be used to explore different aspects of the development and management of insecticide resistance. We have used standard population genetics theory and ecological modelling techniques for developing models to evaluate the spread of resistance in the field.

We started by developing a methodology to quantify the strength of selection for resistance occurring in nature. We used data from Mexico on the mosquito *Aedes aegypti* and a maximum likelihood methodology to estimate the selection and dominance coefficients driving the evolution of resistance in the field. We additionally explored the impact of poor data collection, data that combine information from different locations, and the consequences of selection and dominance coefficients varying over the sampling time period. This analysis highlighted factors highly relevant to field work such as the need for frequent surveillance in discrete sentinel sites.

The use of insecticidal bed nets represents the primary tool for the prevention of malaria worldwide. It is of extreme importance to maintain their efficacy against mosquitoes, which has been undermined by the development of insecticide resistance. We assed the contribution of a novel design of bed nets in delaying insecticide resistance while at the same time determining the important parameters in driving resistance in an heterogeneous environment. We showed that this new bed net can indeed contribute to the delay of the spread of resistance, but surprisingly could have the reverse effect in specific circumstances.

Finally we developed a model for the vector of malaria, that considers the stage-structured nature of the mosquito life cycle and, most importantly, explicitly incorporates insecticide resistance. It can be used to understand the population dynamics of mosquitoes throughout their entire lifecycle while analysing the impact of vector control interventions, alone and in combination, and the spread of insecticide resistance that those interventions induce. We showed that targeting the larval stages has the greatest effect on the adult population followed by targeting non host-seeking female adults. According to our results, low levels of resistance can induce failure of interventions, and the rate of spread of resistance is faster when insecticides target the larval stages.

Publications

Two chapters of this thesis describe content already published in scientific journals:

Chapter 2: Barbosa S, Black WC IV, Hastings I: **Challenges in Estimating Insecticide Selection Pressures from Mosquito Field Data**. *PLoS Negl Trop Dis* 2011, bf5(11): e1387.

I also co-authored (as a minor author) the following manuscript:

Lynd A, Weetman D, Barbosa S, Egyir Yawson A, Mitchell S, Pinto J, Hastings I, Donnelly MJ: **Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s.*.** *Mol Biol Evol* 2010, **27**(5):1117-25.

I collaborated by estimating values for selection and dominance using the methodology developed on Chapter 2, the paper can be found in Annex B.

Chapter 3: Barbosa S, Hastings I: **The spread of insecticide resistance in a heterogeneous environment: the impact of adding synergists to bednets**. *Malar J* 2012, **11**:258.

Chapter 4 will be submitted as soon as possible with authorship Barbosa S, Hastings I and Chitnis N.

Due to publications editorial differences, chapters structure and language style varies sightly.

Contributors Statement

In compliance with the University of Liverpool rules on PhD thesis, I hereby present the contributions to all the chapters:

Chapter 1 S Barbosa wrote the introduction, I Hastings supervised the writing.

Chapter 2 S Barbosa conducted the analysis wrote the draft paper and adapted the paper to a thesis chapter, I Hastings supervised the analysis and the writing. W C Black IV contributed with the field data and reviewed the paper manuscript. Three anonymous referees contributed with suggestions to improve the paper manuscript.

Chapter 3 S Barbosa conducted the analysis and adapted the paper to a thesis chapter, I Hastings supervised the analysis and the writing. Two anonymous referees contributed with suggestions to improve the paper manuscript.

Chapter 4 S Barbosa conducted the analysis and wrote the chapter, I Hastings and N Chitnis supervised the analysis and I Hastings supervised the writing.

Chapter 5 S Barbosa wrote the conclusions, I Hastings supervised the writing.

Acknowledgements

I would like to express my gratitude to the people who have helped me in the making of this thesis. First of all, my supervisor, Ian Hastings, who provided the needed guidance in this research topic completely new to me four years ago. I thank him for giving me the chance to come to Liverpool, for all his patience and the reading of countless drafts. I would also like to thank my co-supervisor, Hilary Ranson, for her valuable help through the years and Bill Black for the field data and comments on the second chapter. I cannot thank enough Nakul Chitnis for all his generous support all these years. Tiago Antão for joyful hours of conversation and his computer assistance throughout. Many thanks are due to the Wolfson Bioinformatics Suite and Simon Wagstaff, who provided additional computer power without which I could not have done most of the work. My warmest thanks to Mary Creegan for her flawless administrative support and unconditional smile. Instituto Gulbenkian de Ciência and the organizers (Jorge Carneiro and Manuela Cordeiro) of the Ph.D Program in Computational Biology and to my colleagues: Afonso, Assunção, Bruno, Hugo Fernandes, Hugo Martins, Zé, Paula, Ricardo, Rodrigo[2] and Sónia, thank you for an unforgettable year. This work was supported by research grant SFRH/BD/33534/2008 from Fundação para a Ciência e Tecnologia, Portugal and Siemens Portugal.

Then there are the people that made Liverpool a home, that helped me get through the difficult times, and provided the invaluable emotional support, entertainment and care. Katherine for being my life saviour during the adaptation period, and Rita for being around since (literally) take off, Bruno Manso for the many hours listening at 'Egg' in the early times and Ana Reimão for creating a really great home and sharing it with me. Alison Mbekeani for the laughs, English lessons, and friendship. Francesca Tamarozzi, an example of perseverance and organisation (and grumpiness), for being a supportive colleague and a very good friend.

My parents, for all their love and comprehension, that have seen me moving further away through the years, and have always been supportive.

Finally to Hans, a graduate school courtesy, the last, but the one I'm leaving with.

List of Figures

1.1 Major biochemical mechanisms conferring resistance to the different classes of insecticides in mosquitoes. 6

2.1 The typical change in insecticide concentration in the field over time. . . 16
2.2 Simulated evolution of resistance allele frequency. 22
2.3 Effect of spatial heterogeneity on estimates of initial allele frequency, dominance and selection parameters with selection coefficient constant at different values. 27
2.4 Effect of spatial heterogeneity on estimates of initial allele frequency, dominance and selection parameters with dominance coefficient constant at different values. 28
2.5 Effect of temporal heterogeneity on estimates of initial allele frequency, dominance and selection coefficients parameters 29

3.1 Plot of partial rank correlation coefficients between each of the parameters in the model and the two outcome variables: mean fitness and ratio of change of resistance allele frequency, in both sexes. 47
3.2 Histogram of the ratio of resistance allele frequency at generation 10 in the extreme case of a fully effective ($k = 0$) compared to an inefficient ($k = 1$) synergist. 48
3.3 How synergists affect the spread of resistance. 49
3.4 An example of a classification tree of outcomes $y > 1$ (class 1) and $y < 1$ (class 0) using all the parameters in the model except k. 52

4.1 The mosquito population stage-structured model 62
4.2 Plot of partial rank correlation coefficients between parameters of the model and the number of adult females at day 600. 77
4.3 Simulations of the impact on female adult population of independently decreasing survival at different stages. 81
4.4 Simulations illustrating the level of resistance required to induce failure of control interventions, aimed at each of the stages independently. . . . 84

4.5 Simulations of the impact on female adult population of decreasing survival at different stages in combination. 87
4.6 Simulations illustrating the level of resistance required to induce failure of control of interventions used in combination 88

List of Tables

2.1 Details of 100 idealized simulated datasets. 21
2.2 Specifications of datasets of Figure 2.2. 24
2.3 Estimated p_0, h and s from field data. 25

3.1 Model structure: niches, exposure and genotype fitnesses within each niche 39
3.2 Parameters, symbols and subscripts used in the construction of the model. 43
3.3 Parameters range of values used in simulations. 44
3.4 Logistic regression: parameter coefficient estimates, standard error, p-value and the odds ratio. 50
3.5 Dynamics of spread of resistance allele frequency. 53

4.1 Description of parameters for the model of mosquito population dynamics. 63
4.2 Mosquito immature life stages summary. 64
4.3 Parameters values used in the fixed points and sensitivity analysis, the values used are identical for males and females when not explicitly stated otherwise. 73
4.4 Parameter values used for the calculation of R_{0m} and derivation of a population threshold level that interrupts malaria transmission (M). . . 80

Abbreviations

CI Confidence Interval

Bt *Bacillus thuringiensis*

DDPR Density-Dependent Population Regulation

DDT Dichlorodiphenyltrichloroethane

DNA Deoxyribonucleic Acid

HWE Hardy-Weinberg Equilibrium

IRS Indoor Residual Spraying

ITN(s) Insecticide Treated Net(s)

kdr Knockdown Resistance

LHS Latin Hypercube Sampling

LLN(s) Long Lasting Net(s)

MACE Modified acetylcholinesterase

ML Maximum Likelihood

PBO Piperonyl Butoxide

PRCC Partial Rank Correlation Coefficient

RR Homozygous Resistant Genotype

RS Heterozygous (Resistant/Susceptible) Genotype

SS Homozygous Susceptible Genotype

WHO World Health Organization

Chapter 1

Introduction

1.1 Vector borne diseases

Vector borne diseases constitute today, as they did in the past centuries, a great source of concern for global public health. The use of preventive and therapeutic measures such as improvements in sanitation and hygiene, implementation of vaccination programs, use of antiparasitic drugs and vector control [1] have significantly decreased their impact in developing countries, and led to their eradication in parts of the developed world, but vector borne diseases still account for around 17% of the infectious diseases burden [2, 3]. They continue to be significant nowadays because despite all past efforts and successes, we have witnessed the reemergence and geographical expansion of diseases like plague, dengue/dengue hemorrhagic fever, yellow fever and West Nile virus, that were once effectively controlled. Additionally, there was the emergence of new diseases such as lyme disease and ehrlichiosis [4].

The reemergence in recent decades is closely tied to unprecedented global changes in population demography, land use, urbanisation, trade and travel. These changes have lead to environmental modifications, creating conditions that favour the proliferation of vector species and increasing contact between humans and vectors [3–8]. Moreover, the burden of human vector borne diseases on regional ecologies and economies is enhanced by plant and animal vector borne diseases [4].

Climate change may also play a role. Temperature, rainfall and humidity changes can affect vector development, reproduction, behaviour, survival rates and geographical range [9–11]. Nevertheless, some authors suggest that lifestyle and public health measures are expected to counterbalance climate effects [12], reinforcing the idea that socioeconomic conditions are the key players in the prevalence of vector borne diseases [2].

The most common vectors are arthropods, with mosquitoes, flies, sand flies, lice, fleas, ticks and mites transmitting a large number of infectious agents such as protozoa, bacteria, viruses and helminths [13]. Mosquitoes are the most important arthropod vector and there are approximately 3500 species widespread around the world [14]. Mosquitoes

of the genera *Culex*, *Aedes*, and *Anopheles* are responsible for the transmission of *Filaria spp*, West Nile virus, japanese encephalitis, dengue/dengue hemorrhagic fever, chikungunya, and malaria [2, 15], that are major causes of morbidity and mortality every year worldwide. This thesis focuses on mosquitoes, but we believe results may have a general application to other vectors.

1.2 Vector control

Vector control represents a significant part of past and current strategies for vector borne diseases control. The methods in use were conceived by considering the mosquito life cycle and behaviour. The life cycle is composed of four distinct stages: egg, larva and pupa, that require water for development, and a free flying adult. Adult females lay their eggs on water bodies and, once hatched, larva, pupa and adult follow, in a time scale that depends on the species and temperature. The biting behaviour of females, that require a blood-meal for fertile egg development, is diverse. Some species prefer animal hosts while others prefer humans. Foraging for a host can occur at dusk, dawn, during the day or during the night. After successfully feeding a female mosquito requires a safe place to rest in order to digest the blood-meal and develop the eggs, a process that can last several days. Some species prefer to rest outdoors (exophily) and others indoors (endophily), in houses and cattle sheds [16]. Following egg maturation the female mosquito begins searching for a suitable breeding site where she will lay her eggs and the cycle repeats. Females can lay eggs more than once in their lifetime, so they will repeatedly forage for a blood meal.
There are several vector control methods available, the main ones are described here:

Environment management: This method involves a range of temporary or permanent environmental manipulations to limit vector reproduction, survival, or abundance, modification of human behaviour to limit the contact with vectors and others like the strategic placement of diversionary hosts such as cattle (zooprophylaxis) [17, 18].

Biological control: Potential interventions include use of vector predators like fish or copepods [19] and bacterial larvicides such as *Bacillus thuringiensis israelensis* (Bti) [20] rending breeding sites unsuitable.

Sterile insect technique: The most used methods use radiation or chemicals to sterilise insects and involve the release of overwhelming numbers of sterile insects to reduce reproduction. It has proved very successful in controlling several agricultural pests although with less impact in mosquito populations [21, 22]. New methods are being developed and include the genetic modification of mosquitoes to render them

refractory to infection, i.e., reduce vector competence [23] and population control using RIDL (release of insects carrying a dominant lethal genetic system) [24]. There are still some important challenges and risks that need to be addressed before use in vector control programmes, but there is support for continued research in this area [25].

Insecticide use is undoubtedly the predominant method (and sometimes the only affordable) [2]. Humans have used pesticides to protect their crops since the advent of agriculture. The first known pesticide was elemental sulphur. Classical ancient Mediterranean writings show that Greeks and Romans were using sulphur to kill pests that afflicted crops [26]. Inorganic arsenic, mercury, lead, and later organic nicotine sulphate, pyrethrum and rotenone (in the 19th century) were the dominant chemicals until Paul Müller discovered that dichlorodiphenyltrichloroethane (DDT) was effective in killing insects in 1939 (it had been first synthesized in 1874) [27, 28].

The current available insecticides are grouped into the following classes according to their chemical properties and origin:

Organochlorines: Chlorinated hydrocarbons. Comprising DDT and analogues, hexachlorohexane and cyclodienes. These are very stable compounds with long term persistence, slow degradation and bioaccumulative properties [29, 30]. DDT targets the voltage-gated sodium channels (which generate nerve action potentials) and has repellency and irritancy effects [30]. Hexachlorohexane and cyclodienes target the gamma-aminobutyric acid (GABA) major neurotransmitter in the nervous system [2, 30].

Organophosphates: Most are esters or amides of organically bound phosphoric or pyrophosphoric acid. They are much less stable and more expensive than organochlorines [2, 29] and are generally not persistent in the environment. Their mode of action is the inhibition of the acetylcholinesterase enzyme by phosphorylation.

Carbamates: Carbamates are acid esters derived from carbamic acid. They inhibit acetylcholinesterase enzyme action buy a process called carbamylation [29].

Pyrethroids: Derived from pyrethrins, they are a group of esters originally extracted from some *Chrysanthemum* species. Currently there are natural and synthetic formulations available. They are highly toxic to insects but present a moderate acute toxicity to humans and other mammals [31] and are easily biodegraded. The primary mode of action is disruption of voltage-gated sodium channels function, similar to that of DDT [32]. They also have an excito-repellent effect [33], reducing biting through feeding inhibition, shorter landing times and undirected flight [34].

During and after World Word II (WWII) organochlorines such as DDT were the main insecticides employed, but their use gradually declined due to toxicological problems and environmental contamination [27] and were replaced by organophosphates, carbamates and pyrethroids. Despite huge efforts to find insecticides with different modes of action, insecticides available today still belong to the four classes described above and, including organochlorine DDT, are used in public health [2].

The main methods of insecticide deployment used worldwide are listed below. Some rely on the dusk and night foraging behaviours and indoor resting habits of important vector species to ensure that a number of potentially lethal interactions with insecticide-treated surfaces occur [35]:

Larviciding: Much like the use of biological interventions mentioned before, the employment of insect growth regulators, oils, carbamates and organophosphates within defined water boundaries is very effective provided the ability to find high proportion of breeding sites.

Adulticides Space Spray: Space spraying of organophosphates and pyrethroids is used mostly in emergency situations to control outbreaks, by achieving a rapid reduction of adult vector density. It is mostly used for control of dengue [2, 36].

Indoor Residual Spraying (IRS): Indoor residual spraying targets indoor resting vectors [36]. IRS on walls is the favoured method due to general applicability and simplicity. All four classes of insecticides are recommended, but there has been a shift in use from DDT to pyrethroids [2]. It has proven very effective against African vectors of malaria [36].

Insecticide Treated Materials (ITM): Includes insecticide treated nets (ITNs) and other materials such as curtains and sheeting. It is a very effective measure if appropriately used at very high levels of coverage. ITNs impregnated with pyrethroids (the only class of insecticide recommend by the World Health Organization - WHO) is the most common and one of the most effective measures in protecting against malaria vectors. As a result, its use has been scaled up upon WHO recommendation of full coverage of all people at risk. [37]. However, insecticide treated nets have a small impact on species that bite during the day.

1.3 Insecticide resistance

The prolonged exposure and widespread use of a small portfolio of compounds against large mosquitoes populations, with short life cycles and abundant progeny, inevitably gives rise to the emergence of resistance to the insecticides. The definition of resistance by WHO is 'the development of an ability in a strain of some organisms to tolerate doses of a toxicant, which would prove lethal to the majority of individuals in a normal population of the same species' [38].

The use of DDT during and after WWII in troops and civilians in Europe and in the USA resulted in the eradication of typhus and malaria in these areas. However, subsequent development of mosquitoes resistant to DDT contributed to the failure of the WHO campaign for global eradication of malaria started at the height of DDT use in 1955. Resistance is still considered to be the most important hindrance in the successful control of vector borne diseases [2, 15].

The first report of insecticide resistance, is believed to be that of Melander in 1914 describing lime sulphur resistance in the San Jose scale, *Quadraspidiotus perniciosus*, in California [2, 39]. Almost a century later, mechanisms allowing survival to insecticide exposures have been selected in many species and to all 4 classes of insecticides. The number of reported resistant species in the Arthropod Pesticide Resistance Database (http://www.pesticideresistance.org) [39] from 1914 to the present (May 2012) is of 575 species, 119 of which are mosquitoes, i.e., belong to the *Culicidae* family.

The mechanisms known to confer some degree of resistance to the action of insecticides can be grouped into the following categories:

Metabolic resistance: Resistant insects may metabolise or destroy the insecticides before they are able to be toxic [40]. They possess enhanced levels or more efficient forms of insecticide degrading enzymes that may also have a broader spectrum of activity. Three major enzyme groups are known to be responsible for metabolically based resistance to all four classes of insecticides: (1) mixed function oxidases, (2) hydrolases or esterases, and (3) glutathione S-transferases [15, 41].

Target site resistance: Target site insensitivity in the nervous system occurs when the insecticide no longer binds effectively to the site of action. Important target sites such as sodium channel, acetylcholinesterase, and gamma-aminobutyric acid receptor have been implicated in insecticide resistance. This type of resistance is associated with non-silent point mutations within structural genes, that do not cause a loss of primary function of the target site. The resistance phenotype known as knockdown resistance (*kdr*, resistance to paralysis whether reversible or not) is due to reduced target site sen-

sitivity in the voltage-gated sodium channels [15, 29, 41]. Figure 1.1 shows the impact of metabolic and target site resistance on the current available insecticides.

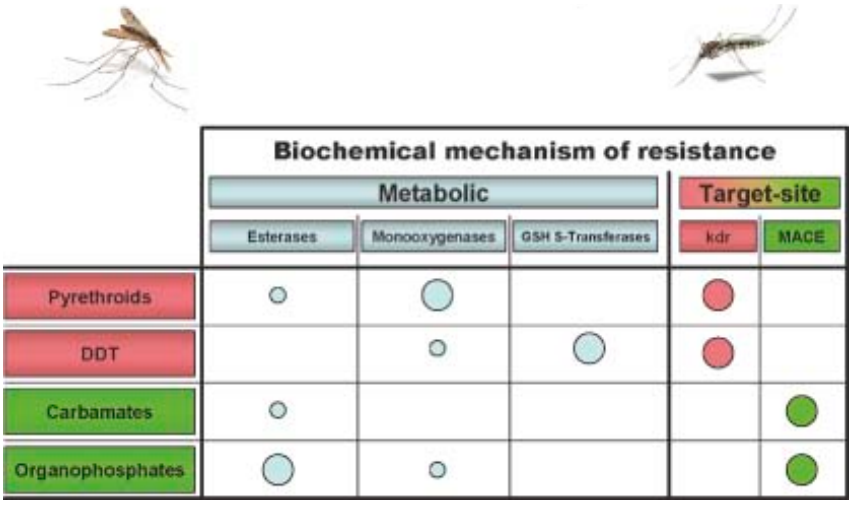

Figure 1.1: Major biochemical mechanisms conferring resistance to the different classes of insecticides in mosquitoes. A large spot indicates an important resistance mechanism; a smaller spot means this mechanism has been described but is considered to be of lesser importance. Adapted from [42] (reprint licence 2941441050130).

Cuticular resistance: Reduced uptake by over expression of cuticular proteins and reduced penetration of insecticide through the cuticle may also play a role in resistance [43], which can enhance the efficiency of metabolic detoxification and excretion [44].

Behavioural resistance: Modification in behaviour that helps to avoid the lethal effects of insecticides. It is stimulus-dependent, such as the repellent and irritant properties of some insecticides. An exposed population can be driven to adopt new behaviours in order to avoid contact with a lethal dose. It is different from stimulus-independent, protective avoidance mechanisms, such as exophily that do not require any contact with insecticides [29]. There are concerns that widespread use of insecticides may have selected for genetic changes in behaviour.

Multiple resistance: Multiple resistance is the development of resistance to more than one insecticide class due to direct exposure of a population to multiple classes of insecticides, normally at different points in time. Consequently it hinders the application of previous used insecticides [2, 41].

Cross resistance: Cross resistance is the phenomenon whereby a mechanism that renders a population resistant to one class of insecticides makes them resistant to others [2]. This is especially likely if the insecticides share the same mode of action. For ex-

ample, the *kdr* phenotype is known to confer resistance to both DDT and pyrethroids. Interactions between different resistance mechanisms are also important because they can act synergistically and provide high levels of resistance. The consequence of multiple and cross resistance is the reduction of the alternative insecticides available [41].

Another difficulty faced by vector control programs is the impact of agriculture. Insecticides are primarily used in agriculture, in fact it accounts for around 90% of all insecticide deployment [2], and can threaten the success of vector control campaigns by selecting for resistance in advance or in parallel, in both the larval and adult stages of the vector. The evidence for agriculture impact, mostly circumstantial, usually falls into one of the following categories: the appearance of resistance in the vector species before insecticide use in vector control; temporary decrease or suppression of vector populations without vector control; seasonal fluctuations of vector resistance following use of insecticide in agriculture (correlation in time); higher levels of resistance in areas with agricultural spraying compared with areas of only vector control (correlation in space); and a correlation between the intensity of insecticides use in agriculture and the resistance level in vector [45–47]. Besides the use of the same chemicals, agriculture can impact disease transmission by creating new breeding sites for vectors.

1.4 Mathematical modelling of insecticide resistance

Effective resistance management strategies require the knowledge of the mechanisms behind resistance in individual insects and some understanding about the dynamics of resistance in the populations. One of the tools available for research of insecticide resistance is mathematical modelling. Although mathematical models are simplified representations of reality they have been used to clarify concepts in the development and management of insecticide resistance, project the consequences of different assumptions, organise data, identify hypothesis and appropriate experiments in a relatively fast, safe and inexpensive way [48, 49].

According to the reviews of Tabashnik (1990) [48] and Hoy (1998) [50] the models used for insecticide resistance research usually differ in the problem addressed, the type of modelling, the basic assumptions and the factors considered.

The choice of the modelling approach depends mostly on the complexity of the model [48]. A mathematical model can represent the underlying processes in either deterministic or stochastic forms. In a deterministic model every set of variable states [1] is

[1] In brief the lexicon used in mathematical modelling: *state variables* are the variables being tracked explicitly in the model, that change over time, e.g. resistance level; *parameters* are the rates that govern movement between classes in the model, that remain constant over time, e.g. level of insecticide use or migration rate. If a particular parameter is not varied during a particular analysis, it may be referred to as a *constant* [51].

uniquely determined by parameters in the model and by previous states of these variables. A deterministic model produces the same results under the same set of initial conditions. On the other hand, in a stochastic model, variable states are described by probability distributions, therefore accounting for randomness.

The dynamics of the process under study can be explored either analytically or using simulation methods. Analytical models, for which a comprehensive toolbox of analysis methods is available, are relatively simple mathematical descriptions that can be solved mathematically and are used to uncover fundamental principles [48, 51]. Simulation models on the other hand, utilise numerical techniques to solve problems for which analytic solutions are impractical or impossible. Simulation models are generally more complex and realistic and are especially valuable for assessing the influence of a large number of factors and for sensitivity analysis [48, 51].

Modelling insecticide resistance evolution is an extremely complex process because it explores the interactions of a multitude of factors that vary with species, population and location. Here we list some of the generally accepted factors that influence the spread of insecticide resistance (reviewed in [29, 52]):

Genetic: Initial frequency of resistance gene, mutation rate, dominance relationship, past selection by other insecticides and fitness of genotypes.
Biological: Duration of life cycle and number of progeny, mating system and population size.
Ecological: Mobility, genetic isolation, migration, variation of ecological conditions.
Operational: Proportion of population exposed, dosage, persistance of insecticide, existence of refugia, life stage exposed, mode and pattern of application.

Many years of research have provided some understanding of the way the factors listed above combine to influence insecticide resistance evolution [53]. A great percentage of the research effort has been devoted to the impact of insecticide resistance on agriculture with findings then being applied in public health.

It was recognised by Georghiou in 1972 [54] that the rate at which development of insecticide resistance proceeds clearly depends on genetic factors. Seminal modelling papers by this author explored genetic factors in combination with what he regarded as 'more subtle influences exerted by the ecology of the population, such as isolation, inbreeding, and reproductive potential'. Dominance, initial gene frequency, refugia, immigration, and reproductive potential were studied using deterministic models and concluded that the rate of resistance evolution was slowed when the population was diluted by susceptible immigrants, the control intensity was high having a great impact on population density and when there were fitness differences between the different genotypes [55]. These were the first general conclusions, many following papers ex-

plored the same factors from different perspectives.

Comins (1977) [56] examined the effect of immigration dependent on dominance, initial allele frequency and population density. He observed as Georghiou had (1977) [55], that for recessive genes migration from untreated populations delays emergence and spread of resistance. Nevertheless, at some critical point, dependent on the initial resistance frequency, resistance in the untreated area increases by diffusion from the treated area and resistance spreads. Taylor and Georghiou (1979) [57] investigated potential adjustments of dominance and migration for suppression of resistance, suggesting control methods such as the release of susceptible individuals and variation of insecticide dosage to affect dominance. More complex finite population models that include randomness by Caprio and Tabashnik (1992) [58] later offered contrasting insights suggesting that in many field situations, gene flow may speed evolution of insecticide resistance. Additionally, Pasteur and Raymond (1996) [59] propose that mutations transforming susceptible genes into resistance genes are extremely rare, implying that the wide distribution of resistance in some species is due to migration. Migration effects on the spread of resistance is still an active research topic today, particularly in agriculture, mostly to study the effects of refugia on Bt crops (transgenic crops containing a gene from *Bacillus thuringiensis*) [60, 61].

Other aspects have been explored, such as the assumption that resistance development depends directly on generation time, because some amount of selection is expected to occur every generation, yet it has been shown to have no impact if selection pressure is daily constant, highlighting the influence of the intensity of selection. There is little doubt that increasing the proportion of individuals treated will increase the frequency of resistance [48]. Moreover, more complex models demonstrated that the influence of generation time should not be generalised because it depends on the interactions of so many factors with diverse effects (linear and non-linear) [62].

The mode of inheritance constitutes yet another fundamental assumption in population genetics resistance models. Most models assume that resistance is monogenic, preferable to avoid complex models and supported by the prevailing view that resistance in the field is conferred by one or two loci of major effects [48, 63, 64] (the term monogenic has been used in some instances to include two loci [65]). There has been an ongoing debate among scientists as to whether resistance is monogenic or polygenic. This is a particularly difficult issue to clarify because the results from field and laboratory experiments seem contradictory. In theory the number of genes selected to confer resistance depends on whether selection acts within or outside the phenotypic distribution of the susceptible population. Selection from within selects polygenic resistance, by combining factors that have minor effects, whereas selection outside the distribution selects for rare mutations in single genes of major effects [66–68]. It is generally postulated that the differences between field and laboratorial experiments are due to

the presumed high selection intensities in the field, that select outside the initial distribution of susceptible phenotypes, therefore favouring resistance alleles of major effects, contrasting to laboratory experiments that have typically lower intensity of selection that favours polygenic control of resistance [69]. Groeters and Tabashnik (2000) [70] present the less polarised view on this contentious issue so far, suggesting that resistance is affected by many genes, but that the distribution of effects across loci is not uniform, which can lead to one or a small number of loci often accounting for most of the resistance but also contemplating a polygenic basis for resistance in the field. Their results suggest that if major genes for resistance are present, they will increase in frequency more rapidly than minor genes under a wide variety of conditions (refugia, immigration, etc) and a possible reason why most laboratory experiments have failed to select for monogenic resistance is that laboratory populations are usually small and unlikely to contain rare resistant mutations. Consequently, to mimic field evolution in the laboratory it is more important to use a large, diverse sample from the field than to use extremely high selection intensities to induce mutations.

Quantitative genetics provide a range of theoretical and empirical tools that avoid this issue of single gene effects. These techniques make no assumptions regarding the number of genes involved, and offer an alternative to predict the speed and potential amount of genetic change involved in resistance and the direction, speed and extent of genetic change in correlated fitness traits [71]. The expression of quantitative traits depends on environmental factors as well as the actions of one or more genes, each with one or more alleles, rendering quantitative tools appropriate independently of the inheritance mechanism [71, 72]. One of the central concepts in quantitative genetics is the heritability of a trait, h^2, defined as the ratio of additive genetic variance to total phenotypic variance. It is an important characteristic, of polygenically determined traits, because it provides a means of predicting future evolutionary responses to selection and also provides insights into the influence of the trait on the organism fitness [73].

Some understanding of the basics of spread of insecticide resistance was crucial for the exploration of resistance management strategies. Management strategies focus on operational factors that are or can be brought largely under human control. The modelling of management strategies requires the track of resistance frequencies and population size in time and space. In his review on the 'Principles of insecticide resistance management' Georghiou in 1994 [53] considered that the strategies used or suggested for management of resistance can be grouped under three principal categories: (i) management by saturation, (ii) management by moderation and (iii) management by multiple attack.

Management by saturation uses tactics that completely overcome insects defences by for example using high dosages of insecticides that kill individuals that would be resis-

tant to lower dosages, or the use of synergist to potentiate the insecticide. Nonetheless, Tabashnik in 1990 [74] using three and four allele models that assumed resistance based on gene amplification showed that in some situations resistance can be potentiated by overwhelming high concentrations of insecticides.

Management by moderation is more conservative and suggests the maintenance of susceptibles in the population with low insecticide pressure. The model developed by Georghiou and Taylor in 1977 [75] was the first to examine the evolution of resistance considering some operational factors in combination: dosage, population density threshold, refugia and alternation of schedules of application. This model showed that in some cases lower insecticide dose levels and less intense treatments can help delay the spread of resistance particularly if resistance confers reproductive disadvantage. In 2001 Carriere and Tabashnik [76] revisited this subject and concluded the factors favouring reversal of resistance in cases of high dosage and refuge strategies are non–recessive costs of resistance, low initial resistance allele frequency, large refuges, incomplete resistance and density–independent population growth in refuges.

The multiple attack strategy considers that control can be achieved by exerting pressure on the population from multiple sources. Many models explored the success of strategies like the use of mixtures, rotations and mosaics. The rational behind the simultaneous use of unrelated insecticides is based on the principle that if resistance to each of the insecticides is independent and initially rare the arising of double resistance is very unlikely [77]. The deployment of insecticides in rotation relies on the assumption that resistance confers some degree of fitness disadvantage and there is no cross-resistance between insecticides used, so that the frequency of resistance to one insecticide will decline during deployment of another insecticide [78]. Success of mixtures requires a small population size or an untreated portion of the population [78] and depends on the initial allele frequencies, recombination, effective dominance, escape and linkage disequilibrium, but a major problem according to Mani (1985) [79] is the choice of the two insecticides mostly due to cross-resistance. Further to these caveats, polygenic models have concluded that application of more than one insecticide in a mixture might not be a good strategy because it can increase the overall selection intensity and hasten the evolution of resistance [78]. Nonetheless it is advocated that under certain conditions mixtures can retard resistance, e.g. provided resistance is not fully dominant [77], more effectively than rotations or mosaics. Birch and Saw (1997) [80] intended to unite the debate around the impact of mixtures by developing a model that identifies mixtures that have potential for delaying the spread of resistance. The spatial application of insecticides in mosaics was considered by Mani [81] to have limited application for retarding resistance.

Lenormand and Raymond in 1998 [82] proposed a different strategy for management. Instead of attempting to delay the appearance and spread of resistance suggested aim-

ing to maintain it at low equilibrium values. They used a two-locus model, considered treated areas size, dominance, fitness disadvantage of resistance and concluded that it is possible to achieve an optimal size for the treated area where a minimal and stable density reaches equilibrium, and where resistance genes cannot invade.

In summary, decades of modelling of resistance management came to variable conclusions clearly depending on the assumptions that are made, although most suggest that resistance selection is slowed but not completely stopped by the management tactics described. Despite these inconsistent results WHO [83, 84] recommends the use of insecticides in rotations, mosaics or mixtures during interventions, and/or to combine interventions when possible in a effort to maximise the time period for which current available insecticides provide useful disease control. There are field based data, like the success of the onchocerciasis control programme, that uses pre-planned rotations of insecticides in larviciding and a trial in Mexico, that tested fine scale mosaic and rotation strategies directly, to support these recommendations [84]. In the Mexico trial, DDT resistance did not revert completely towards susceptibility, while pyrethroid resistance increased more in areas under pyrethroid treatment alone than those in the rotation and mosaic areas. No major difference in the performance of the mosaic and rotation strategies were found.

Recently Read (2009) *et al.* [85] proposed a new approach to effective vector control, in particular malaria control. They argue that exerting weak selection for insecticide resistance by targeting only old mosquitoes that have already laid most of their eggs can decrease developing of resistance and extend the useful lifespan of insecticides and at the same time decrease transmission of disease since most malarial mosquitoes do not live long enough to transmit the disease. Following this idea Koella *et al.* (2009) [86] discussed potential options to use insecticides in alternative ways, such as the combination of a late-acting insecticide and larvicide to sustainably control malaria. Further exploration of this subject has been made by Gourley *et al.* (2011) [87] by developing a model that explores the effects of late-acting insecticides (by including a time delay on insecticide effect after exposure) and concluded that, although late-acting insecticides could not prevent the rise of resistance it can be delayed considerably. Glunt *et al.* (2011) [88] explored this concept in an experimental setting and concluded that low concentration formulations targeting older mosquitoes might have the capacity to reduce disease transmission without strong selection for resistance. [2]

Currently major international efforts are in progress to control and even attempt to globally eradicate malaria by scaling-up significantly IRS and ITNs. These efforts will surely drive resistance evolution and considering current restrictions on approved chemicals, there are virtually no options for resistance management for ITNs (since

[2]We posted our concerns regarding this conclusions in a comment posted online on the Journal's website, that can be found in Annex A.

there have not yet been approved any suitable insecticide mixture for ITNs, seen as the most promising tactic for management of resistance in bed nets [89]). To keep ITNs effective is the next big challenge in public health vector control. The focus of management is shifting from the sole use of insecticides to the use of combined strategies as part of a global integrated vector management [3]. This process for vector populations management, aimed at reducing or interrupting transmission of disease [90], considers the knowledge of factors influencing local vector biology, disease transmission and morbidity, and the use of a range of interventions. It also advocates the implementation of monitoring systems to allow for immediate action and the anticipation of resistance when planning interventions [83].

1.5 Thesis outline

This thesis uses mathematical modelling to explore different aspects of the spread of insecticide resistance in mosquito vectors of disease. The work was conducted in the context of malaria transmission by *Anopheles gambiae*, except for the first chapter, where we used field data from *Aedes aegypti*, the vector of dengue (among others).

The core of the mathematical methods used in this thesis is based on population genetics theory. The field of population genetics studies the temporal and spatial changes of frequencies of types (alleles, genes, genotypes, gametes) in entire populations of organisms subject to various ecological and genetic influences [91] using Mendel's laws and other genetic principles [92]. Along the way we integrated population genetics with ecological modelling to explore the dynamics of insecticide resistance in mosquito populations under vector control.

In the introduction we have discussed how mathematical tools have been used to assess different aspects of resistance, from fundamental questions such as the factors that drive resistance emergence to optimisation of management strategies to mitigate its impact. One aspect that has been overlooked, and is the the focus of Chapter 2, is the quantification of the spread of resistance in the field. In order to efficiently manage resistance one of the first questions to answer, once resistance onset has been acknowledged, should be: how fast is it evolving? To provide an answer to this question we developed a simple maximum likelihood method to measure the strength of selection acting on the different genotypes, using changes in allele frequencies through time. It also provides an estimate of the dominance relationship between genotypes, that has a substantial influence on the rate of spread. We discuss all the challenges associated with such a task, and assess the impact of temporal and spatial heterogeneities on the estimations. This methodology could be used for surveillance purposes, monitoring the effectiveness of interventions to tackle disease transmission and resistance.

In Chapter 3 we moved away from using empirical data and examined the contribution

of a new generation of long lasting insecticidal nets (usually refereed to as LLNs) in delaying the spread of insecticide resistance, through the development of a genetic model. Treated bed nets are one of the most widespread tools for combating malaria transmission, thus innovative approaches are required to address the growing challenge that insecticide resistance development poses. This net prototype incorporates an insecticide synergist on the roof besides being impregnated with the usual pyrethroid insecticide. The manufactures believe that it gives increased efficacy against pyrethroid-resistant malaria vectors. We provided a theoretical viewpoint on the (dis)advantages of such a new tool while exploring the dynamics of resistance considering the spatial structure in which LLNs are usually deployed. We give some insights on the parameters that are implied in driving resistance in such heterogeneous environments, considering, as well, sexual heterogeneities in insecticides exposures.

Chapter 4 originated from the need to incorporate the effects of insecticide resistance in models of malaria epidemiology. There are modelling works on mosquitoes in the literature in malaria transmission, however, the metamorphic structure of mosquitoes populations has, for the great part, been ignored by assuming homogeneous mosquito populations. We developed a stage-structured model of *Anopheles gambiae* life cycle, based on a system of difference equations that, most importantly, explicitly includes insecticide resistance by tracking different genotypes for males and females. We performed an analysis of the dynamics of the model by numerically estimate fixed points and investigating their stability. We proceed by performing a sensitivity analysis by executing the model for a collection of simulated parameters and observe the resulting change in model behaviour, namely the impact on the number of adult female mosquitoes. We used an hypothetical parameter setting to explore the impact on the population of using insecticides targeted at different stages of mosquitoes development. In order to link the developed model with malaria transmission we determined a threshold number of female adult mosquitoes below which transmission is expected to be interrupted. We simulated the most common insecticidal interventions used for vector control: larvicides, pupacides, LLNs and indoor residual spraying while examining the effect, on vector control, of the emergence of resistance. Besides the investigation of the use of single interventions we also explored the efficacy on vector control of interventions being used in combination and the impact of resistance in the effectiveness of such combinations.

In Chapter 5 we summarise our results, policy implications and future research.

Chapter 2

Challenges in estimating insecticide selection pressures from mosquito field data

Abstract

Insecticide resistance has the potential to compromise the enormous effort put into the control of dengue and malaria vector populations. It is therefore important to quantify the amount of selection acting on resistance alleles, their contributions to fitness in heterozygotes (dominance) and their initial frequencies, as a means to predict the rate of spread of resistance in natural populations. We investigate practical problems of obtaining such estimates, with particular emphasis on Mexican populations of the dengue vector *Aedes (Stegomyia) aegypti*. Selection and dominance coefficients can be estimated by fitting genetic models to field data using maximum likelihood (ML) methodology. This methodology, although widely used, makes many assumptions so we investigated how well such models perform when data are sparse or when spatial and temporal heterogeneity occur. As expected, ML methodologies reliably estimated selection and dominance coefficients under idealised conditions but it was difficult to recover the true values when datasets were sparse during the time that resistance alleles increased in frequency, or when spatial and temporal heterogeneity occurred. We analysed published data on pyrethroid resistance in Mexico that consists of the frequency of a Ile1,016 mutation. The estimates for selection coefficient and initial allele frequency on the field dataset were in the expected range, dominance coefficient points to incomplete dominance as observed in the laboratory, although these estimates are accompanied by strong caveats about the possible impact of spatial and temporal heterogeneity in selection.

2.1 Introduction

We consider the problem of measuring the strength of selection pressure for insecticide resistance in mosquito field populations and show how changes in the frequencies of the alleles at a single locus can be used to estimate the selection acting on each genotype. This type of data is collected for the identification of genetic mechanisms of resistance and/or during monitoring programs of vector control campaigns. The method we developed extends that described earlier by DuMouchel and Anderson in 1968 [93] for laboratory populations. Laboratory based conditions differ significantly from the field. In the laboratory insecticide assays are conducted over standardized range of doses

and concentrations that may not account for field situations such as decay rates and exposure characteristics. Following insecticide deployment in the field, concentration decreases and there is a selective window of time at lower concentrations (Figure 2.1), where resistant heterozygotes do not die but susceptible homozygotes are still killed, therefore acting as dominant when under more standardised conditions it may appear to be recessive. This is relevant because dominance relationships between susceptible and resistance alleles affect the rate of spread of resistance.

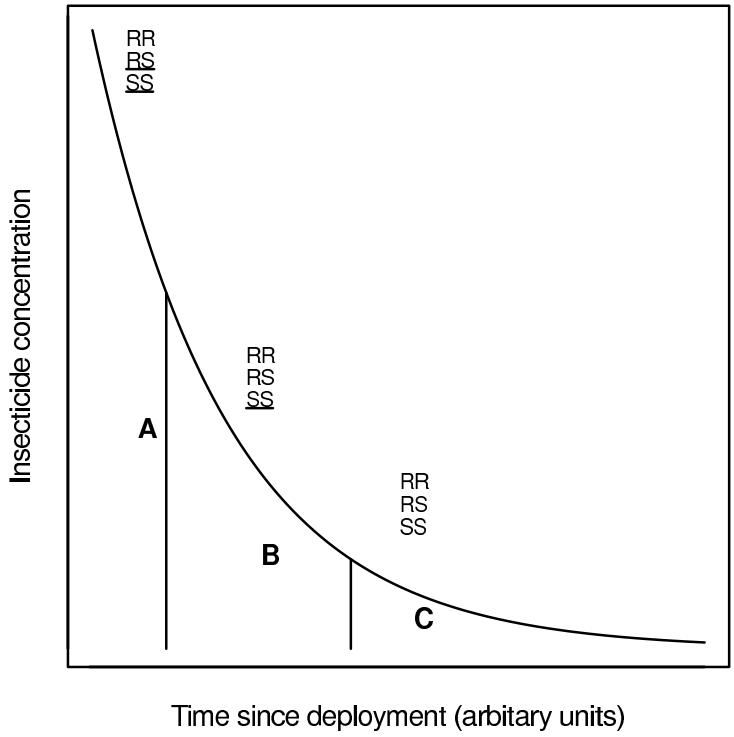

Figure 2.1: The typical change in insecticide concentration in the field over time. As concentration decays with time after deployment there is a differential survival of genotypes. In period A the RR genotype (homozygote resistant) will survive while the RS (heterozygote) and SS (homozygote susceptible) dies: this makes the R allele recessive in this period. In period B both RR and RS survive making the R allele dominant in this period. In period C all genotypes can survive so no selection occurs. These are windows of selection, adapted from Hastings and Watkins (2006) [94].

Using a maximum likelihood (ML) procedure and a recursive genetic model that tracks the changes in the resistance allele frequencies at a single locus it is possible to estimate a selection coefficient (s), a coefficient quantifying dominance (h) and the initial frequency of the resistance allele (p_0), key parameters that determine the dynamics of resistance. The model provides a straightforward way to obtain these values with the least complex dataset possible. However, field data on the spread of resistance is often

suboptimal: datasets may be small, may only track one period of dynamics (typically early and late stage of spread) or may be pooled from different locations. In this paper we discuss the challenges associated with this approach. We used published data on pyrethroid resistance from *Aedes (Stegomyia) aegyptii*, throughout Mexico [95] on the frequency of the Ile1,016 mutation, one of the mutations in the voltage-gated sodium channel gene known to confer resistance to pyrethroids (this is known as knockdown resistance, a term applied to insects that fail to lose coordinated activity immediately after exposure).

2.2 Model and methods

The genetic model we employ assumes a single autosomal locus conferring insecticide resistance in a diploid sexually reproducing population, with non-overlapping generations and assuming random mating; these are standard assumptions in population genetics models. There are two possible alleles, resistant (R) or susceptible (S), and three possible genotypes (SR, RR, SS). The fitness coefficient, which is a measure of survival and reproduction of the different genotypes, was defined as 1 for the susceptible homozygotes SS, $1+s$ (s is the selection coefficient) for resistant homozygotes RR and $1+hs$ (h is the dominance coefficient) for heterozygotes SR. The level of dominance is a measure of the relative position of the phenotype of the heterozygote relative to the phenotype of the two corresponding homozygotes. Complete dominance for a susceptible allele is represented by $h = 0$ and complete dominance for a resistance allele by $h = 1$, alleles are codominant or additive when $h = 0.5$. The fitness coefficients are composite measures of fitness in both the exposed and unexposed mosquitoes groups and are assumed to be the same for males and females.

We also assume a large population, so that genetic drift can be ignored, which enabled us to predict the frequency of the resistant allele at any time t according to the recursion expression:

$$p_{t+1} = \frac{p_t^2(1+s) + p_t q_t(1+hs)}{1 + s(p_t^2 + 2hp_t q_t)} \tag{2.1}$$

Where:

p_t : frequency of the resistant allele at time/generation t
$q_t = 1 - p_t$: frequency of the susceptible allele at time/generation t

This recursive equation is the basic formula of selection of a favourable gene [96, 97]. We defined the allele initial frequency p_0, as the frequency in the first sampling time point, set as generation 0. Each subsequent generation can be converted onto a real timescale of years by assuming a constant number of generations per calendar year.

The ML approach to estimate the unknown parameters h and s and p_o, based on this genetic model, involved selecting initial values of s, h, and p_0 and then testing how well the predicted allele frequencies matched those observed in the dataset.

Field datasets usually consist of the number of resistant alleles x_t and the total number of sampled alleles n at different time points t. The probability of observing x resistant alleles among n alleles follows a binomial distribution, with a probability of success (being a resistance allele) p for each sampled time point t.

$$f(x_t|n_t, p_t) = \binom{n_t}{x_t} p_t^{x_t}(1-p_t)^{n_t - x_t} \tag{2.2}$$

Where:
p_t : probability of sampling an R allele, *i.e*, probability of a success
$\binom{n_t}{x_t}$: combinatorial term to account for the number of ways of sampling x resistant alleles among n total alleles

The corresponding binomial likelihood function is:

$$L(p_t|x_t, n_t) = f(x_t|n_t, p_t) \tag{2.3}$$

The likelihood function returns the likelihood of the value p_t given the observed data of x_t resistant alleles among the sample of n_t at each generation. Essentially, it tells us how consistent the data are with predicted values of p_t (Equation 2.1). The likelihood value for the dataset is the product of the likelihoods across the entire sample:

$$L(p|x, n) = \prod_{i=1}^{t} \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i} \tag{2.4}$$

We implemented this ML methodology in *R* [98] using *constrOptim()* function (from *stats* package) for which it is not necessary to provide analytic derivatives and that can minimize/maximize a function subject to linear inequality constraints. Three constraints on the parameter values were enforced: $0 < p < 1$, $0 < s < 1$, $0 < h < 1$, except when analysing the field data when a constraint on h ($0-1.5$) was imposed. The Nelder-Mead optimisation method algorithm was used, that generates a new test position by extrapolating the behaviour of the objective function measured at each test point arranged as a simplex. The algorithm then chooses to replace one of these test points

with the new test point and the algorithm progresses. The simplest step is to replace the worst point with a point reflected through the centroid of the remaining points. If this point is better than the best current point, then it will expand exponentially along this line. On the other hand, if this new point is not much better than the previous value the simplex returns the previous point. The standard error (s.e) of the estimates was determined by inverting the Hessian matrix evaluated at the ML estimate and the 95% confidence interval endpoints were calculated as $Parameter\ estimate \pm 1.96 * s.e.$

Maximum likelihood estimation is an optimisation technique and there is no guarantee that the set of parameters that uniquely maximises the likelihood will always be found because the algorithm may converge onto local optima whose likelihood is below the global maximum. To overcome this problem 1,000 runs of the ML iteration procedure were performed in every estimation, with random starting values of the parameter estimates used to initialise the optimisation routine [99]. In the analyses described here, the runs that converged to other estimates had ML values sufficiently less than the global maximum that a likelihood ratio test considered them different, so that the set of parameters could be safely discarded. However a small percentage of the runs converged to a set of different parameters with a similar likelihood value that could not be considered different using a likelihood ratio test. The criteria used to exclude these results as potential best estimates was that the estimated value of h lay on the boundary of the constrained parameter range and is expected to reflect erratic behaviour of the algorithm when using a small sample.

We tested the algorithm and program by analysing 100 datasets simulated under 'idealised' conditions using Equations 2.1 and 2.2. Initial frequency, dominance and selection coefficient were in the ranges 0.01-0.04, 0.2 to 0.8 and 0.1 to 0.3 respectively, all distributions were uniform. Three parameter values were selected for each dataset and held constant during the simulation, i.e., there was no temporal or spatial variation in parameter values and population sizes were sufficiently large that stochastic changes in alleles frequencies could be ignored. Data were available for each generation, 100 alleles (50 mosquitoes) were sampled each generation (Equation 2.2) and the simulations were run until the resistance allele frequency exceeded 0.99. Accuracy of analysis was gauged by the correlation coefficient between true and estimated parameter values, and by checking how frequently the true values fell within the estimated 95% confidence intervals.

Next we examined the impact of suboptimal datasets. Equation 2.1 was used to predict allele frequency for 120 generations and we assumed that 100 alleles were sampled in each generation. Two optimal datasets with different dominance values were produced to check if the ML method accurately recovered the parameters when data from all generations was available (as above) and to investigate the effect of different degrees of dominance on estimations. Subsets of the data were used to examine the

influence of incomplete sampling when only a few generations of data are available, or when only the initial stages of spread are available for analysis (Table 2.2 and Figure 2.2).

Field data, collected and analysed by Garcia *et al.* (2009) [95], was available for analysis. There were a total of 78 field collections containing $3,808$ *A. aegypti* (some as much as 2000 km apart). Each mosquito was genotyped at the Ile1,016 locus.

We pooled data from the different locations and analysed it assuming different number of generations of mosquitoes per year (6,9,12,16 and 20), to check the consistency of the estimations. Intuitively, we would expect spatial and temporal variation in the selection parameter in the Garcia *et al.* (2009) [95] dataset and in many other datasets obtained under field conditions. It was therefore vital to ascertain how heterogeneity would affect the algorithm's ability to recover the underlying parameters from pooled data.

Spatial heterogeneity was investigated by simulating allele frequencies for 80 different locations over 50 generations using Equation 2.1; 100 alleles were sampled from each generation (Equation 2.2) and data from each generation in each location were used in the analyses. Parameters p_0 and h were randomly selected from a uniform probability distribution ($p_0 \sim \cup(0,1), h \sim \cup(0,1)$) while s was randomly drawn from a normal distribution ($s \sim \mathcal{N}(0.15, 0.025)$), the constraints on s coefficient are within a reasonable range for a field setting. Once selected for a location, the values of h and s did not change, i.e., there was no temporal heterogeneity.

Two simulation strategies were used: (i) p_0 and h were allowed to vary while s was held constant at 0.1, 0.3, 0.6, 0.8 or 1, (ii) p_0 and s was allowed to vary while h was held constant at 0, 0.25, 0.5. 0.75 or 1. The data across the simulated locations were pooled for analyses. Each simulation strategy was run 300 times giving a total of 300x 5=1500 per strategy. The mean values of each parameter over all simulated locations was assumed to be the true value and the accuracy of the program was, as before, gauged by the correlation coefficient between the estimated and true values, and by the proportion of the true values falling within the 95% CI.

The effect of temporal heterogeneity in estimations was also investigated by varying s and h over 50 generations in a single location, i.e., different s and h values in different generations. The distribution of values were the same as those used for spatial heterogeneity. Three scenarios were considered: (i) s and h both varied over generations, (ii) h could vary while s was held constant, (iii) h could vary while s was held constant. In the simulations of spatial heterogeneity the values of h and s had to be fixed across locations (e.g. h= 0, 0.25, 0.5. 0.75 or 1) but in the simulation of temporal heterogeneity only one location was examined in each simulation so the values could be drawn from the underlying distributions. As before, 300 datasets were produced for each scenario but because the fixed values of h and s could be drawn from a distribution, the total

number of runs was 300x 3=900. As before, the performance of the algorithm under conditions of temporal heterogeneity was assessed by defining the 'true' value as mean over the generations, and calculating the correlation coefficient between the estimated and true values and how frequently the true value was included in the 95% confidence interval.

Finally, it is important to note two features of our analyses that may not be obvious to non-specialists. Firstly, that the genetic parameters h and s describe the overall, net rate of spread of resistance alleles through natural populations and cannot formally distinguish where selection is acting. For example, they cannot determine whether selection was acting differentially on the adult or larval stages, whether fitness costs were associated with resistance, whether there was differential selection on the sexes, nor whether killing was likely to be in early or later adult stages, the latter being a topic of contemporary interest given suggestion by Koella *et al.* (2009) [86] that killing older adults will reduce the selective pressures for insecticide resistance. Secondly, the analyses were designed to recover the genetic parameters that resulted from past control program and, as such, they cannot explore the issue of how differing patterns of insecticide deployment drive resistance. This require a separate, formal modelling approach explicitly designed to investigate the differing impact of deployment strategies on driving resistance. These analyses have been described elsewhere, particularly for the agriculture pesticides [50, 82, 100, 101].

2.3 Results

The analysis of idealised datasets (Table 2.1) suggest ML can accurately recover the underlying parameter values from optimal simulated data.

Table 2.1: Details of 100 idealized simulated datasets.

	p_0	h	s
Parameter range	0.01 - 0.04	0.2-0.8	0.1-0.3
r $*$	0.94	0.99	0.99
TV (%)$*$	91	92	97
[ ]$*$	0.021	0.014	0.002

The simulated data sets were used to check the precision and accuracy of the ML procedure. $*$r correlation coefficient between original value and estimate, TV percentage of true values in the estimates 95% confidence interval and [ ] mean range value of the confidence interval.

Figure 2.2 shows six example simulations of the increase in resistance allele frequencies over 120 generations, under two dominance conditions (semi-recessive, h=0.2 bottom

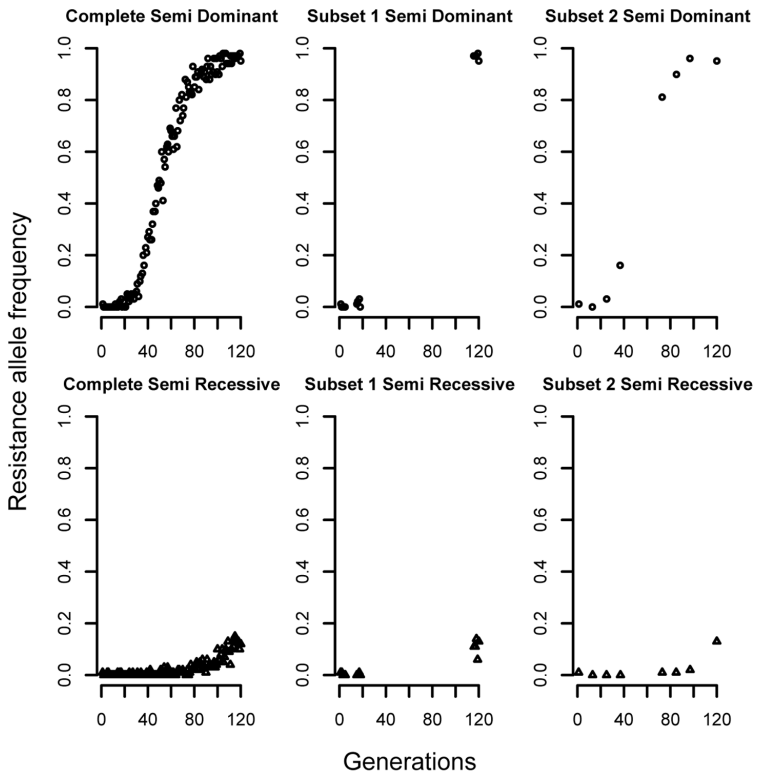

Figure 2.2: Simulated evolution of resistance allele frequency over 120 generations under two different scenarios of dominance relationship and analysing the full dataset or subsets of data. Specifications in Table 2.2.

panel, and semi-dominant, h=0.8 top panel). Values of p_0, s, and h appear in Table 2.2. The program appears less accurate when analysing subsets of the original data, particularly if the resistance allele is semi-recessive. When the resistance allele was semi-dominant, resistance increased rapidly and the true estimates were recovered if the subset included points that captured the pattern of increase, such as the subset 1. When the resistance allele was semi-recessive, the frequency was maintained at low levels for a long period, the true parameters values were either recovered (Subset 1) but within confidence intervals that were so large as to be uninformative, or were not even contained in the confidence intervals (Subset 2) even with the inclusion of the last generation, the only sampling point in the subset that captures the incipient frequency rise. The ML parameter estimates in Table 2.2 were achieved in 34 to 83% of the 1,000 ML runs indicating that a significant proportion of the estimation routines converged onto local maxima.

Analysis of the *A. aegypti* dataset resulted in the parameter estimates in Table 2.3. These ML estimations were obtained assuming 6, 9, 12, 16 and 20 generations per year. The estimation converged on the same ML value 76 to 96% of the runs. With a small percentage of the runs (0.005 to 0.16%) that converged to a set of different parameters with a similar likelihood value but were excluded because the estimated value of h was on the boundary of the constrained parameter range. The estimates of p_0 and h were highly consistent irrespective of assumed number of generations per year and ranged from 0.0032 to 0.0035 and from 0.77 to 0.78, respectively. As expected the s was strongly dependent on the assumed number of generations per year and ranged from 0.042 to 0.15.

Results from spatially heterogeneous datasets pooling data from 80 different locations are shown in Figures 2.3 and 2.4. The algorithm appears unable to consistently obtain accurate estimations of the parameters s and h under such heterogeneous settings, manifested by low values of correlation coefficients and many true values outside the 95% confidence interval of the estimate. For example, with the dominance estimations in Figure 2.3 when selection was constant at 0.6, only 12% of the true values fell within the confidence interval. Initial frequency values were accurately recovered in all simulated scenarios, possibly due to the the recursion dependency on the initial frequency. However, the estimation of selection and dominance coefficients was achieved with very low values of correlation coefficients between the estimates and the mean of the parameter over the 80 simulated locations (not very precise), in all different hypothetical scenarios.

Additionally, if most of the values were in the confidence interval, the mean range of the interval was as wide as the parameter range. For example in Figure 2.4 note that when dominance is constant at 0.75, 100% of the true values are in the confidence interval, but the average mean range is 0.86 (the range is 0-1). The plotted simulated

Table 2.2: Specifications of datasets of Figure 2.2.

		True			Estimates [95% CI]		
Dataset	Generations	p_0	h	s	p_0	h	s
Complete	1:120	0.001	0.2	0.2	0.0010 [0, 0.0018]	0.17 [0.13, 0.38]	0.24 [0, 0.48]
		0.001	0.8	0.2	0.0009 [0.0005, 0.0013]	0.81 [0.77, 0.84]	0.20 [0.19, 0.21]
Subset 1	1:5,15:18,116:120	0.001	0.2	0.2	0.0025 [0, 0.0062]	0.45 [0, 1.5]	0.07 [0, 1]
		0.001	0.8	0.2	0.0016 [-0.0017, 0.0049]	0.77 [0.36, 1.18]	0.19 [0.08, 0.29]
Subset 2	1,13,25,37,73,85,97,120	0.001	0.2	0.2	0.0014 [0, 0.0030]	0.02 [0, 0.18]	1.00 [0.96, 1]
		0.001	0.8	0.2	0.0009 [-0.0007, 0.0025]	0.80 [0.64, 0.96]	0.20 [0.16, 0.25]

Sampled generations and true parameter values and ML parameter estimates with 95% confidence intervals.

Table 2.3: Estimated p_0, h and s from field data.

Parameter	Generations/year	Best value	95% Confidence interval	
p_0	6	0.0032	0.0032	0.0032
	9	0.0033	0.0033	0.0033
	12	0.0034	0.0034	0.0034
	16	0.0034	0.0034	0.0034
	20	0.0035	0.0035	0.0035
h	6	0.77	0.76	0.78
	9	0.77	0.76	0.78
	12	0.77	0.76	0.78
	16	0.78	0.77	0.78
	20	0.78	0.77	0.78
s	6	0.15	0.14	0.16
	9	0.096	0.090	0.101
	12	0.071	0.060	0.081
	16	0.053	0.048	0.057
	20	0.042	0.038	0.046

The data set corresponds to field collected data on Ile1,016 resistance allele frequencies in *A. aegypti* from Mexico. Assuming 6, 9, 12, 16 and 20 generations per year.

data and estimates do not traverse the entire range of the parameters values because they are the mean over the 80 locations, the central limit theorem predicts that these estimates will converge to the center of the distribution.

Simulations of a location with temporal heterogeneous selection pressure (dominance and/or selection changing in every generation) are shown on Figure 2.5. Again, the model does not accurately recover the true parameters under conditions of temporal heterogeneity. The exception was the dominance parameter when it was held constant in a particular location with selection varying in each generation, the correlation coefficient between the estimate and the mean dominance value over the 80 locations was 0.86.

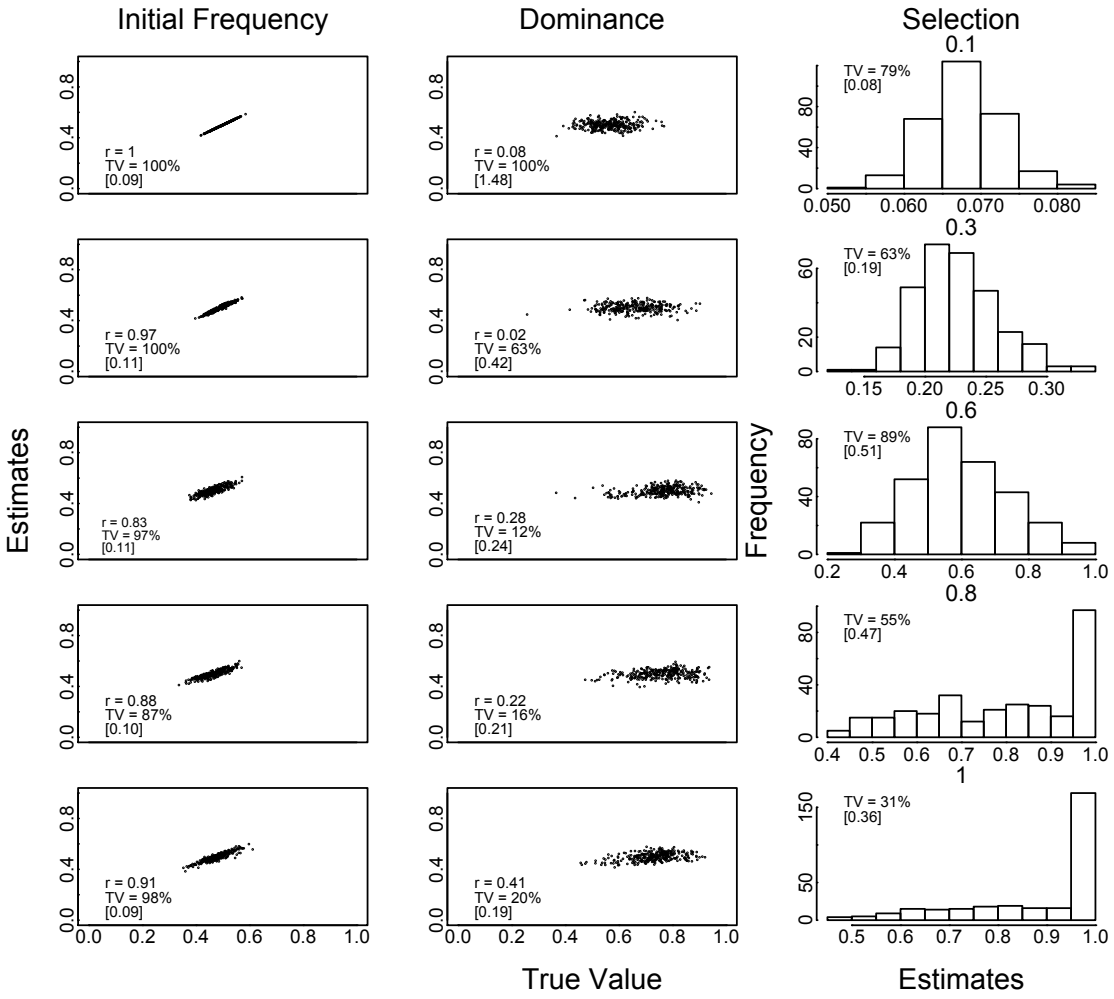

Figure 2.3: Effect of spatial heterogeneity (pooled data from 80 simulated locations) on estimates of initial allele frequency, dominance and selection parameters. The value of the selection coefficient was held constant at 0.1,0.3,0.6,0.8 or 1 in all locations and in every generation, hence there are five rows of results corresponding to each of the 5 values of the selection coefficient. Dominance ($h \sim \cup(0,1)$) varied between simulated locations, but was constant over time within each location. The 'true' value is the mean parameter value over all locations. The Pearson correlation coefficient (r) is between estimated and true values. TV is the percentage of the true values that are included in the 95% confidence interval of the estimate. [ ] is the mean width of the 95% confidence interval in all runs.

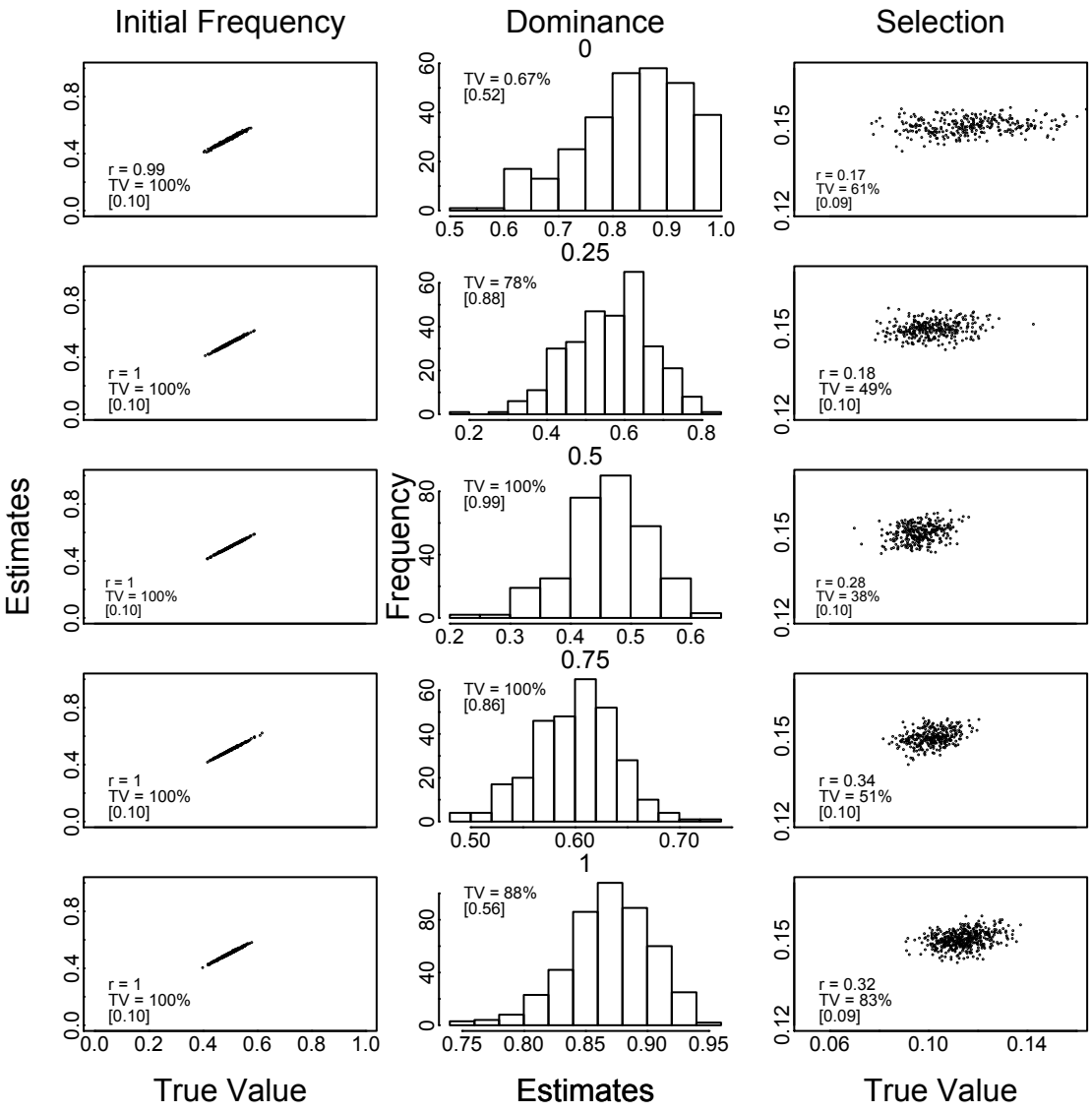

Figure 2.4: Effect of spatial heterogeneity on estimates of initial allele frequency, dominance and selection parameters. The value of the dominance coefficient was held constant at 0,0.25,0.5,0.75 or 1 in all locations and in every generation, hence there are five rows of results corresponding to each of the 5 values of the dominance coefficient. The value of the selection coefficient ($s \sim \mathcal{N}(0.15, 0.025)$) varied between locations, but was held constant over time in each location. The 'true' value is the mean parameter value over all locations. The Pearson correlation coefficient (r) is between estimated and true values. TV is the percentage of the true values that are included in the 95% confidence interval of the estimate. [ ] is the mean width of the 95% confidence interval in all runs.

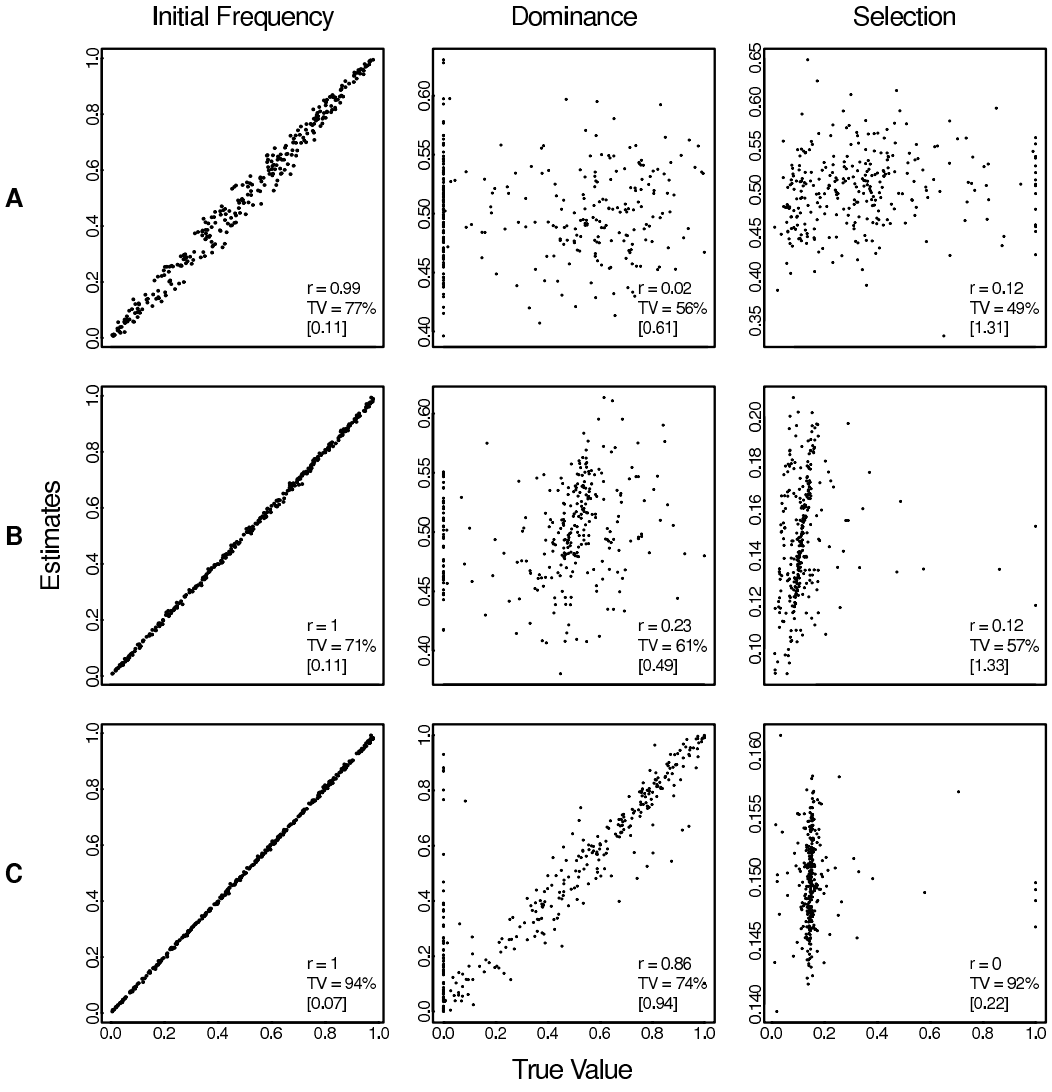

Figure 2.5: Effect of temporal heterogeneity on estimates of initial allele frequency, dominance and selection coefficients parameters. Three different scenarios were simulated: (A) dominance and selection are different in every generation, (B) selection coefficient was held constant in all generations but dominance was allowed to vary, (C) dominance was held constant in all generations while the selection coefficient was allowed to vary. The Pearson correlation coefficient (r) between estimate and true value is shown. TV refers to the percentage of the true values that are included in the 95% confidence interval of the estimate. [ ] is the mean range of the 95% confidence interval in all runs.

2.4 Discussion

Insecticide resistance research is largely focused on the identification of the mechanisms responsible for resistance, and whether the genetic mechanism is monogenic or polygenic, general or population specific and if there are associated fitness costs and developmental patterns [67]. The emergence and spread of resistance is well documented, but there is still a worrying lack of quantification of the evolution dynamics in populations under control [102] and its persistence in populations following cessation of control.

The quantification of the strength of selection acting in the wild has previously been attempted using direct laboratory and field trials, and indirect approaches using a variety of data, including patterns of DNA variability and spatial and temporal changes in allele frequencies [102, 103]. Selection acting on insecticide resistance genes in the field was first estimated using genetic models for species in the genera *Anopheles* by Curtis *et al.* (1978) [104] and Wood and Cook (1983) [105]. Both were based on the observed changes in gene frequency over regular intervals and the latter also discussed estimation by deviations from the expected Hardy-Weinberg equilibrium frequencies. Both methods assumed a fixed level of effective dominance under field conditions.

Another example is the estimation of relative fitness by Livingston and Fackler (2002) [106] for pyrethroid resistance in insects that infest crops. In this case the magnitude of the estimates were similar to those obtained using traditional laboratorial direct approaches using non linear least squares estimation.

The most refined work that we are aware of, quantifies selection coefficients and costs associated with resistance for *Culex pipiens* in southern France, using spatial information from clines to estimate selective advantages and costs, and temporal information from a long-term survey to estimate the selection coefficients of alleles in each environment using a standard ML estimation approach [107, 108].

We have described a ML method for simultaneously estimating the selection and dominance coefficients and an initial resistance allele frequency similar to that of DuMouchel in 1968 [93], but we also tackled the effects of spatial and temporal differences in selection intensity that can arise as a result of different strategies of deployment of the insecticide, migration patterns and/or infrequent and sparse field sampling of mosquitoes.

The approach described in this paper was accurate with simulated data but proved less robust when analysing few intermediate allele frequencies, especially when the resistance allele is recessive. The reason is that all resistance dynamics start from the same point (very low frequency) and end at the same point (very high frequencies) but in the absence of intermediate time points it is impossible to reconstruct the dynamics in between. If the sampling period covers only the onset of resistance or the final stages,

when resistance is close to fixation, the accurate estimation of selection and dominance coefficients can be difficult.

The estimation is problematic because in the early stages heterozygotes prevail in the population, with a fitness $W_{rs} = 1 + hs$, for which there are a range of values of h and s that yield the same product hs. This is illustrated using subsets in Figure 2.2. The true values of s and h were 0.2 and 0.2 but the estimates were 0.45 and 0.07 (Table 2.2).

The situation was even worse for subset 2 of the data (Figure 2.2) where the analysis inferred a completely different trajectory of resistance spread and the true values of $h = s = 0.2$ were estimated as $h = 0.02$ and s=1.0 (Table 2.2). Once again, note that the fitness of the heterozygote was estimated as $1 + hs = 1.02$ which was relatively close to the true value of 1.04 and that it is the predicted value of the homozygotes, which were largely absent from this subset of the data, that were badly estimated (as $W_{rr} = 2.0$ rather than the true value of 1.2). Nevertheless, the calculated fitness (1+hs) is very similar (1.04 and 1.03).

On the other extreme, when resistance is almost fixed, there will be mainly homozygotes in the dataset (with fitness $W_{rr} = 1 + s$), so estimating a dominance value will also be problematic because of the lack of heterozygotes (with fitness $W_{rs} = 1 + hs$).

Unfortunately, this is a very common type of data where genetic surveys initially indicate resistance was absent then, once its presence was detected, a second survey was run and higher levels were detected. Our analyses indicate that it is highly unlikely that any robust genetic parameters can be obtained from these kind of fragmented datasets. Future surveillance surveys should consequently be optimised by choice of a proper sampling strategy and timeframe. It is therefore of extreme importance to sample as many generations as possible, even if it means collecting fewer individuals.

There is an important difference between standard statistics and ML estimation. In standard statistics, the 95% CI should capture the likely variation in magnitude of parameter estimates. In ML it only captures the likely variation provided the model has identified the correct trajectory of allele frequency changes. This is problematic in incomplete datasets where many trajectories may provide similar fits to the observed data. It is absolutely essential to run numerous analyses from randomly selected starting parameter values to check for the presence of numerous trajectories of similar ML but with widely different parameter estimates.

Pooling data from different locations can be seen as a reasonable option to minimise the lack of sampled generations and small sample size. The Ile1,016 mutation frequency dataset of Garcia *et al.* (2009) [95] provided the opportunity to apply the model to real field data. These data contains allele frequencies of mosquitoes collected in 78 different locations around Mexico since 1999. Insecticide use was not uniform across cities and towns in Mexico and will probably differ between years and in addition migration

will probably lead to different initial resistance allele frequency. The estimates obtained from simulated pooled data demonstrated that this kind of data pooling, which is probably inevitable in most surveys, is not very robust. The coefficients reported in Table 2.3 should simply be recognised as a rough estimate between the years 1999 to 2008 and that they may vary, albeit by an unknown amount, over time and space.

Equation 2.1 describes a highly idealised population, i.e., one that is large, randomly mating, and homogenous in time and space. It is therefore important to consider the extent to which our population differs from this paradigm and what consequence this may have for the results.

A large population is required so that we can ignore genetic drift, i.e., random fluctuations in allele frequency around our predicted values. Drift is important in laboratory studies (see [109] for discussion) but in natural populations there is a consensus that genetic drift can be ignored provided $4Ne\tilde{S} > 10$ where Ne is effective population size [110] and $\tilde{\mathrm{S}}$ is the weighted mean fitness of the resistance heterozygotes and homozygotes. Estimates of Ne provided by Gorrochotegui-Escalante *et al.* (2002) [111] for *A. aegypti* ranged from 10-22 in different regions of Mexico. These estimates seem intuitively to be very small. The most likely explanation is that they are measure of historical population size, so may have been caused by founder effects and population bottlenecks in the distant past. Estimates of contemporary population sizes are more appropriate in the current context and most estimates of contemporary effective population sizes of vectors are much higher, for example, in the region of 1,000+ for *Anopheles gambiae* [112–114]. It would be possible to introduce the effects of drift by simulating small populations sizes and sampling (with replacement) the parents of the next generation. However one of the key conclusions of this study is the difficulty of obtaining good quality estimates of genetic parameters from field data, so we prefer to ignore the effects of drift, and simply point out that the stochastic variation introduced by drift will likely further decrease our ability to recover accurate genetic parameter values from field data.

The second requirement, that mating occurs at random is unlikely to be true given the geographical scale of our surveys. It would be relatively straightforward to incorporate this effect by including Wrights F statistics in Equation 2.1. However, there was no evidence of significant departure from Hardy-Weinberg in our dataset (results not shown) so this strategy was not required.

The assumption that the population is homogenous in space is clearly untrue. Pooling of data from different regions was required to increase sample size and frequency because mosquito collections were not uniform at the same location. The simulation results demonstrate the dangers of this approach and work on malaria vectors in Africa show unpredictably high levels of heterogeneities in resistance even across relatively small distances [115]. As mentioned by [93] simple models cannot account for the al-

teration of selection pressure by long term changes in the environment. More complex models that consider geographic clines and the antagonist effect of selection-migration, should be more accurate, but the amount of data necessary make the implementation unlikely in most settings. This model in its simplicity presents a straightforward way to obtain estimates of fitness parameters. The fact that only information about resistant allele frequencies is necessary should make it easier to apply, and yet even such a simple data design is difficult to implement.

Nevertheless the estimated value for p_0 $(0.0032 - 0.0035)$, was in the higher range of 10^{-2} to 10^{-13} expected when a pesticide is first introduced, based on mutation-selection equilibrium [48]. This initial p_0 value reflects the frequency prior to the sampling period. Since 1950, vector control programs in Mexico have used a series of insecticides. DDT was used extensively for indoor house spraying from 1950-1960 and was still used in some locations up until 1998. Malathion was later used for ultra-low volume space spraying of wide areas from 1981 to 1999. In 2000, programs switched to permethrin-based insecticides [116]. The spread of resistance genes in a treated region will depend on the initial resistant allele frequency and it is known that resistance development in pest organisms can occur within 5-100 generations [48]. The relatively high initial frequency estimated explains the immediate, dramatic increase in frequencies of Ile1,016 from the late 1990s to 2006-2008 [95] neglecting genetic drift.

As expected, the strength of selection increased as the number of generations per year decreased, whereas there was less time to get to the same frequency of resistance allele. Selection coefficients ranging from 0.042 to 0.053 (assuming 20 and 16 generations per year) are similar to the selection coefficients of DDT and dieldrin resistant phenotypes in *Anopheles* mosquitoes that have been previously estimated to be on the order of 0.013-0.061 [67]. The values 0.071 and 0.097 (12 and 9 generations per year) are in the range of what was estimated for antimalarial drug resistance: 0.05-0.1 [117], however the value of 0.147 for 6 generations was higher than any previous estimates. This is the first time selection for insecticide resistance has been quantified in this species and should be seen as a preliminary estimate.

The estimated values of h, 0.77 to 0.78, point to partial dominance of the resistance allele under field settings. Alleles conferring knockdown resistance were found to be to be recessive or semi-recessive in their influence in *Anopheles gambiae s.s.* [15], but there is strong evidence for partial dominance or additive effects of Ile1,016 from two laboratory studies of knockdown and survival in strains or families of *A. aegypti* segregating for the Ile1,016 allele. Saavedra-Rodriguez *et al.* (2007) [118] found that 127 of 221 heterozygotes recovered from permethrin knockdown and showed later [119] that when considering overall survival the differences among the three phenotypes appear additive.

Dominance in the field is dependent on the concentration and decay of the insecticide

(see Figure 2.1), under this situation the resistant allele will be effectively dominant and we think that our results of intermediate dominance of Ile1,016 reflect this effect [94]. This interpretation is supported by Roush and Tabashnik (1990) [48], who reported the same situation of partial dominance for cyclodienes and lindane, diazinon, malathion and also for pyrethroids, where 20-60% of the heterozygotes survived exposure in a field setting. There is ongoing debate about differences between laboratory and field settings that extended to the evolution of insecticide resistance itself, some suggesting that resistance in the fields tends to be based on an allele of major effect at a single locus whereas resistance obtained in the laboratory is usually polygenically based [69]. Our results show rapid selection of mutations at a single locus.

The number of generations under natural conditions for this species was estimated at 20 or more among strains in field conditions in Brazil, this leads us to consider the results with the highest number of generations as the most likely, but because *A. aegypti* eggs can survive desiccation for months and hatch once submerged in water [120], the number of generations is variable. Nevertheless, the predicted resistance frequency trajectory using equation 2.1 and the different estimates obtained assuming different generations per year will be approximately the same in a timescale of 20 years.

Most mutations encoding insecticide resistance are expected to incur a fitness penalty, compared to unmutated genes, in the absence of insecticide. There is some field evidence of reduced fitness of Ile1,016 mutations in *A. aegypti* in permethrin free environments [95] which leads us to make two technical points. Firstly, that the selection and dominance coefficients reported here are overall values that combine the mutation's benefit when encountering insecticide and any fitness effect in insecticide-free areas. Secondly, the method can equally be applied to measure negative selection pressures, (i.e., when a mutation is being lost from a population) from field data on the mutation after insecticide is withdrawn.

Two factors are of particular relevance to field work. Firstly, that surveillance needs to be continuous so that a full dataset covering the whole period of resistance spread becomes available upon which to base these estimates. This may mean monitoring sentinel sites for long periods when resistance is rare or absent, but a continuous dataset is a prerequisite for accurately estimating the dynamics underlying the spread of resistance. Note that a continuous dataset does not necessarily mean collecting samples every generation. The reason the analysis could fail to recover the true parameters (Table 2.2) was because of large gaps in the survey: simulations of semi-dominant mutations lacked samples from periods of intermediate frequency, while simulations of semi-recessive mutations only contained data from the early stages (Figure 2.2). Operationally, this suggests that regular, rather than intensive but periodic, sampling is the best strategy. As an example, we re-analysed the semi-dominant dataset but just incorporated samples every 10 generation, i.e., at generations 1, 10, 20, 30:120.

This resulted in estimates of $p_o = 0.0006$ (95% CI: 0.0001-0.0010), h=0.89 (95% CI: 0.84-0.95), s= 0.20 (95% CI: 0.18-0.21) which are similar to those obtained using data from all generations (Table 2.2, the dominance coefficient is higher but the confidence intervals overlap).

The second point is that dominance levels acting in the field may be much higher than those observed in the laboratory. The most plausible explanation is that mosquitoes in the wild are encountering low levels of insecticide that are insufficient to kill heterozygotes. Increasing dominance greatly increases the rate at which resistance develops. This suggests that insecticide applications should be enforced in such a way that ensure high coverage with high doses. Our results suggest that the doses being applied may be inadequate and that pursuing the current deployment settings will lead to the rapid increase of resistant mosquitoes and eventually to the complete inefficiency of permethrin in the combat of dengue in Mexico.

Chapter 3

The importance of modelling the spread of insecticide resistance in a heterogeneous environment: the example of adding synergists to bed nets

Abstract

Insecticides are an effective and practical tool for reducing malaria transmission but the development of resistance to the insecticides can potentially compromise controls efforts. A new generation of long-lasting insecticidal bednets is being developed that incorporates a chemical synergist on the roof panel of the net. We use mathematical modelling to explore the contribution of such nets in delaying insecticide resistance while determining the important parameters in driving resistance in an heterogeneous environment, i.e., an environment in which insecticides can be encountered in different ways. A genetic model is developed to predict changes in mosquito fitness and resistance allele frequency. Parameters describing insecticide selection, fitness cost and the additional use of synergist were incorporated. We performed uncertainty and sensitivity analysis followed by investigating the evolution of resistance under scenarios of fully effective or ineffective synergists. The spread of resistance was most sensitive to selection coefficients, fitness cost and dominance coefficients while mean fitness was most affected by baseline fitness levels. Using a synergist delayed the spread of resistance but could, in specific circumstances, actually increase the rate of spread. We observed different spread dynamics, with simulations leading to fixation, loss and most interestingly, equilibrium (without explicit overdominance) of the resistance allele. This strategy has the potential to delay the spread of resistance but note that in an heterogeneous environment it can also lead to the opposite effect, i.e., increasing the rate of spread. This clearly emphasises that selection pressure acting inside the house cannot be treated in isolation but must be placed in context of overall insecticide use in an heterogeneous environment.

3.1 Introduction

Malaria is one of the most important parasitic infection in humans. Several initiatives from the international health community in the past decade have lead to an estimated drop in malaria associated mortality from around 1 million in 2000 to about 655,000 in

2010 according to the world health organization [34], although an independent recent study reported the decrease to be from 1.82 million in 2004 to 1.24 million in 2010 [121].

Among the current recommended interventions to control the disease is the use of insecticidal nets or indoor residual spraying with insecticide to control vector mosquitoes populations [34]. A major issue arising from the intense deployment of insecticides is the development of resistance to the chemical agents [42]. It is a ubiquitous problem, regarded as a major hindrance in the control of malaria. Furthermore, the use of insecticides is not restricted to public health, in fact, around 90% of all insecticide is deployed in agriculture [2]. This potential spatial heterogeneity of insecticide deployment can give rise to a mixed environment for mosquito populations.

In the past mathematical models have been used to inform resistance management practices [52], determining the impact of different mosquito control intervention strategies including the protection conferred by bed nets [122], and, recently, to develop new approaches such as the idea of evolution-proof insecticides [85–87]. However, few [123] have considered the spread of resistance in a variable selection pressure context. Consequently a model is presented here that considers different niches in an environment, that can offer some insights on the importance of different parameters and their interactions in the dynamics of insecticide resistance.

This approach is particularly suitable for investigating the impact of a specific long-lasting insecticidal net that is being developed. Vestergaard Frandsen has submitted a bed net prototype (PermaNet 3.0) for formal evaluation to the WHO Pesticide Evaluation Scheme that incorporates insecticide plus synergist.

Synergists are natural or synthetic chemicals, which increase the lethality and effectiveness of currently available insecticides, but that are nontoxic to insects on their own. They block the metabolic systems that would otherwise break down insecticide molecules, helping to restore chemical susceptibility that would require higher levels of the insecticide [124]. For this reason they are proposed for use in overcoming metabolic resistance and also to delay the manifestation and/or spread of resistance [125].

ln this bed net, synergist (Piperonyl butoxide - PBO) together with the pyrethroid deltamethrin are incorporated into the fibres on the roof panel of the net, while incorporating only deltamethrin on a lower dosage in the side panels. The rationale behind this approach approximates the "two-in-one" concept for bed nets. Treating different parts of bed nets with different insecticides (e.g. combining pyrethroid insecticide, applied to the side panels of the bed net, together with carbamate insecticide on the roof [126]) has been suggested to confer advantages over the use of insecticides alone [127]. The assumption is that foraging female mosquitoes explore an occupied bed net from the top downwards (as the warm air and carbon dioxide that emanate from the sleeper move upwards), i.e., will land on the roof first and make their way down the side panels.

Restricting the synergist to the roof also allows the sides of the net to be made of a softer and more comfortable fiber for the user [128].

Here we developed a general and flexible model by expanding the usual genetic models to account for spatial and sexual heterogeneities in insecticides exposures. Statistical tools as partial rank correlation coefficients, logistic regression and classification trees were used to explore specific situations of synergist application and to uncover the dynamics of resistance.

3.2 Methods

Model

A population genetic model was designed that predicts changes in mean fitness and resistant allele frequency as outcome variables to explore the relative contribution of each different environmental niche to the dynamics of the population insecticide resistance status. Mean fitness assesses the potential effectiveness of control strategies at decreasing the population while change of allele frequency between generations quantifies selection pressure for resistance.

The model is deterministic, i.e., based on the approximation of an infinitely large population size so that stochastic fluctuations of allele frequencies can be neglected. Investigations of the changes in allele frequency caused by natural selection are based upon the assumption that selection operates through differential survival of the zygote from birth to maturity. It assumes that random mating occurs among all adults pooled across all niches, and that progeny are then randomly distributed among the niches. Resistance is determined by one allele at one locus [48, 63, 64](S: insecticide susceptible allele; R: insecticide resistance allele).

Table 3.1 defines the fitness of each genotype for each different niche. It also defines the proportions exposed to each niche, that sum to 1, which implies that a mosquito can only encounter a single niche in a generation.

Four niches were considered:
1- Insecticide free (n): it can be an area either inside or outside a household;
2- Non public-health related insecticide deployment (o): typically insecticide use in agriculture and households. These are deployed outwith public health mosquito control campaigns, and generally out of the control of public health officials; The subscript 'o' is used for brevity, noting that casual use inside the house, e.g. mosquito coils, would also be included in this class;
3- Insecticide-treated bed nets (ITN);
4- Insecticide-treated bed nets with synergist on top of the net (ITN + Synergist);

Table 3.1: Model structure: niches, exposure and genotype fitnesses within each niche. See Table 3.2 for parameters.

	Niches			
	Insecticide free	Non public	ITN	ITN + Synergist
Exposure males	$1-(\alpha_{mo}+\alpha_{mi})$	α_{mo}	$\alpha_{mi}(1-\beta_m)$	$\alpha_{mi}\beta_m$
Exposure females	$1-(\alpha_{fo}+\alpha_{fi})$	α_{fo}	$\alpha_{fi}(1-\beta_f)$	$\alpha_{fi}\beta_f$
Fitness SS	1	$1-\varphi_o$	$1-\varphi_i$	$(1-\varphi_i)k$
Fitness RS	$1-h_n z$	$(1-\varphi_o)+h_o s_o$	$(1-\varphi_i)+h_I s_I$	$[(1-\varphi_i)+h_i s_i]k$
Fitness RR	$1-z$	$(1-\varphi_o)+s_o$	$(1-\varphi_i)+s_i$	$[(1-\varphi_i)+s_i]k$

There is likely to be differential exposure to insecticide and hence different selection pressure in the sexes, since only females feed on humans and are, therefore, the ones most likely to enter human habitations and encounter insecticides. Consequently, the proportion of mosquitoes (α) that encounter each niche was differentiated by sexes and genotype fitness was calculated separately. Fitness of the genotype SS in the insecticide free niche was considered as the reference fitness level, and other genotypic fitnesses were measured relative to this fitness, which was taken to be 1.

Different selection (s) and dominance (h) parameters were defined for each niche, except for the insecticide free. There is by definition no insecticide exposure in the insecticide free niche, so s is replaced by z (the cost of carrying a resistance allele). Dominance is not an intrinsic property of the alleles, it depends on the environment in which they are expressed, thus the differences in dominance coefficients between niches. High levels of insecticide may render the resistance allele recessive, because only homozygotes survive, while low levels may allow survival of both heterozygotes and resistant homozygotes rendering the allele dominant [129, 130].

This is a highly flexible genetic model, that includes a baseline fitness level φ for niches where insecticide is deployed, that captures the variable effects on fitness of being fully susceptible to insecticides [131]. For example, setting $\varphi_o = \varphi_i = 1$ means SS genotypes are always killed when contacting insecticides, while setting $\varphi_o = 0.9$ means 10% of SS will survive exposure in the non-public niche. It also allows the fitness of a resistance homozygote meeting a ITN to be less than 1 and therefore smaller than a susceptible homozygote in an insecticide free niche, reflecting the fact that a fully resistance genotype may not be completely impervious to the insecticide. For example setting $\varphi_o = 0.9, h_o = 0.2, s_o = 0.6$ means that 10% of SS genotypes, 22% of RS and 70% of RR survive exposure in the non public-health related insecticide deployment niche.

Two parameters were included that relate to the additional use of synergist: k that

quantifies synergist efficiency and β the proportion of mosquitos meeting both insecticide and synergist in the bed net. It is assumed that synergist exposure is equally efficient across genotypes. For example, if the probability of surviving bed net contact for SS, RS and RR genotypes is 10%, 22%, 70% (see above) and $k = 0.1$, the proportion surviving bed net plus synergist falls to 1%, 22% and 7%, respectively. It would be straightforward to include separate k values for each genotype if the synergist impact differed between genotypes. Description of all parameters on Table 3.2.

Based on the model described on Table 3.1 the fitness (W with appropriate subscripts) across all niches of each genotype will be [132]:
Males,

$$\begin{aligned} W_{m,ss} =& 1 - (\alpha_{mo} + \alpha_{mi}) + \alpha_{mo}\,(1 - \varphi_o) + \alpha_{mi}\,(1 - \beta_m)\,(1 - \varphi_i) \\ &+ \alpha_{mi}\,\beta_m\,(1 - \varphi_i)\,k \end{aligned} \tag{3.1}$$

$$\begin{aligned} W_{m,rs} =& [1 - (\alpha_{mo} + \alpha_{mi})]\,(1 - h_n\,z) + \alpha_{mo}[(1 - \varphi_o) + h_o\,s_o] + \alpha_{mi}\,(1 - \beta_m)\times \\ &\times [(1 - \varphi_i) + h_i\,s_i] + \alpha_{mi}\,\beta_m\ [\,(1 - \varphi_i) + h_i\,s_i\,]\,k \end{aligned} \tag{3.2}$$

$$\begin{aligned} W_{m,rr} =& [1 - (\alpha_{mo} + \alpha_{mi})]\,(1 - z) + \alpha_{mo}[(1 - \varphi_o) + s_o] + \alpha_{mi}\,(1 - \beta_m)\times \\ &\times [(1 - \varphi_i) + s_i] + \alpha_{mi}\,\beta_m\ [\,(1 - \varphi_i) + s_i\,]\,k \end{aligned} \tag{3.3}$$

Females,

$$\begin{aligned} W_{f,ss} =& 1 - (\alpha_{fo} + \alpha_{fi}) + \alpha_{fo}\,(1 - \varphi_o) + \alpha_{fi}\,(1 - \beta_f)\,(1 - \varphi_i) \\ &+ \alpha_{fi}\,\beta_f\,(1 - \varphi_i)\,k \end{aligned} \tag{3.4}$$

$$\begin{aligned} W_{f,rs} =& [1 - (\alpha_{fo} + \alpha_{fi})]\,(1 - h_n\,z) + \alpha_{fo}[(1 - \varphi_o) + h_o\,s_o] + \alpha_{fi}\,(1 - \beta_f)\times \\ &\times [(1 - \varphi_i) + h_i\,s_i] + \alpha_{fi}\,\beta_f\ [\,(1 - \varphi_i) + h_i\,s_i\,]\,k \end{aligned} \tag{3.5}$$

$$\begin{aligned} W_{f,rr} =& [1 - (\alpha_{fo} + \alpha_{fi})]\,(1 - z) + \alpha_{fo}[(1 - \varphi_o) + s_o] + \alpha_{fi}\,(1 - \beta_f)\times \\ &\times [(1 - \varphi_i) + s_i] + \alpha_{fi}\,\beta_f\ [\,(1 - \varphi_i) + s_i\,]\,k \end{aligned} \tag{3.6}$$

If resistance allele frequency is $p_m\ /p_f$ and frequency of susceptible allele is $q_m\ /q_f$, after selection the genotypic frequencies will be:

$$
\begin{aligned}
RR_m &= \frac{W_{m,rr}\ p_m\ p_f}{\bar{W}_m} \\
RS_m &= \frac{W_{m,rs}\ (p_m\ q_f\ +\ p_f\ q_m)}{\bar{W}_m} \\
SS_m &= \frac{W_{m,ss}\ q_m\ q_f}{\bar{W}_m} \\
RR_f &= \frac{W_{f,rr}\ p_m\ p_f}{\bar{W}_f} \\
RS_f &= \frac{W_{f,rs}\ (p_m\ q_f\ +\ p_f\ q_m)}{\bar{W}_f} \\
SS_f &= \frac{W_{f,ss}\ q_m\ q_f}{\bar{W}_f}
\end{aligned}
\tag{3.7}
$$

Where $\bar{W}$ are the mean fitness, given as the sum of the numerators:

$$
\bar{W}_m = W_{m,rr}\ p_m\ p_f + W_{m,rs}\ (p_m\ q_f\ +\ p_f\ q_m) + W_{m,ss}\ q_m\ q_f \tag{3.8}
$$

$$
\bar{W}_f = W_{f,rr}\ p_m\ p_f + W_{f,rs}\ (p_m\ q_f\ +\ p_f\ q_m) + W_{f,ss}\ q_m\ q_f \tag{3.9}
$$

The frequency of the resistance allele in males after selection, i.e., in the mating pool for the next generation $(t+1)$, is

$$
p_{m,t+1} = \frac{W_{m,rr}\ p_m\ p_f + 0.5\ W_{m,rs}\ (p_m\ q_f + p_f\ q_m)}{\bar{W}_m} \tag{3.10}
$$

and the corresponding frequency in females following selection is

$$
p_{f,t+1} = \frac{W_{f,rr}\ p_m\ p_f + 0.5\ W_{f,rs}\ (p_m\ q_f + p_f\ q_m)}{\bar{W}_f} \tag{3.11}
$$

Under this model, the ratio of change of the gene frequency per generation is given by

$$
\Delta\ p_m = \frac{p_{m,t+1}}{p_{m,t}} \tag{3.12}
$$

$$
\Delta\ p_f = \frac{p_{f,t+1}}{p_{f,t}} \tag{3.13}
$$

All simulations started assuming Hardy-Weinberg equilibrium (HWE), but genotypes will move away from HWE due to differential selection on the sexes. This reflects their different degrees of exposure to different environments so that the resistance allele frequency will diverge slightly in the breeding individuals of each sex and their progeny genotypes will no longer be in HWE. Consequently, to allow for redistribution of resistance between the genotypes the chosen census point was at generations 10-11, based solely on intuition.

Parameter values

The subjective part of the analysis lies in identifying plausible values and distributions for the parameters in Table 3.2.

Initial resistance allele frequency value, p_0=0.001, was used in all calculations and was selected to reflect the initial stages of insecticide resistance (where most of the individual are expected to be heterozygotes). There is little field information available for the parameter values appropriate for *Anopheles gambiae* species complex and parameter values vary depending on the species and the local environment. The range of values and distributions chosen were very broad to investigate general properties of the system; narrow distributions can be used to investigate specific situations.

The proportion of mosquitoes subject to a particular niche (α) were randomly selected from a uniform distribution but subjected to the constraint that the sum over all niches by sex is 1; values were randomly selected from the uniform distribution and then divided by the overall sum. The proportion of males that meet the ITN, α_{mi} was constrained to always be smaller than the proportion of females α_{fi} and smaller than 0.2 (less than 20% of the males of the population enter the household and contact the bed net) to reflect the belief that only a small proportion of males enter a household since they do not seek to blood feed on humans. The proportion of males that is expected to contact the top of the bed net, and be exposed to both insecticide and synergist is assumed to be very small, so we restricted the maximum value of β_m to 0.2.

Uncertainty and sensitivity analysis

Simulations to understand the influence of each parameter on the outcome variables (mean fitness and change in resistance allele frequency) were performed using latin hypercube sampling (LHS) to generate a data set and partial rank correlation coefficients (PRCC) calculated to provide a quantitative measure of the impact of each parameter [133]. LHS techniques were first developed to explore the behavior of complex models in economics, engineering, chemistry and physics and have been used in models predicting the impact of insecticide-treated nets on malaria transmission [134].

The analysis was performed using R software [98] and implementation of LHS using

Table 3.2: Parameters, symbols and subscripts used in the construction of the model.

Parameter	Symbol
Dominance coefficient in each niche	h
Fitness cost of carrying a resistance allele	z
Selection coefficient in each niche	s
Baseline fitness level of susceptible homozygote in niches were insecticide is deployed	φ
Proportion of mosquitoes encountering a particular niche	α
Proportion of mosquitoes encountering ITN that also encounter the synergist	β
Impact of synergist	k
k=0; synergist completely effective: all mosquitoes encountering insecticide plus synergist die k=1; synergist completely ineffective: mosquitoes encountering insecticide plus synergist die at the same rate as those encountering insecticide alone (the model tracks survival in different niches -Table 3.1, so the impact is on a reverse scale)	
Subscripts	
Male	m
Female	f
Insecticide free	n
Deployment of insecticide outside the house	o
Deployment of insecticide inside the house	i

Table 3.3: Parameters range of values used in simulations.

Parameter	Range of values			Distribution
	Minimum	Peak	Maximum	
$p_m = p_f$		0.001		Constant
h_n	0	0.5	1	Triangular
h_o	0	0.5	1	Triangular
h_I	0	0.5	1	Triangular
z	0	0.5	1	Triangular
s_o	0	0.5	1	Triangular
s_I	0	0.5	1	Triangular
φ_o	0		1	Uniform
φ_I	0		1	Uniform
β_f	0		1	Uniform
β_m	0		0.2	Uniform
α_f*	0	0.5	1	Triangular
α_m*	0	0.5	1	Triangular
α_{mi}	0		0.2	Uniform
k	0		1	Uniform

$*$ Females: all female niches; Males: all male niches expect ITN/ITN+synergist.

package *lhs*. It does not allow for the specification of each variable distribution beforehand, so sampling was performed assuming a uniform distribution. Once the sample was generated, the uniform sample from a column (variable) could be transformed to the required distribution (Table 3.3) by using quantile functions (using the *qtriangle* command in R).

A data set of 3,000 replications was generated, with random parameters and the corresponding values of the outcome variables using equations 3.8, 3.9, 3.12 and 3.13. Ten replicates of this procedure were performed as suggested in [133] to investigate the predictive precision of model using LHS as the sampling method. This was achieved by analysing each replicate separately and verifying that results were consistent across ten replicates.

Allele frequency ratio under two extreme scenarios of synergist effectiveness

Following uncertainty and sensitivity analysis the evolution of the frequency of the resistance allele was investigated. This was achieved by simulating a scenario with a fully

effective synergist ($k = 0$), and the other extreme, a scenario where encountering the synergist had absolutely no effect ($k = 1$). The dataset consisted of 3,000 individual simulations that were run drawing values from Table 3.3 for the parameters. Each simulation was run twice, once with $k = 0$ and another one with $k = 1$. The ratio between the resistance allele frequency in the population in both scenarios at generation 10 ($y = \frac{p_{10}|k=0}{p_{10}|k=1}$) indicates how fully effective synergists increase ($y > 1$) or decrease ($y < 1$) the spread of resistance.

Results included a counter-intuitive outcome that the inclusion of a synergist could lead to an increase in the rate of the spread of resistance (i.e. $y > 1$). Further investigation of this result was pursued by performing a logistic regression with a binary dependent variable (1 if $y > 1$ and 0 if $y < 1$), therefore quantifying how changes in the parameters values affect the odds of getting the unexpected outcome $y > 1$. In this regression only 14 parameters out of the 16 could be included. The parameters α were excluded since they are codependent. They must sum to unity, so α_{mo}, α_{fo} were excluded from the regression since they achieve the smaller PRCC values (see later).

Classification trees

The model has a substantial number of parameters, 16, so the logistic regression becomes inefficient when considering all possible interactions between them. An alternative approach to logistic regression is classification trees, that sub-divide the parameter space into smaller regions, where the interactions are more manageable.

Classification trees are used to predict membership of cases in the classes of a categorical dependent variable (1 if $y > 1$ or 0 if $y < 1$) from their input parameters and were implemented using an algorithm that grows a binary tree [135]. At each internal node in the tree, a test is applied to the input parameters to identify the binary distinction which gives the most information about the class membership. The process is repeated at each resulting node, continuing the recursion until some stopping criterion is reached where it makes a prediction [135]. The threshold of complexity parameter (cp) was one of the stopping criteria used here, it ensures that any split that does not decrease the overall lack of fit by a factor of cp is not attempted; it can be preset or estimated using cross-validation. Here it was used cross-validation, which is a method for validating a procedure for model building, without an independent validation dataset. It includes any given random divisions of the data into 90% learning and 10% test sets [136]. The optimally sized tree was obtained by running 10-fold cross-validations on the data and by including another stopping criterion, a minimum of 50 observations in a node in order for a split to be attempted.

3.3 Results

Uncertainty and sensitivity analysis

LHS was used to generate a dataset for sensitivity analysis. The procedure was first replicated 10 times so that the model predictive precision could be assessed. The standard errors (se) and coefficient of variation (cv) of the outcome variables between replications are small (se: 0.001-0.3; cv: 0-0.005), suggesting that the predictive precision of the model does not depend on the LHS generated dataset. Statistical evidence (t-test, p-value < 0.05 in all replications) indicates a difference between sexes on both fitness and rate of change of resistance.

Parameter sensitivity was performed to quantify how a change in an input parameter value causes a change in the outcome variables. Partial rank correlation coefficients calculated between each of the input parameters and the outcome variables are shown in Figure 3.1, in black circles in all panels. This analysis allows to assess the relative importance of the parameters in driving resistance and how it affects fitness, especially the magnitude of the correlation with the synergist, k.

The rate of spread of resistance (Figure 3.1, A and B) is most sensitive to parameter values of selection and dominance coefficients and fitness cost (s, h, z). The correlation is negative in niches where insecticide is not employed and positive when it is present. In males (A), dominance and selection coefficients inside the house (h_i, s_i) have little effect on the ratio of change, presumably because only a small fraction is exposed.

The negative correlation between mean fitness and baseline fitness levels penalties (φ_o and φ_i) is the strongest of all in both sexes. Male (C) and female (D) mean fitness are also sensitive to the parameters α: male PRCC coefficients are positive with mean fitness in all niches and females PRCC coefficients are positive in the insecticide free niche and negative in the other two.

Overall, changes in parameter k do not appear to have a big impact in the mean fitness of the population. In females the parameter k is positively correlated but small in magnitude and β_f (the proportion that meet both the synergist and insecticide) shows also only a small negative correlation. Both k and β_f show no correlation with mean fitness in the male population.

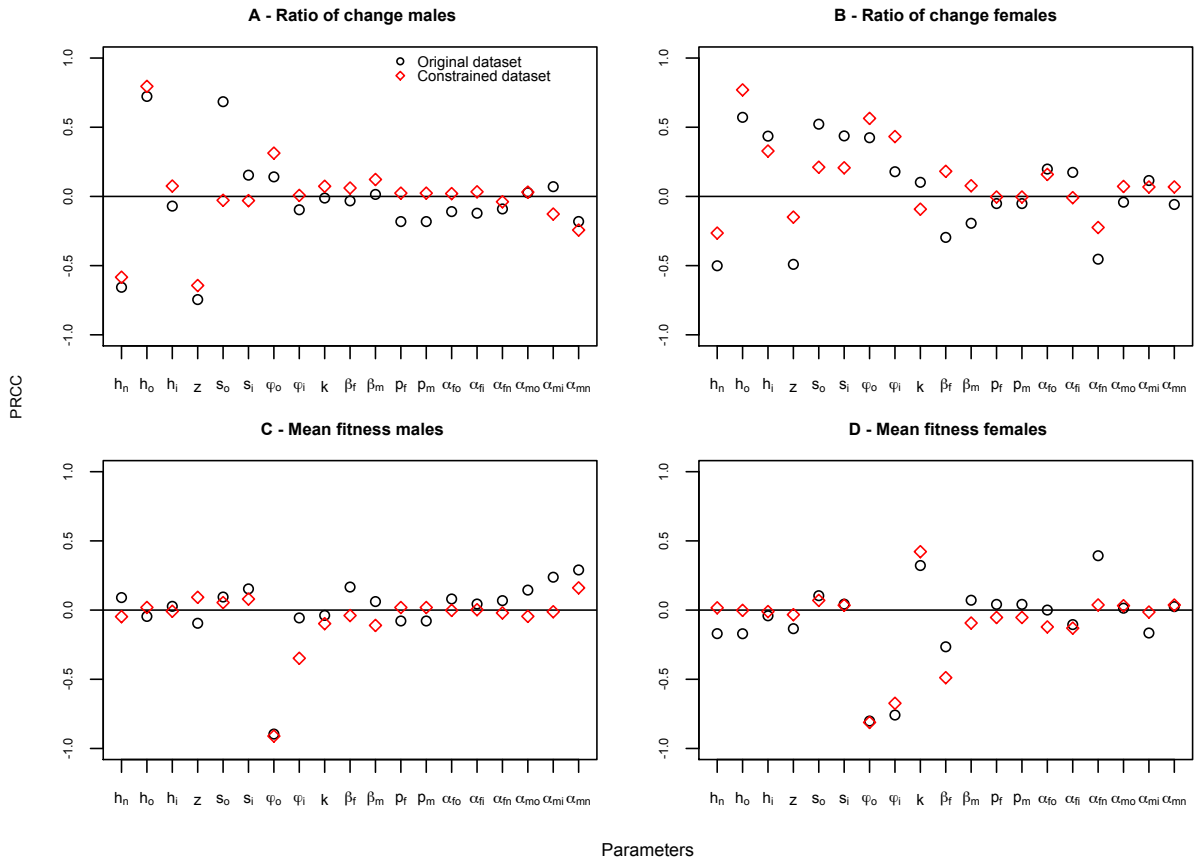

Figure 3.1: Plot of Partial Rank Correlation Coefficients between each of the parameters in the model and the two outcome variables: mean fitness and ratio of change of resistance allele frequency, in both sexes (above zero increasing positive correlation, below zero increasing negative correlation). The black circles refer to the coefficients calculated using the original dataset and the red diamonds coefficients were calculated using a dataset generated with the constraints: $\alpha_{fn} < 0.2$, $s_o > 0.5$, $\varphi_i < 0.8$, $s_i < 0.5$, derived from the classification trees analysis. The parameter symbols in the x-axis are defined in Table 3.2.

Allele frequency ratio under two extreme scenarios of synergist effectiveness

Figure 3.2 shows the rate change of allele frequency comparing the extremes cases of a fully effective ($k = 0$) and of a inefficient ($k = 1$) synergist. In most cases (90%) y is smaller than 1, which is intuitively the most likely outcome, i.e., resistance spreads slower in the presence of the synergist. The effect of the synergist on males and females is not strictly comparable but is overall similar. Most importantly, Figure 3.2 shows that there is little difference between the two scenarios, most of the ratios are between 0.8 and 1, implying that the delay in the spread of resistance caused by the synergist is not very large. Nevertheless, what was unexpected was that in approximately 10% of the cases the rate of allele spread can be higher when the synergist is fully effective ($y > 1$). Figure 3.3 shows the predicted frequency of the resistance allele under different values of k (ranging from 0 to 1) in a scenario which $y > 1$ to illustrate the difference in the spread of resistance. As an example, at generation 70 the predicted frequency when the synergist is inefficient ($k = 1$) is 0.11 and when is fully effective ($k = 0$) is 0.26.

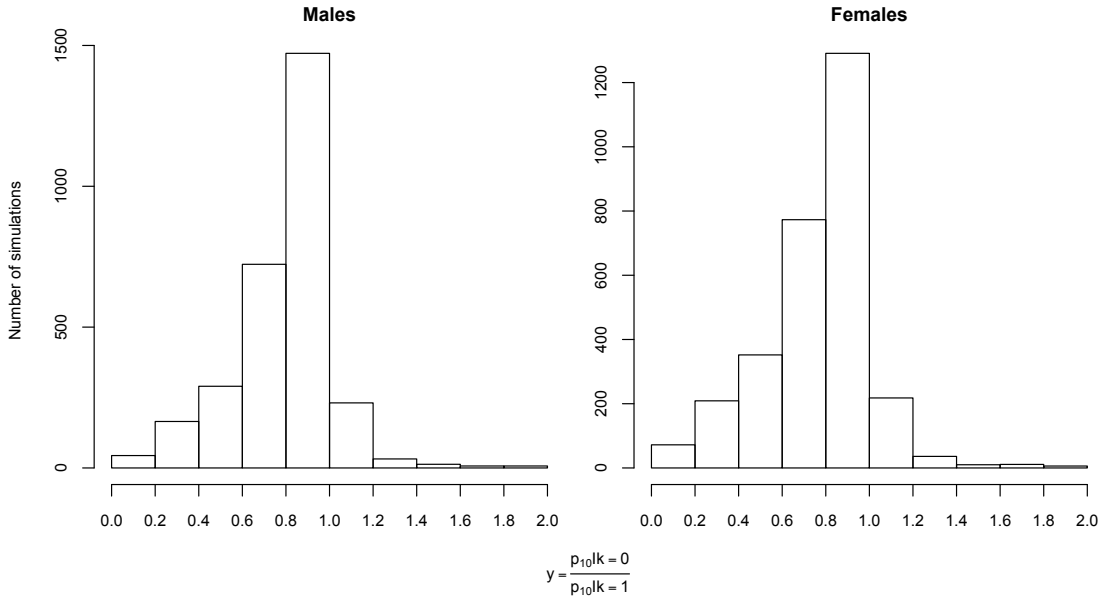

Figure 3.2: Histogram of the ratio of resistance allele frequency at generation 10 in the extremes cases of a fully effective ($k = 0$) compared to an inefficient ($k = 1$) synergist: $y = \frac{p_{10}|k=0}{p_{10}|k=1}$). Only values of $y < 2$ shown, which constitute 99.3% of the number of simulations. Values higher than 1 ($y > 1$) indicate the counter-intuitive result, i.e., that the synergist presence drives resistance faster.

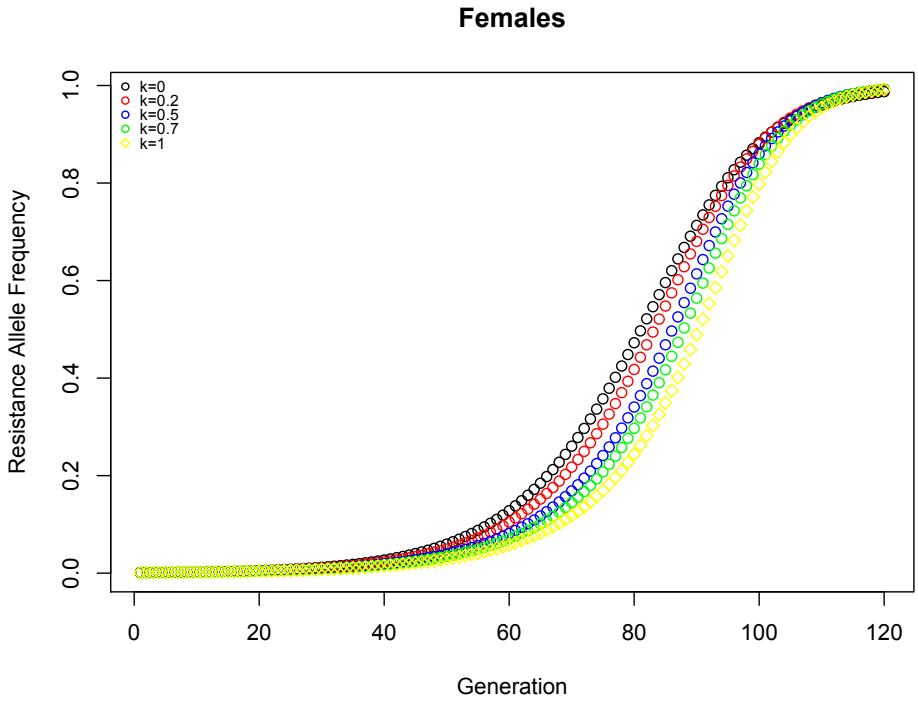

Figure 3.3: How synergists affect the spread of resistance: Predicted resistance allele frequency in females for different values of synergist impact ($k = 0, 0.2, 0.5, 0.7, 1$) in a specific setting ($h_n = 0.37, h_o = 0.50, h_i = 0.07, z = 0.87, s_o = 0.47, s_i = 0.46, \varphi_o = 0.11, \varphi_i = 0.33, \beta_m = 0.02, \beta_f = 0.82, \alpha_{fo} = 0.41, \alpha_{fi} = 0.50, \alpha_{fn} = 0.10, \alpha_{mo} = 0.38, \alpha_{mi} = 0.34, \alpha_{mn} = 0.28$) that scored $y > 1$.

Logistic regression

Results of the logistic regression are presented in Table 3.4. Each one unit of increase in the parameter value in question will increase/decrease ($+/-$ signal of the estimate) the log odds of the unexpected event ($y > 1$), the odds are the exponentiated values of the estimates, shown also in Table 3.4. The parameters β_f, β_m and α_{mi} are not significant (p-values > 0.05) and appear to have no impact on the outcome. An increase in the parameters s_o, h_o, and φ_o increases considerably the odds off the counter-intuitive result ($y > 1$) and a increase of α_{fn}, h_i, s_i, φ_i, α_{fi}, h_n and z decreases the odds. Parameters related to the niche where insecticide is not deployed (n) have little impact on the counter-intuitive result, which appears to be governed mainly by the values of the parameters in the niche where insecticide is being applied for other reasons outside the house (o) and the niche insecticide inside the house (i).

Additionally, to compare these two niches, a series of Wilcoxon signed-rank tests were performed. From the results of the tests it was determined that all the parameters in the niche where insecticide was encountered outside the house (α_{mo}, α_{fo}, φ_o, s_o and h_o) were significantly higher than the equivalent parameters in the niche inside the house in the simulations that led to the unexpected outcome.

Table 3.4: Logistic regression: parameter coefficient estimates, standard error, p-value produced by the Wald test (to check that the parameter coefficients are different from zero) and the odds ratio ($OR = e^{Estimates}$) associated with each parameter.

	Estimate	Std. Error	p-value	OR
Intercept	7.614	0.975	< 0.05	2.026e+03
h_n	-3.215	0.553	< 0.05	4.016e-02
h_o	7.828	0.685	< 0.05	2.511e+03
h_i	-8.761	0.696	< 0.05	1.567e-04
z	-3.057	0.515	< 0.05	4.703e-02
s_o	8.109	0.688	< 0.05	3.324e+03
s_i	-7.645	0.681	< 0.05	4.785e-04
φ_o	3.797	0.416	< 0.05	4.455e+01
φ_i	-6.929	0.544	< 0.05	9.794e-04
β_f	0.190	0.387	0.6	1.209e+00
β_m	0.378	1.833	0.8	1.460e+00
α_{fi}	-6.584	0.753	< 0.05	1.383e-03
α_{fn}	-22.538	1.447	< 0.05	1.628e-10
α_{mi}	-2.205	3.016	0.46	1.103e-01
α_{mn}	-0.948	0.472	0.04	3.875e-01

Classification trees

The classification tree in Figure 3.4 is a tree pruned to avoid over fitting the data, that minimises the cross-validated error [136]. The parameters actually selected by the algorithm (shown to have discriminant value) to construct the tree shown were: h_i, h_o, s_i, s_o, α_{fn}, α_{fo}, φ_i and φ_o. The proportion of observations correctly classified at each leaf can be used to represent the likely proportion of similarly classified observations of unsampled data at the field conditions defined by that terminal node [137]. The proportions of classifications on the five terminal nodes that predicted the class 1 ranged from 0.059 ($\frac{1}{(1+16}$) to 0.35 ($\frac{9}{9+17}$). Further simulated datasets, with higher number of observations, produced slightly different trees, but they agree on the parameters selected for their construction and have the same basic structure.

The parameters most closely associated with the counter intuitive results are consistent between logistic regression and classification trees. Logistic regression is considered to be a more potent method, [138] but the schematic nature of the trees provides a clearer understanding of the interactions between the parameters and does offer a set of rules to follow in order to try and achieve a particular outcome.

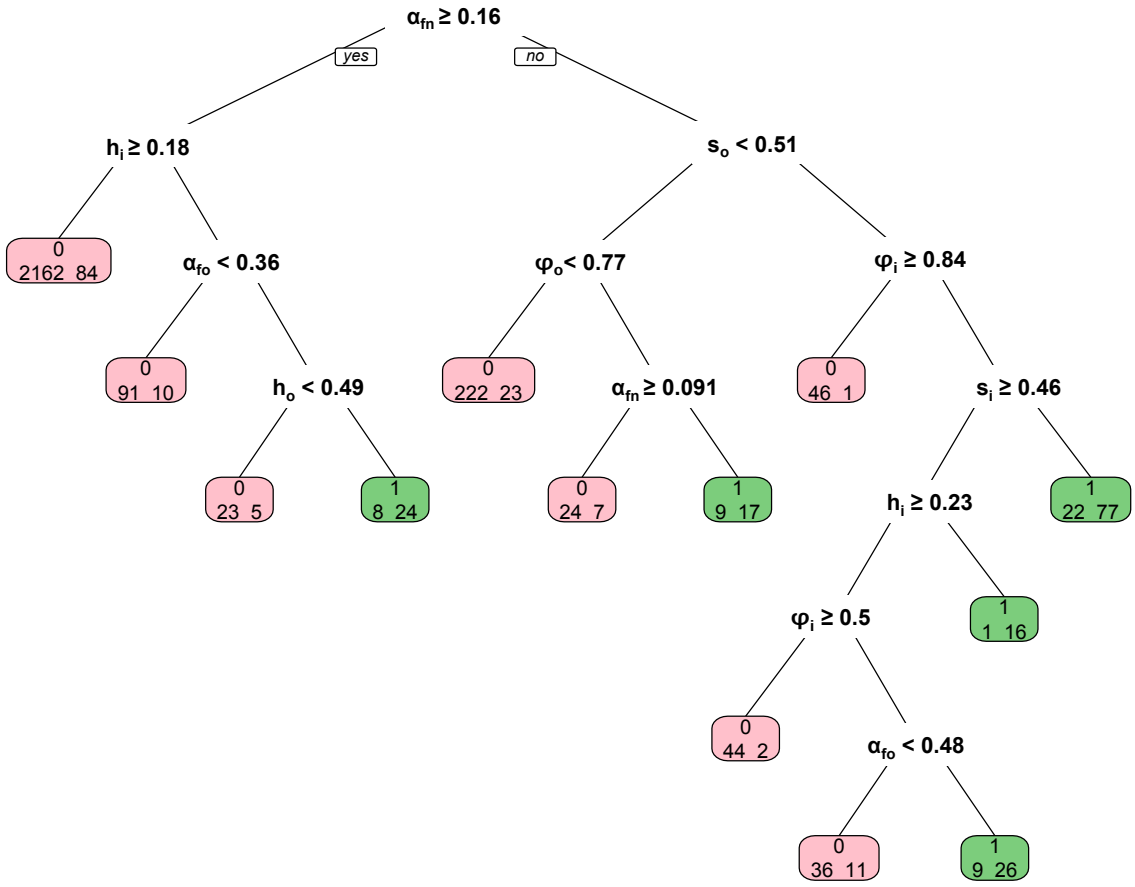

Figure 3.4: An example of a classification tree of outcomes $y > 1$ (class 1) and $y < 1$ (class 0) using all the parameters in the model except k. The items displayed in the nodes in the tree diagram are: the criterion for making the decision (e.g.: $\alpha_{fn} \geq 0.1626$), the predicted class for that terminal node (0 or 1) and the number of observations correctly classified to the class versus the number misclassified, achieved by cross-validation (e.g.: 2162/84). To proceed through the diagram at a given node move to the left branch if the stated condition is true (yes) and to the right if false (no).

Constrained datasets

Objective analysis using logistic regression and classification trees led to a subjective conclusion as to why counter-intuitive results ($y > 1$) occur (see later). To validate the results new datasets with constrains on the parameters based on the tree decision criteria (e.g.: $\alpha_{fn} < 0.2$, $s_o > 0.5$, $\varphi_i < 0.8$, $s_i < 0.5$), were generated with the expectation of reducing substantially the number of observations where $y > 1$. From 3,000 simulations produced with the previous constrain none resulted in $y > 1$. To end the analysis, a new dataset was generated with the above constrains to examine the impact of the different parameters on mean fitness when all observations lead to $y < 1$. Figure 3.1 shows in red the PRCC results calculated with this constrained dataset. The estimates values for the original dataset (black circles) and the constrained dataset (red diamonds) are similar, which implies that the presence of the counter-intuitive scenarios where $y > 1$ did not influence the overall correlation between the parameters and mean fitness. There was, however, a difference in the PRCC between the parameter s_o (selection in the environment with insecticide outside the household) and the ratio of change of resistance allele frequency in the male population, it shows no correlation in this dataset while it did in the original dataset. The most plausible explanation is that male selective pressure occurred mostly in the niche with insecticide employed outside. These selection coefficients were reasonable high in the simulations that led to the unexpected outcome, so it can be considered normal that it showed impact in the original dataset and not in the new dataset.

Dynamics of spread

The dynamics of spread was investigated by checking allele frequency at generation 1000 and determining resistance status (allele fixed if frequency greater than 0.99, lost if smaller than 0.001 and at equilibrium if change of frequency smaller than 0.001 in the last 50 generations) in the original dataset, subset with simulations that scored $y > 1$ and constrained dataset. Results are shown in Table 3.5.

Table 3.5: Dynamics of spread of resistance allele frequency. Percentage of simulations that lead to loss, fixation and equilibrium of resistance allele in the original dataset, subset of the original dataset with simulations that scored $y > 1$ and constrained dataset based on the classification tree criteria ($\alpha_{fn} < 0.2$, $s_o > 0.5$, $\varphi_i < 0.8$, $s_i < 0.5$).

	% Loss	%Fixation	%Equilibrium
Original dataset	49	30	21
Original subset with y >1	3	72	25
Constrained dataset	8	79	13

3.4 Discussion

Protection against vector borne diseases predominantly depends upon the usage of insecticides. Different strategies of delivery, in combination or independently, can be enforced while trying to minimise the emergence or impact of resistance. This study presents a model where mosquitoes face an heterogeneous environment of four different niches, considering the use of insecticides outside the household and the existence of insecticide free areas (refugia).

The model also allows one to check the effect on resistance spread of new generation long-lasting insecticidal nets, that incorporate a synergist, reported to have improved increased efficacy against pyrethroid-resistant malaria vectors [128, 139]. It would be simple to add more niches, essentially adding extra columns to Table 3.1, which is a major benefit of developing such a flexible methodology. For example, the outside niche could have high or low levels of insecticide. Under these conditions of different insecticide concentrations the resistance allele may be recessive or dominant respectively, with a huge impact on rate of resistance. It would also be possible to allow for mosquitoes to be exposed to more than one niche, by multiplying the fitness in each niche. This would however increase the complexity of the model and, as presented, the model demonstrates how interesting results can be derived from simple approaches.

Here it was only considered the existence of different niches, each with a different use of a single insecticide, not allowing for deployment of a second insecticide with a different mode of action. This would require a new model assuming 2 loci, each encoding resistance to a single insecticide. These 2 locus models are simple in principle [77] but in the present model of multiple niches it would increase the number of parameters significantly (total of 9 genotypes for each sex and increase the number of potential niches to reflect different combinations of insecticides) and would not be beneficial in the current exploration of the effect of the synergist.

The calibration of the model with field data proved to be problematic, hence the decision to sweep a range of values and check the outcomes. Restricting the male proportion inside the house to less than 20%, was based on personal communication since the numbers of males that enter households is rarely reported in hut trials, consequently it must be seen as a rough estimate. This work, in spite of the simplicity of the model, is illustrative of the advantages of modelling; an overall understanding that would not be possible by specific calibration and also the emergence of non-intuitive outcomes. Mathematical models have been used to expose other counter-intuitive results such that indoor residual spray (IRS) of insecticides in conjunction with bed nets can show antagonism, arising via interference of their modes of action while it is mostly assumed that the two tools have synergistic benefits in reducing malaria transmission [140].

Experimental studies have been conducted in the field to assess the potential of pro-

totypes of bed nets that incorporate a pyrethroid insecticide in the side panels and the synergist PBO plus insecticide on the roof. An experimental hut trial in Tanzania failed to demonstrate improvements in mosquito mortality, passage through holes and feeding rates, when compared to standard insecticide impregnated nets against *Culex quinquefasciatus* [139] and only moderate performance against pyrethroid-resistant *Anopheles gambiae* M taxon was reported in another trial in southern Benin [141]. However, Killeen and colleagues [142] noted that the manufacturer of the prototype claims only that the net has greater efficacy than its predecessor and that their simulations corroborates this claim, which the findings of this work also support.

From the PRCC analysis is straightforward to infer that the success of control campaigns depends mostly on the proportion of mosquitoes that encounters insecticides and on the fitness scaling factors φ, that define survival of SS genotypes when contacting insecticides. These last parameters were introduced to emphasise possible differences between niches on the impact on fully susceptible SS genotypes when facing insecticides. Unsurprisingly the control measures are more effective the higher their values. The inclusion of this parameter was crucial because it considers the complexity of fitness, and incorporates the differential environmental effects of insecticides across different genotypes.

The present model includes two parameters that relate to the effects of application of a synergist in combination with a insecticide: k, the reduced survival due the synergist and the proportions of mosquitoes, β (males and females), that meet both chemicals. The magnitude of the correlation between males mean fitness and k and β_m is very small, presumably due to the small proportion of males that is exposed to insecticide in an household, but k and β only correlates moderately with females mean fitness and ratio of change of allele frequency. This indicates that the synergist has a small impact in controlling the population, but even small values of k will help to recover the effect of the insecticide, and possibly this is the main contribution of the synergist. Nevertheless adding synergists to bed nets does decrease the rate at which resistance spreads in about 90% of scenarios.

It was not possible to use the model to investigate the overall impact of changes in fitness in the mosquito population dynamics, which would require a more elaborate model incorporating demography (e.g. [143]), and more specifically to investigate the effects of targeting mainly females, decreasing their fitness more than that of males, which are generally regarded as determining overall population regulation.

The finding that situations can arise when having a fully effective synergist contributes to intensify the spread of resistance is the most interesting result of this work. The setting in which it emerges is very specific, it encompasses a strong selective pressure for resistance in the niche outside the household (mean $s_o = 0.62$ while mean $s_i = 0.41$; these and the following figures come from the simulated subset where $y > 1$),

most mosquitoes are exposed to insecticides (mean α_n for females is 0.16 and 0.44 for males) and there are different dominance values in the niches were insecticide is deployed (mean $h_o = 0.61$ and mean $h_i = 0.37$).

In these circumstances the insecticide in the bed nets will be largely ineffective and the pressure for selection is weak. It seems as if this niche is acting as a refugia for susceptibles, that will contribute with their susceptibles genes for the next generation, therefore decreasing the resistance allele frequency. If a fully effective synergist ($k = 0$) is present, the fitness of all genotypes inside the house will be zero (k affects the 3 genotypes equally, so all mosquitoes die irrespective of their genotype) and the next generation will be mostly composed of the progeny of survivors from the niche outside the household, where selection for resistance was high. An hypothesis is that in this particular case the synergist is removing the refugia of weak selection in the house thereby magnifying the effects of selection for resistance outside the house.

Males and females showed the same patterns of spread, 49% converged to fixation, 30% to loss and 21% to equilibrium. Equilibrium in single niche models can only be achieved if there is heterozygote advantage [92] which was not postulated in the model because dominance coefficient h lie between 0 and 1 (Table 3.3). The balancing effects of different selection and dominance acting in different environments in the same population seems to be able to keep the resistance allele frequency at equilibrium. As an hypothetical example, a mutation which is dominant for insecticide resistance but has a large, recessive effect on fitness. As resistance starts to spread, most R alleles are in heterozygotes which resist insecticides and do not pay the fitness penalty. As R increases in frequency the proportion of RR homozygotes increases, so fitness penalties escalate until an equilibrium occurs. In effect, the marginal fitnesses (i.e. the average over all niches) generated by Table 3.1 and Equations 1 to 7 result in the heterozygote being the most fit in some simulations. As far as it was possible to verify it has not been reported in the field, possibly because it is not currently regarded as a likely occurrence.

The three dynamics of spread were predicted in subsequent analysis (Table 3.5) by estimating allele frequency for 1,000 generations. In the simulations that scored $y > 1$, 3% eventually lead to loss, 72% to fixation and 25% to equilibrium. It is overall a worse picture than with the original dataset. The choice of parameter values is important, for example constraining the dataset reduces the possibilities of reaching equilibrium, i.e., fixation of resistance was much more likely. Analysing only the simulations that lead to fixation in the original dataset generated results very similar (not shown) to Figure 3.1 and Table 3.4.

These results emphasise a very important fact often overlooked in modelling resistance: that it is highly dangerous to consider selection in only a single niche, isolated from other selection pressures, and then extrapolate the results from the single niche to

the whole population. In this case it seems reasonable to conclude that adding effective synergists will reduce selection for resistance in the household niche because all three genotypes are killed. The impact that a fully effective synergist will have in disease transmission is a fundamental question, that cannot be directly answered by the results presented here, because it is not clear how the genetic concept of fitness translates into demographic factors such as mosquito population size and longevity that determine the intensity of disease transmission. On the other hand, as noted above, if use of a synergist throws most of the selection pressure onto another niche then overall the rate of selection for resistance may increase. Consequently the public use of insecticide within the home (predominantly as wall sprays and/or bed nets) cannot be investigated isolated from other insecticide applications that mosquitoes may encounter during their lifetime. This suggest that the malaria community is correct in being alarmed at the often uncontrolled use of insecticides in applications such as agriculture [46, 144].

Chapter 4

Modelling insecticide impact and resistance in a malaria vector mosquito population

Abstract

Many current strategies to control malaria rely on the use of insecticide. These can be targeted at various stages of mosquito development but control is undermined by the continual evolution of resistant mosquitoes. Here we present a model that considers the stage-structured nature of the mosquito life cycle and, most importantly, allows for the tracking of insecticide resistant genotypes. In this way it is possible to understand the population dynamics of mosquitoes throughout their whole lifecycle while assessing the impact of the most common vector control interventions, alone and in combination, and the spread of insecticide resistance that those interventions induce. The model consists of a system of difference equations, that describes the immature (eggs, larvae and pupae) and adults stages, for males and females separately and that incorporates density-dependent regulation of mosquito larvae in breeding sites. We determined a threshold level of mosquitoes below which transmission of malaria is interrupted, based on a classic derivation of the malaria reproductive rate and used it to assess the effectiveness of different control strategies. Using equilibrium and stability analysis we concluded that the model can have two locally stable fixed points, one trivial (extinction) and one positive (stable population). We employed a sensitivity analysis technique to explore the impact of the parameters on the number of individuals in each stage, with particular interest on the number of adult females. We simulated different scenarios of insecticide deployment by changing some key parameters in the model. The analysis we performed explored the comparative impact of insecticide treated nets, indoor residual spraying, larvicides and pupacides, the benefits that can be achieved by using these intervention alone and in combinations and the level of resistance that arises from different amounts of insecticide usage. We concluded that targeting the larval stages achieves the greatest reduction on the adult population followed by targeting of the non-seeking females stage, as provided by indoor residual spraying. According to our results, low levels of resistance can induce failure of interventions, and the rate of spread of resistance is faster when insecticides target the larval stages.

4.1 Introduction

Modelling malaria goes back to the beginning of the past century, to the pioneering works of Ross [145] and later extended by Macdonald [146], whose assumptions have been thoroughly explored by Smith and McKenzie (2004) [147]. Since then, numerous

models have been developed, predominantly studying the dynamics of transmission. Many influenced by the seminal work of Kermack and McKendrick (1927) [148] use a deterministic epidemiological compartment approach, employing differential equations. A few others focused on the within host dynamics and in the study of the evolution and spread of the parasite in the populations [149]. From these models, we have learned for example that malaria can only exist in a population beyond a critical threshold of mosquito density, that adulticides are generally more efficient at controlling malaria than larvicides and that the duration of efficacy of a prospective vaccine has the biggest impact in an eventual vaccination programme [150].

Models have also enhanced our understanding of the impact of different vector control strategies on transmission. Ronald Ross (1905) [151] was the first to investigate the effects of larval control relating it to mosquito dispersal and density. Larviciding and environmental management were the only tools available in the beginning of the past century and anti-larval measures were used with some success before the advent of insecticides that could target adults. Nevertheless, this strategy was mostly discontinued in favour of indoor residual spraying (IRS) and insecticide treated nets (ITNs) [152]. One of the reasons for this change were the conclusions of the Ross-Macdonald model that postulated that the greatest reductions in malaria transmission can be achieved by reducing the longevity of the adult female vector population, best achieved by killing adult vectors indoors. However, more recent models showed that larval interventions have potential in reducing transmission intensity and incidence of malaria [153]. Contrary to adults that can actively avoid many intervention measures reducing the impact on malaria transmission, immature stages cannot escape control measures [154]. In addition, others have shown that the effects of reducing adult emergence is multiplicative and has an even greater effect on the basic reproductive rate than reducing survival alone [152, 155].

Saul *et al.* [156] developed a feeding cycle model of mosquito and malaria transmission that has since been extended to model several vector control interventions. Le Menach *et al.* [157] demonstrated, using a spatially explicit feeding cycle model, that ITNs simultaneously reduces lifespans, lengthen the feeding cycle and divert bites onto non-human hosts. Killeen and Smith (2007) [158] explored the impact of ITNs in the community. They found that ITNs have a big impact in reducing transmission at individual and community levels but that excito-repellent properties of the pyrethroid insecticide can increase exposure of unprotected humans if there is a lack of alternative hosts. Chitnis *et al.* (2010) [159] explored the effects of both strategies with different insecticides, concluding that IRS with bendiocarb provides the best community protection, IRS with DDT provides good personal protection and ITNs provide the best personal protection. Recently the model was extended to include seasonality [160]. Some models have also assessed the impact of different combinations of interventions

[140, 159, 161] and explored the potential of interventions like zooprophylaxis [158, 162], genetic manipulation of mosquitoes [163] and the use of fungal entomopathogens [164].

The model by Depinay *et al.* (2004) [165] was the first vector population dynamics model to integrate biological and environmental factors. They explored the effects of temperature, moisture, nutrient competition (as a regulatory mechanism in larval stages), dispersal of adults and predation and diseases (in all different stages) on mosquito abundance, while allowing for some exploration of control interventions. More recently Eckhoff (2011) [166] presented a model that also considers vector ecology, the impact of different interventions and in addition disease transmission. The emphasis of that work was on local tailoring and design of models in order to pursue local elimination goals.

Only a few models considered the dynamics of the mosquito population and modelled the entire life cycle of the mosquito. Most models of mosquito population dynamics have focussed on *Aedes* mosquitoes and have used ordinary differential equations (ODEs) [167], delay differential equations (DDEs) [168, 169], stochastic individual based models [165, 170, 171] or difference equations [172, 173]. Lu and Li (2011) [174] developed a stage-structured population dynamics model for mosquitoes in general, derived a net reproductive number and examined analytically the dynamics of the system. White *et al.* (2011) present an ecological model for the whole lifecycle of *Anopheles gambiae* populations using a Bayesian approach, with density-dependent regulation of the population at larval stages also with the aim of exploring the impact of vector control interventions.

However, there has been no mosquito population dynamics modelling that simultaneously analyses the impact of insecticidal vector control interventions and its inevitable consequence in driving the spread of insecticide resistance. It is necessary to explore the impact on vector populations, and on disease transmission, of intensive use of insecticides and intensive selection pressure for resistance at the same time. The model that we describe here is intended to be linked in the future with the model for malaria in mosquitoes described in [122]. Therefore, we focus on the vector of malaria *Anopheles spp.* even though the model could easily be modified to accommodate the specificities of other species.

Our model for population dynamics considers overlapping generations and a discrete time step of one day. It includes the dependence of the emergence rate of new mosquitoes on the number of eggs laid so it includes the nonlinear effects of adulticides on reducing the population size. We give particular emphasis to the demographic impact of insecticidal interventions and the emergence of resistance and some insights on how it might translate in terms of disease transmission.

4.2 Description of the model

We present a schematic outline of the model structure and the role of its parameters in Figure 4.1. We formulated a discrete-time stage-structured model, based on a system of difference equations. The inclusion of the stage-structure allows more realistic modelling of the life cycle, the different stages can have different responses to environment and we included regulating factors of the population. We use discrete time steps of 1 day that better capture the circadian nature of the mosquito's life.

Anopheline mosquitoes undergo complete metamorphosis going through four distinct stages of development during a lifetime: egg, pupa, larva, and adult. Adult females lay eggs after a blood meal in permanent water or temporary sites that have been flooded. Within one or two days to a week or more, the eggs hatch into larvae that breathe air through tubes, eating floating organic matter. Larvae moult four times until they became pupae. Pupae live near the surface of the water and do not eat, breathing through siphons on their back, after a few days as a pupa the adult emerges. The adult lives for a few days to several weeks [120]. The juvenile stages are similar in males and females, but the adult stage differs significantly.

We replicate this model three times for the homozygous susceptible, SS, heterozygous, RS, and homozygous resistant, RR, genotypes to allow for insecticide resistance incorporation. We assume that males can mate multiple times but female mosquitoes only mate once immediately after emergence and carry the sperm of the male with them for the rest of their lives. We explicitly track the genotype of the male a given female mated with.

We use superscript f to denote females and m to denote males. We also append the superscript j, where $j \in \{SS, RS, RR\}$ denotes the genotype of the mosquito. For adult female mosquitoes, we append an additional superscript k, where $k \in \{SS, RS, RR\}$ denotes the genotype of the male mosquito that she mated with. We describe the parameters of the model in Table 4.1. The parameter values used for the life cycle are presented in Table 4.3 and the derived equations that govern the cycle are presented next.

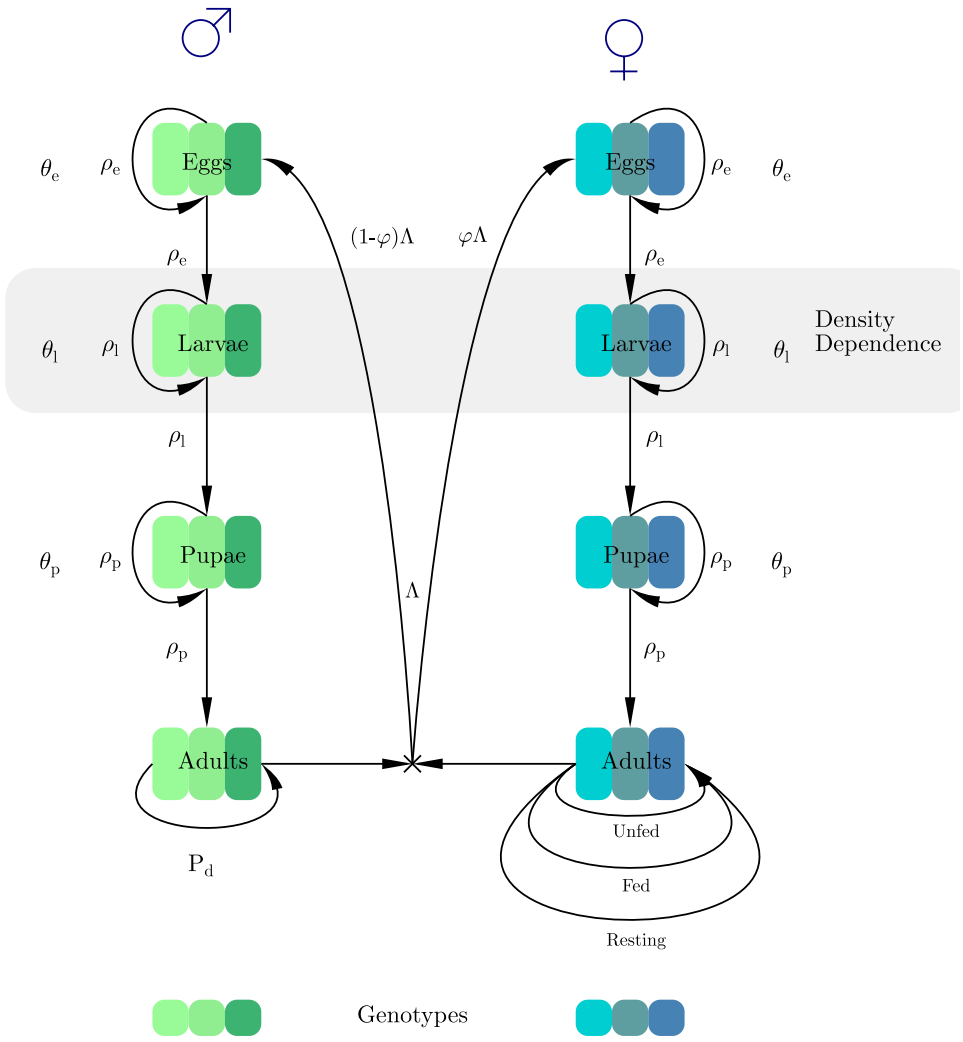

Figure 4.1: The mosquito population stage-structured model. The adult stage dynamics is considerably different in males and females: male adults are composed of newly emerged individuals plus the adult males that survived the previous day (P_d); female adults are grouped in 3 classes: (i) unfed individuals that are currently host-seeking (newly emerged individuals, individuals that did not find a host the previous days and individuals that laid eggs the previous day and are starting a new gonotrophic cycle, Equation 4.6a); (ii) fed individuals, Equation 4.6b and (iii) resting individuals, Equation 4.6c. The model tracks the 3 potential genotypes $j \in (SS, RS, RR)$ of the individuals through their developmental stages. The total number of eggs laid by all females is Λ (Equations 4.7), of which $(1-\varphi)\Lambda$ are males and $\varphi\Lambda$ are females. We assume adult females mate once upon emergence, while males can mate multiple times. The θ parameters refer to the duration of each stage in days, and ρ to the proportion of individuals that survive per day in a given stage (e eggs, l larvae, p pupae).

Table 4.1: Description of parameters for the model of mosquito population dynamics. The superscripts, m and f denote males and females, j and k, denote genotype and can be any of SS, representing homozygous susceptible, RS, representing heterozygous, and RR, representing homozygous resistant.

θ_e:	Duration of the egg stage. Dimension: Time. $\theta_e \in \mathbb{N}$.
ρ_e^{fj}:	Proportion of female eggs of genotype j that survive one day. $0 < \rho_e^{fj} < 1$.
ρ_e^{mj}:	Proportion of male eggs of genotype j that survive one day. $0 < \rho_e^{mj} < 1$.
θ_l:	Duration of the larval stage. Dimension: Time. $\theta_l \in \mathbb{N}$.
ρ_l^{fj}:	Density-independent proportion of female larvae of genotype j that survive one day. $0 < \rho_l^{fj} < 1$.
ρ_l^{mj}:	Density-independent proportion of male larvae of genotype j that survive one day. $0 < \rho_l^{mj} < 1$.
$\gamma(t)$:	Resource availability at time t. Dimension: 1/Animals. $\gamma(t) > 0$.
c_i^{fj}:	Effect of larval competition on female larvae of genotype j in stage i where $i \in \mathbb{N}$ and $\theta_e < i \leq \theta_e + \theta_l$. $0 < c_i^{fj} \leq 1$.
c_i^{mj}:	Effect of larval competition on male larvae of genotype j in stage i where $i \in \mathbb{N}$ and $\theta_e < i \leq \theta_e + \theta_l$. $0 < c_i^{mj} \leq 1$.
ω_i^{fj}:	Relative resource consumption of female larvae of genotype j in stage i where $i \in \mathbb{N}$ and $\theta_e < i \leq \theta_e + \theta_l$. $0 < \omega_i^{fj} \leq 1$.
ω_i^{mj}:	Relative resource consumption of male larvae of genotype j in stage i where $i \in \mathbb{N}$ and $\theta_e < i \leq \theta_e + \theta_l$. $0 < \omega_i^{mj} \leq 1$.
θ_p:	Duration of the pupal stage. Dimension: Time. $\theta_p \in \mathbb{N}$.
ρ_p^{fj}:	Proportion of female pupae of genotype j that survive one day. $0 < \rho_p^{fj} < 1$.
ρ_p^{mj}:	Proportion of male pupae of genotype j that survive one day. $0 < \rho_p^{mj} < 1$.
τ:	Duration of the resting period of the gonotrophic cycle of a female adult mosquito. Dimension: Time. $\tau \in \mathbb{N}$.
H^j:	Proportion of adult females of genotype j that find a host and successfully feed while seeking.
ρ_s^j:	Proportion of adult females of genotype j that survive while host-seeking per day.
ρ_n^j:	Proportion of adult females of genotype j that survive while resting per day.
P_d^j:	Proportion of male adults of genotype j that survive one day. $0 < P_d^j < 1$.
b^j:	Number of eggs laid per oviposit by female mosquitoes of genotype j. $b^j > 0$.
σ^k:	Mating viability of a male of genotype k. $0 < \sigma^k \leq 1$.
φ:	Proportion of eggs that are female. $0 < \varphi < 1$.
μ:	Probability of a susceptible allele S mutating to a resistant allele R in any generation.
λ^j:	Number of eggs of genotype j that migrated from a different population of mosquitoes.

Immature Stages

We define the duration of the juvenile stages by ζ:

$$\zeta = \theta_e + \theta_l + \theta_p$$

Where θ_e is the duration of the egg stage, θ_l the duration of the larval stage and θ_p the duration of the pupal stage.

We track stage development in i days and describe the female juvenile mosquito population of genotype j at time t by $x_i^{fj}(t) \in \bar{\mathbb{R}}_+^{\zeta}$, where

- $x_i^{fj}(t)$ for $1 \leq i \leq \theta_e$ denote the female egg stages of age i at time t.
- $x_i^{fj}(t)$ for $\theta_e + 1 \leq i \leq \theta_e + \theta_l$ denote the female larval stages of age i at time t.
- $x_i^{fj}(t)$ for $\theta_e + \theta_l + 1 \leq i \leq \zeta$ denote the female pupal stages of age i at time t.

The male juvenile population is described in an analogous manner.

In Table 4.2 we present a summary of the immature stages of the mosquito life cycle. It shows the duration of each stage and the correspondent age of mosquitoes. It should facilitate the understanding of the dynamics of the juvenile male and female mosquito population of genotype j that we describe in Equations (4.1) and (4.2). At each iteration mosquitoes are moved forward in chronological time (to $t+1$) and in the developmental time (to $i+1$).

Table 4.2: Mosquito immature life stages summary.

Stage	Duration	Timespan i (age in days)
New eggs	-	1
Developing eggs	θ_e	$2 : \theta_e$
Larvae	θ_l	$\theta_e + 1 : \theta_e + \theta_l$
Pupae	θ_p	$\theta_e + \theta_l + 1 : \zeta$

The juvenile female mosquito population of genotype j at time t, $x_i^{fj}(t) \in \bar{\mathbb{R}}_+^{\zeta}$ is,

$$x_1^{fj}(t+1) = \varphi \Lambda^{j'}(t) \tag{4.1a}$$

$$x_{i+1}^{fj}(t+1) = \rho_e^{fj} x_i^{fj}(t) \quad \text{for } 1 \leq i \leq \theta_e \tag{4.1b}$$

$$x_{i+1}^{fj}(t+1) = \frac{\rho_l^{fj}}{1 + \gamma(t) c_i^{fj} L(t)} x_i^{fj}(t) \quad \text{for } \theta_e + 1 \leq i \leq \theta_e + \theta_l \tag{4.1c}$$

$$x_{i+1}^{fj}(t+1) = \rho_p^{fj} x_i^{fj}(t) \quad \text{for } \theta_e + \theta_l + 1 \leq i \leq \zeta - 1 \tag{4.1d}$$

The juvenile male mosquito population of genotype j at time t, $x_i^{mj}(t) \in \bar{\mathbb{R}}_+^{\zeta}$ is similarly defined,

$$x_1^{mj}(t+1) = (1-\varphi)\Lambda^{j'}(t) \tag{4.2a}$$

$$x_{i+1}^{mj}(t+1) = \rho_e^{mj} x_i^{mj}(t) \quad \text{for } 1 \leq i \leq \theta_e \tag{4.2b}$$

$$x_{i+1}^{mj}(t+1) = \frac{\rho_l^{mj}}{1+\gamma(t)c_i^{mj}L(t)} x_i^{mj}(t) \quad \text{for } \theta_e + 1 \leq i \leq \theta_e + \theta_l \tag{4.2c}$$

$$x_{i+1}^{mj}(t+1) = \rho_p^{mj} x_i^{mj}(t) \quad \text{for } \theta_e + \theta_l + 1 \leq i \leq \zeta - 1 \tag{4.2d}$$

Here, $L(t)$ is the total larval consumption of resources by male and female larvae of all genotypes,

$$L(t) = \sum_{j \in \{SS,RS,RR\}} \left(\sum_{k=\theta_e+1}^{\theta_e+\theta_l} \omega_k^{fj} x_k^{fj}(t) + \omega_k^{mj} x_k^{mj}(t) \right). \tag{4.3}$$

Equations 4.1a and 4.2a are the input of newly laid eggs, determining which proportion of the eggs laid are male or female, according to the parameter φ, and of which of the 3 possible genotypes. These eggs will then go through development that will last θ_e days. Equations 4.1b and 4.2b describe the progress of the eggs through the stage that depends on a surviving probability ρ_e.

Equations 4.1c and 4.2c characterise the larval stage that follows, with a duration of θ_l days. Density-dependent population regulation (DDPR) in this model occurs in both sexes and only in the larval stages. There is no DDPR of adults, and pupae mosquito aquatic stages undergo morphological development but do not feed. Equations 4.1c, 4.2c and 4.3 incorporate density-dependent regulation that is analogous to Beverton-Holt type (B-H), which is a classic discrete time population growth model whose continuous-time equivalent is logistic growth towards a carrying capacity.

Classically the Beverton-Holt equation is:

$$x_{t+1} = \frac{R_0}{1+\frac{x_t}{M}} x_t \tag{4.4}$$

Where R_0 is the growth rate per generation and $K = (R_0 - 1)M$, is the carrying capacity, the maximum population size that the environment can sustain indefinitely.

Our Equations 4.1c and 4.2c can be regarded as equivalent to the B-H as will now be explained. Classically the B-H models does not distinguish between different 'types' in the population (e.g. sexes). It can do so, as in this case, by summating the number of individuals of each 'type' in the next generation, hence we use the number of each type x_i^{mj} and x_i^{fj} in the numerator of 4.1c and 4.2c rather than X_t. Overall R_0 is determined by other parts of the demographic model so here it is replaced simply by the survival probability (ρ_l^{mj} or ρ_l^{fj}). The parameter $\gamma(t)$ is equivalent to 1/M so needs no discussion. Setting $c_i^{f/mj}$=1 in Equations 4.1c and 4.2c means it disappears from the equation, and setting $\omega^{m/fj}$ =1 in Equation 4.3 means L(t) effectively becomes x_t in the denominator; under these conditions Equations 4.1c and 4.2c become equivalent to the B-H. The reason why we do not constrain $c_i^{f/mj} = \omega^{m/fj} = 1$ is for biological flexibility and plausibility. The competitive ability, $c_i^{f/mj}$, of larvae may differ depending on their genotype (for example resistant forms may pay a 'fitness penalty') and the resource consumption, $\omega^{m/fj}$, of each type may vary (for example, resistant forms may be larger and consume more resources).

The individuals that survive the larval stage will undergo pupal development for θ_p days. Progress in this stage is dependent on a survival probability ρ_p.

Adults

We denote the population of adult male mosquitoes of genotype j at time t by $y^{mj}(t) \in \bar{\mathbb{R}}_+$. The equation for the adult male population of genotype j is,

$$y^{mj}(t+1) = P_d^j y^{mj}(t) + \rho_p^{mj} x_\zeta^{mj}(t), \tag{4.5}$$

Adult male population at a given time is the result of the sum of adults that emerged from pupae that day ($\rho_p^{mj} x_\zeta^{mj}(t)$) and male adults that survived from the previous day ($P_d^j y^{mj}(t)$).

An adult mosquito emerges from pupae of stage x_ζ. Only females anophelines need to blood feed, because they need blood components in order to produce eggs, so the processes that govern the adult female population differ significantly from those of males. Upon emergence there will be mating, that happens during flight (unlike males the females mate only once). Once fertilised, females initiate the gonotrophic cycle that consists of 3 phases: foraging for a host and blood-feeding, digestion of the blood and egg maturation and the search for a suitable oviposition site and oviposition (Figure 4.1). The gonotrophic cycle will be repeated throughout the female lifespan.

The female adult population at a given time $t+1$ is given by Equations 4.6 (see Table 4.1 for parameters description). Females unfed in the current gonotrophic cycle are

given in Equation 4.6a. The first term are the newly emerged female adults of genotype j $[\rho_p^{fj} x_\zeta^{fj}(t)]$ that will mate with a male of genotype k $(P_d^k y^{mk}(t) + \rho_p^{mk} x_\zeta^{mk}(t))$, which has a mating viability σ^k [this term is normalised by dividing by the total male population]. The second term refers to other female individuals that were unfed adults in the previous day, that had not been successful in finding a host and therefore are still in the host-seeking state $[\rho_s^{fj}(1-H)y_1^{fjk}(t)]$. The third term are females that successfully laid eggs an are now seeking a host in their new gonotrophic cycle $[\rho_n^{fj} y_\tau^{fjk}(t)]$.

Equation 4.6b gives the second state of the adult population, this corresponds to individuals in y_1 that survived and successfully feed $(\rho_s^{fj} H^j)$.

For the rest of the duration of the gonotrophic cycle the female adult population will be governed by Equation 4.6c, that considers the survival probability while resting.

$$y_1^{fjk}(t+1) = \rho_p^{fj} x_\zeta^{fj}(t) \frac{\sigma^k (P_d^k y^{mk}(t) + \rho_p^{mk} x_\zeta^{mk}(t))}{\sum_{h\in\{SS,RS,RR\}} \sigma^h (P_d^h y^{mh}(t) + \rho_p^{mh} x_\zeta^{mh}(t))} + \rho_s^j (1-H^j) y_1^{fjk}(t) + \rho_n^j y_\tau^{fjk}(t), \tag{4.6a}$$

$$y_2^{fjk}(t+1) = \rho_s^j H^j y_1^{fjk}(t), \tag{4.6b}$$

$$y_{i+1}^{fjk}(t+1) = \rho_n^j y_i^{fjk}(t) \quad \text{for } 2 \le i \le \tau - 1. \tag{4.6c}$$

Egg laying, mutation and migration

The number of eggs laid of each genotype is calculated assuming random mating between the different genotypes.

The number of homozygous susceptible eggs laid at time t is,

$$\Lambda^{SS}(t) = b^{SS} \rho_n^{SS} \Big(y_\tau^{fSSSS}(t) + \tfrac{1}{2} y_\tau^{fSSRS}(t) \Big) + b^{RS} \rho_n^{RS} \Big(\tfrac{1}{2} y_\tau^{fRSSS}(t) + \tfrac{1}{4} y_\tau^{fRSRS}(t) \Big), \tag{4.7a}$$

The number of heterozygous eggs laid at time t is,

$$\begin{aligned} \Lambda^{RS}(t) = {} & b^{SS} \rho_n^{SS} \Big(\tfrac{1}{2} y_\tau^{fSSRS}(t) + y_\tau^{fSSRR}(t) \Big) \\ & + b^{RS} \rho_n^{RS} \Big(\tfrac{1}{2} y_\tau^{fRSSS}(t) + \tfrac{1}{2} y_\tau^{fRSRS}(t) + \tfrac{1}{2} y_\tau^{fRSRR}(t) \Big) \\ & + b^{RR} \rho_n^{RR} \Big(y_\tau^{fRRSS}(t) + \tfrac{1}{2} y_\tau^{fRRRS}(t) \Big), \end{aligned} \tag{4.7b}$$

The number of homozygous resistant eggs laid at time t is,

$$\Lambda^{RR}(t) = b^{RS}\rho_n^{RS}\Big(\tfrac{1}{4}y_\tau^{fRSRS}(t) + \tfrac{1}{2}y_\tau^{fRSRR}(t)\Big) + b^{RR}\rho_n^{RR}\Big(\tfrac{1}{2}y_\tau^{fRRRS}(t) + y_\tau^{fRRRR}(t)\Big). \tag{4.7c}$$

To incorporate the possibility of mutation we introduce μ, as the probability of a susceptible allele S mutating to a resistant allele R in any generation and vice versa (equal mutation rate). We assume that μ is so small that μ^2 is negligible, i.e., that no mosquito has a double mutation at the locus. We also include a parameter for migration of mosquitoes of a particular genotype to the environment, λ^j, given by a increase in the number of eggs of that particular genotype.

The number of homozygous susceptible eggs laid at time t, is,

$$\Lambda^{SS'}(t) = \frac{\Lambda^{SS}(t)(1-2\mu) + \Lambda^{RS}(t)\mu + \lambda^{SS}}{\Lambda_N}, \tag{4.8a}$$

The number of heterozygous eggs laid at time t is,

$$\Lambda^{RS'}(t) = \frac{\Lambda^{RS}(t)(1-2\mu) + \Lambda^{SS}(t)2\mu + \Lambda^{RR}(t)2\mu + \lambda^{RS}}{\Lambda_N}, \tag{4.8b}$$

The number of homozygous resistant eggs laid at time t is,

$$\Lambda^{RR'}(t) = \frac{\Lambda^{RR}(t)(1-2\mu) + \Lambda^{RS}(t)\mu + \lambda^{RR}}{\Lambda_N}. \tag{4.8c}$$

Where Λ_N is equal to the sum of the numerators in Equations 4.8.

4.3 Population reproductive rate

We also computed the vector population reproductive rate R_{0v}, that we defined as the number of newly emerged female adults produced by a single newly emerged female (Equation 4.9).

We derived this population reproductive rate because, due to the nonlinear effects of density-dependence on the larval stages, we could not use standard matrix population modelling (for example Lefkovitch projection matrices), that for linear models enables the calculation of a reproductive value using the eigenvalues and eigenvectors of the matrix [175].

At this point we will drop the subscripts related to genotypes, and compute just for the overall population.

$$R_{0v} = m\, E\, b\, \varphi\, \rho_e^{\theta_e}\, \rho_l^{\theta_l}\, \rho_p^{\theta_p} \tag{4.9}$$

Where:
m is the probability of a female finding a mate, we assume it to be 1;
E is the expected number of broods produced by a newly emerged female;
b is the number of eggs laid per oviposit.
φ is the proportion of eggs in a batch that are female.
$\rho_e^{\theta_e}\ \rho_l^{\theta_l}\ \rho_p^{\theta_p}$ are the probabilities of survival the imature stage (eggs, larvae, pupae) depending on their durations $(\theta_e, \theta_l, \theta_p)$.

We ignored the effects of density dependence on this calculation, because R_{0v} is defined at very low population densities when density-dependent regulation will be absent.

E is calculated as follows:

First determine $p(f)$, the probability that a female finds a host and feeds in her lifetime (Equation 4.10). The first term, $H\rho_s$, is the probability that at any given day she survived the seeking stage (ρ_s) and successfully found a host and fed (H). The second term is the probability that she has not found a host in a previous day ,$(1-H)$, but survived the seeking stage, ρ_s and will find a host and feed one of the next days.

$$p(f) = H\rho_s + H\sum_{i=1}^{\infty}[(1-H)\rho_s]^i \tag{4.10}$$

the sum in Equation 4.10 can be evaluated as $\frac{(1-H)\rho_s}{1-(1-H)\rho_s}$ given that $|(1-H)\rho_s| < 1$ [49] so that equation 4.10 becomes:

$$p(f) = H[\rho_s + \frac{(1-H)\rho_s}{1-(1-H)\rho_s}] \tag{4.11}$$

After successfully finding a host and feeding with probability $p(f)$, a female will go through a non seeking stage for egg maturation and oviposition (assumed to last τ days), with probability of survival ρ_n, after which she has completed a gonotrophic cycle. Equation 4.12, p(f)', is the probability of finding a host and surviving incubation.

$$p(f)' = p(f)\rho_n^\tau \tag{4.12}$$

Female mosquitoes go through this process more than once in their lifetime so that the number of broods produced by one adult female in her lifetime becomes:

$$E = p(f)' + p(f)'^2 + p(f)'^3 + ... \tag{4.13}$$

The first item is the probability of survival to produce one brood of eggs, the second the probability of survival to produce a second brood of eggs and so forth.

Equation 4.13 simplifies to:

$$E = \sum_{i=1}^{\infty} [H\rho_n^\tau [\rho_s + \frac{(1-H)\rho_s}{1-(1-H)\rho_s}]]^i \tag{4.14}$$

With the vector population reproductive rate R_{0v} we can predict the behaviour of the population under a given parameterisation, since the population will become extinct when $R_{0v} < 1$. This is an algebraic result that can be used to validate the results of the computational model; the same results suggest an absence of programming errors in the same simulation.

4.4 Malaria transmission

The main function of the model developed here is to study the dynamics of mosquito populations, and the response to perturbations caused by the use of insecticides. Nevertheless it is important to demonstrate the link with the basic reproductive rate of malaria, a classical quantity that has been used by vector entomologists, related with the potential of malaria transmission by a mosquito vector population. We determined the basic reproductive rate of malaria in Equation 4.15 as developed by Ross and Macdonald [176].

$$R_{0m} = \frac{ma^2 b_1 b_2}{gr} \tag{4.15}$$

In Equation 4.15:
$m = \frac{M}{N}$ is the number of female mosquitoes per human host where N is the size of the human population and M is the size of the female adult mosquito population;
a is the rate of biting on humans by a single mosquito (number of bites per unit time);
b_1 is the proportion of infected bites on human that produce an infection;
b_2 is the proportion of infected bites on mosquitoes that produce an infection;
r is the per capita rate of recovery for humans ($\frac{1}{r}$ is the average duration of infection in the human host);
g is the per capita mortality rate for mosquitoes ($\frac{1}{g}$ is the average life time of a mosquito).

In the context of the current work R_{0m} can be used to determine the size of the female mosquito population M that leads to the control of the disease. In principle to control the disease, we should reduce $R_{0m} < 1$, therefore the female adult population should be maintained at size $M < \frac{rg}{a^2 b_1 b_2} N$. We use simulations to compare the effects of reducing survival at different life stages on the time it takes to reach this population threshold size M.

4.5 Analysis

The model was implemented in R [98]. Verification was done to ensure that the model was programmed correctly, the algorithm has been implemented properly and that it does not contain errors or oversights. We started the population with 1000 eggs and zero individuals in all other stages (i.e., zero larvae, zero pupae, zero adults and zero eggs from day 1 onwards) and setting the parameters to discriminant values (ex: all survival proportion to 1) so that it was possible to monitor the population progression (both in time and stage) and to confirm the calculations.

We defined conditions on which to declare the population at equilibrium or in severe decline towards extinction. For the population to be considered in equilibrium the ratio of the number of adult females (Equations 4.6) of two consecutive days should be higher than 0.9999, for at least 20 days (there is expected to be a turnover of at least one generation in this period [177]). The population is considered in severe decline or eliminated if for 50 consecutive days the number of eggs and female adults is smaller than 1.

4.5.1 Fixed points

The core part of any analysis of a population dynamic model is to determine the asymptotic behaviour of the model. It is usually done by the identification of steady states, i.e. states that a population would maintain for all times, if initially started in it, and the determination of the local stability of these steady states, i.e. analysing whether the population abundance would return to the steady state, if displaced away from it by a (infinitesimally) small, but otherwise arbitrary amount.

Steady states are solutions to the equations where the state variables are not changing with respect to a dimension of interest, in this case, the number of individuals in each stage remains constant in time. These steady states are generally denominated fixed points and as they are good candidates for where the system might eventually end up, an analysis of this type could also validate the results obtained by simulation. Nevertheless, the model we developed is so complex (numerous parameters, numerous equations and non-linearity in the larval stages) that is not possible to treat it analytically and solve the system of equations for fixed points. Consequently, to simplify the analysis, we decided to check the dynamic of the general system without the distinction between genotypes and proceed by evaluating the equations for a set of parameters values. The parameters for which there is empirical data were assigned those values, while a range of values was assigned to the remaining and randomly sampled from it (Table 4.3). This leads to the generation of a set of 3000 combinations of parameters values and solving for the equilibrium points for each of the possible combinations of parameter values. We analysed the model for fixed points using the software *Maple 13*.

Table 4.3: Parameters values used in the fixed points and sensitivity analysis, the values used are identical for males and females when not explicitly stated otherwise.

Parameter	Short description	Value	Reference
θ_e	Duration of the egg stage	2	[177]
ρ_e	Proportion of eggs that survive one day	0 -1	This work
θ_l	Duration of the larval stage	10	[143, 177]
ρ_l	Density-independent proportion of larvae that survive one day	0 -1	This work
$\gamma(t)$	Resource availability at time t	3×10^{-11}-1×10^{-6}	This work
c_i	Effect of larval competition on larvae of all genotypes in stage i	0-1	This work
ω_i	Relative resource consumption of larvae of all genotypes in stage i	0-1	This work
θ_p	Duration of the pupal stage	2	[177]
ρ_p	Proportion of pupae of all genotypes that survive one day	0-1	This work
τ	Duration of the resting period of the gonotrophic cycle of a female adult mosquito	3	[159]
P_d^j	Proportion of male adults that survive one day	0-1	This work
b^j	Number of eggs laid per oviposit by female mosquitoes	100	This work
σ^k	Mating viability of males of all genotypes	0.5	This work
φ	Proportion of eggs that are female	0.5	[177]
μ	Probability of a susceptible allele S mutating to a resistant allele R in any generation	1	This work
λ	Number of eggs that migrated from a different population of mosquitoes	0	This work

Briefly, the fixed points can be found by determining the values of the variables that cause all of the variables to be the same in the next time step:

$$x_1^m = (1-\varphi)\Lambda \tag{4.16a}$$
$$x_1^f = \varphi\Lambda \tag{4.16b}$$
$$x_2^m = \rho_e^m x_1^m \tag{4.16c}$$
$$x_2^f = \rho_e^f x_1^f \tag{4.16d}$$
$$x_i^m = \frac{\rho_l^m}{1+\gamma c_{i-1}^m L} x_{i-1}^m \quad \text{for } 3 \le i \le 12 \tag{4.16e}$$
$$x_i^f = \frac{\rho_l^f}{1+\gamma c_{i-1}^f L} x_{i-1}^f \quad \text{for } 3 \le i \le 12 \tag{4.16f}$$
$$x_i^m = \rho_p^m x_{i-1}^m \quad \text{for } 13 \le i \le 14 \tag{4.16g}$$
$$x_i^f = \rho_p^f x_{i-1}^f \quad \text{for } 13 \le i \le 14 \tag{4.16h}$$
$$\vdots \tag{4.16i}$$

This results in a system of 32 equations; Male= 14 (ζ) immature stages + 1 adult stage plus Female= 14 (ζ) immature stages + 3 adult stages (Assuming values of $\theta_e = 2; \theta_l = 10; \theta_p = 2$).

The next step is to factor each equation:

$$x_1^m - (1-\varphi)\Lambda = 0 \tag{4.17a}$$
$$x_1^f - \varphi\Lambda = 0 \tag{4.17b}$$
$$x_2^m - \rho_e^m x_1^m = 0 \tag{4.17c}$$
$$x_2^f - \rho_e^f x_1^f = 0 \tag{4.17d}$$
$$x_i^m - \frac{\rho_l^m}{1+\gamma c_{i-1}^m L} x_{i-1}^m = 0 \quad \text{for } 3 \le i \le 12 \tag{4.17e}$$
$$x_i^f - \frac{\rho_l^f}{1+\gamma c_{i-1}^f L} x_{i-1}^f = 0 \quad \text{for } 3 \le i \le 12 \tag{4.17f}$$
$$x_i^m - \rho_p^m x_{i-1}^m = 0 \quad \text{for } 13 \le i \le 14 \tag{4.17g}$$
$$x_i^f - \rho_p^f x_{i-1}^f = 0 \quad \text{for } 13 \le i \le 14 \tag{4.17h}$$
$$\vdots \tag{4.17i}$$

and identify all possible solutions for all variables for all equations. This is a non-linear model so there may be more than one fixed point.

After identifying the fixed point(s) the next step is to determine whether the fixed

point(s) is (are) stable. To perform a local stability analysis we determine the Jacobian matrix of the system. The Jacobian matrix is the matrix of all first-order partial derivatives of the system of equations. Its importance lies in the fact that it represents the best linear approximation of the non-linear model near the fixed points. So that when we evaluate the Jacobian matrix at each fixed point we use the linearisation as a proxy of the behaviour of the non-linear model in the vicinity of the fixed point. We finish by determining the eigenvalues of this matrix.

For a non-linear multivariate model in discrete time a fixed point will be locally asymptotically stable if the absolute values of all eigenvalues (e) are less than one. For real eigenvalues, stability requires that both $e < 1$ and $-1 < e$. For complex eigenvalues $e = A \pm Bi$, stability requires that the absolute value $\sqrt{A^2 + B^2}$ be less than one [49].

The analysis showed that the model can have two fixed points: a trivial fixed point, where all the state variables are zero and therefore the population goes to extinction, and one positive.

Regarding stability both possible fixed points were found to be locally stable. Some of the eigenvalues calculated are complex, which shows that the system can spiral around the fixed point and a thorough analysis would be necessary (that we did not pursue) to explore this behaviour.

From the 3,000 simulations, only 103 (3.4%) have both trivial and positive fixed points, the majority have only a trivial fixed point, suggesting that the combinations of parameter values that lead to equilibrium is considerably narrow. Running the model with the 3,000 different parameterisations in R (running it for 600 days and check status, i.e., in equilibrium or eliminated) and identifying the fixed points in *Maple* achieved the same results, implying that both have been implemented correctly. The choice of 600 for the simulations iteration time (days) is a good compromise between computation time and the convergence of most of the simulations to a state (equilibrium or extinction).

We compared the parameters values in the simulations that reached a positive equilibrium and that lead to extinction of the population, and verified that the parameters ρ_e, ρ_l, ρ_p, ρ_s, ρ_n and H are statistically significantly higher in a t-test in the simulations that reach an equilibrium (alternative hypothesis: true difference in means is greater than 0; p-value < 0.05), while the parameters c, ω, γ and P_d were not statistically different (p-value> 0.05). The later are associated with larval competition which is absent near extinction and therefore expected to have no effect.

We compared the predicted long term status of the population using the R_{0v} with the outcome of running the model for 600 days, for the 3,000 parameter combinations. The R_{0v} predicted correctly 95% of the simulations that lead to extinction and correctly all the simulations that lead to a positive equilibrium. Considering these results our R_{0v} seems to be overestimating the reproductive rate. We did not included in the

derivation of the R_{0v} the density-dependence regulation in the larval stages, in principle the density-dependence will be important in populations that increase in size and is responsible for a constant rate of emergence of larvae, but it is not expected to influence the outcome of a population that is supposed to be in decline. Nevertheless, the discordant values of R_{0v} (140 simulations) ranged from 1 to 3.14 (mean 1.6), very close to the discriminant threshold.

4.5.2 Sensitivity analysis

We proceeded by performing a sensitivity analysis of the model using the partial rank correlation coefficient (PRCC) between each parameter (we included in this analysis only the parameters that are expected to be directly influenced by vector control measures) and the total number of adult females (total of Equations 4.6) obtained by running the model for $t = 600$ days. It would be quite laborious to perform and inspect a graph for each of the stages, consequently we choose to use the total number of adult females for the analysis given they are the main target of vector control interventions. The PRCC was performed separately for the simulations that converged to extinction and to equilibrium.

Figure 4.2, that shows the PRCC results, shows that the parameters that have the biggest impact in regulating the size of the adult female population are the survival in larval and pupal stages followed by the survival while non host and host-seeking. The parameters that are implicated in density dependence in the larval stages (c, ω and γ) have little impact in the adult female populations that eventually declined towards extinction but seem very important in regulating the populations that achieved a positive equilibrium value.

As reviewed in the introduction to this chapter, the impact of larviciding interventions has recently regained interest and we contribute with some insights with the results of this sensitivity analysis. Nonetheless the outcomes of sensitivity analysis can be influenced by the method used, in general is considered that no one method is clearly best, and the use of two or more methods is good practice, preferably with dissimilar theoretical foundations. There are some caveats to the use of PRCC, a disadvantage is that it is a local technique so that it will not be able to allocate the output variance to the variance in the inputs, however, the strongest concern is that the nonlinear features of the model imply that interactions between parameters are not account for in the PRCC. The most advisable method in this case would be the extended Fourier Amplitude Sensitivity Test (eFAST) method, which is superior to local sensitivity analysis and works for both monotonic and non-monotonic models [178], however in order to implement this method in R software a redesign of the model code would be necessary and we decided not to pursue it at this stage but to consider it as further work. Consequently the results of this section are to be considered as a preliminary analysis.

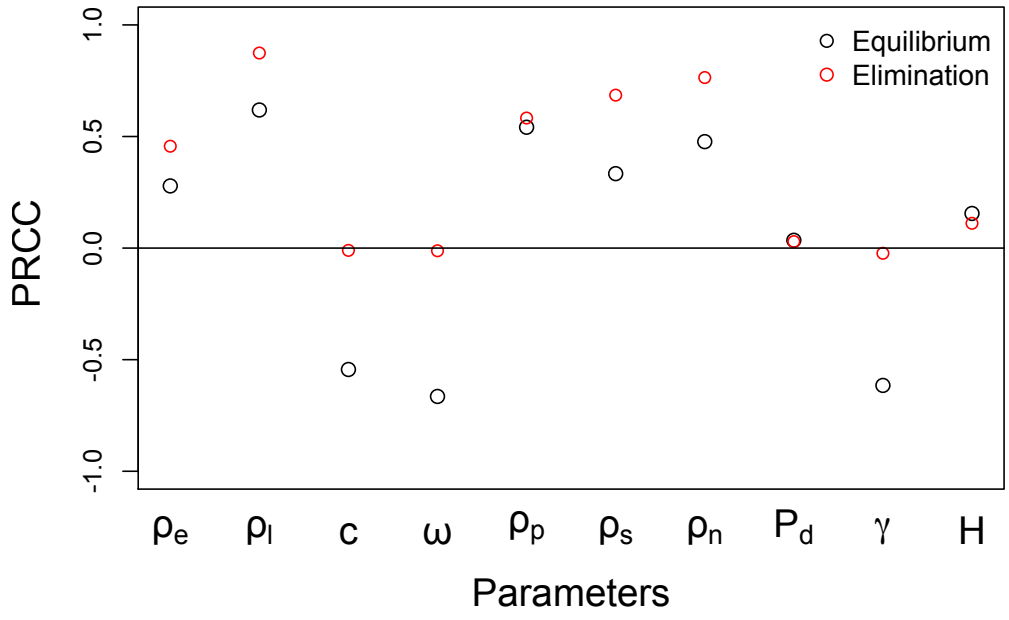

Figure 4.2: Plot of Partial Rank Correlation Coefficients between some parameters of the model and the number of adult females at day 600 (above zero increasing positive correlation, below zero increasing negative correlation). The black circles refer to simulations that resulted in a viable population and the red circles to simulations that resulted in the extinction of the population. The parameters symbols in the x-axis are defined in Table 4.1, however, in this analysis we did not distinguish between genotypes. We assume there is no insecticide resistance in the population.

4.6 Scenarios of insecticide deployment

After analysis of the dynamics of the model we can finally proceed to the section of actually asking questions of the model. The aim of this section is to explore generic issues concerning the application of insecticide rather than to parametrize to a particular setting because unfortunately, data on many of the parameters used are very limited, particularly data on the differential survival of the different genotypes when resistance is present. We found only a few values in the literature for some of the survival rates (without considering resistance), for *Anopheles gambiae*. Under laboratory conditions Olayemi and Ande (2009) [177] reported the following daily survival probabilities in the four larval instars: $0.84 \pm 0.09, 0.91 \pm 0.15, 0.82 \pm 0.06$ and 0.87 ± 0.09 and in the pupal stage: 0.76 ± 0.21. An estimate of daily survival probability of adults conducted in the field in Kilomero, Tanzania, in a capture-recapture experiment was of 0.78 [179].

Life table analysis of mosquito species are challenging. In the laboratory environmental conditions change between experiments generating different estimates [177]. In field settings many factors such as natural heterogeneities on rainfall, wind speed, desiccation, temperature, relative humidity, density-dependence on vector abundance and difficulties in adequate sampling procedures complicate the estimations.

Considering these confounding factors, we arbitrarily choose to use one of the random generated combination of parameters for the exploration of some scenarios of insecticide deployment and insecticide resistance: $\rho_e = 0.72$, $\rho_l = 0.94$, $\rho_p = 0.55$, $P_d = 0.5$, $\sigma = 0.5$, $h = 0.67$, $\rho_n = 0.96$, $\rho_s = 0.71$, $c = 0.67$, $\omega = 0.66$, $\gamma = 4 \times 10^{-8}$, $\varphi = 0.5$, $b = 100$. This combination of parameters results in a viable population in the absence of resistance. The mean of each parameter in the remaining simulations that resulted in viable populations were: $\rho_e = 0.76; \rho_l = 0.88; c = 0.50; \omega = 0.45; \rho_p = 0.76; \rho_s = 0.78; \rho_n = 0.81; P_d = 0.61; \gamma = 4.9 \times 10^{-7}; H = 0.62$). In all simulations we assume that immigration and mutation do not occur, so $\mu = 0$ and $\lambda = 0$.

We calculated a population reproductive number of 7.97, for the chosen parametrisation, and equilibrium was reached at day 240, with a starting population of 1000 individuals in each stage (hence the value of 27,000 adult female individuals at time zero in all of the simulations/scenarios presented below).

We proceed by showing simulations that mimic insecticide based interventions, followed by our interpretation of the results. Here larvicides, pupacides, ITNs and IRS are assumed to have just one effect on the mosquito population, killing. ITNs and IRS are, however, known to contribute with 2 additional effects: repelling and possibly diverting mosquitoes to an alternative hosts due to either insecticide irritation or the physical barrier of the net and lengthening the duration of the gonotrophic cycle leading to a reduced oviposition rate [143].

4.6.1 Single interventions

We start by simulating the dynamics of the mosquito population under reduced survival imposed by the use of insecticides without the emergence of resistance. We assume that ITNs act by decreasing the survival probability of female adults while host-seeking (ρ_s) and IRS by influencing the female adult survival while resting (ρ_n). Larvicides and pupacides reduce the survival probabilities ρ_l and ρ_p, respectively. As some male mosquitoes might contact with insecticide in sprayed house walls we consider that IRS also affects P_d, the proportion of male adults that survive per day. Henceforth we will be using the intervention name and the parameter that we assume it affects interchangeably.

Figure 4.3 shows the impact on adult female population of reducing the values of these parameters to mimic the additional mortality imposed by the use of insecticide at any of these stages, for now independently. We included in each graph a red horizontal line to mark the level of mosquito population that would be necessary to interrupt transmission of malaria in this setting, based on the theoretical formula for the reproductive rate of malaria R_{0m}.

The parameter values used in the calculation of R_{0m} and consequently of M (the threshold mosquito population for control of disease) are shown in Table 4.4, these values were chosen to mimic an infection by *Plasmodium falciparum* carried by *Anopheles gambiae* in an adult.

Parameter m was derived by using $M = 38,000$, which is the equilibrium number of host-seeking mosquitoes per day in Chitnis *et al.* (2008) model, assuming $N = 1,000$ as the human population (as in previous models [158]). We also estimated the value of 113 from the capture-recapture trial in [179]. In this field trial the highest value of mosquitoes found per day inside an house was 900 so if we assume that 8 is the average number of humans per house we get $m = \frac{900}{8} = 113$. To calculate the threshold value for this setting we first determined the total adult female population in equilibrium under the combination of parameters shown above, $A_f = 135,878$. Then to determine the number of humans for this seeting assuming $m = 38$: $m = \frac{M}{N} \equiv N = \frac{135,878}{38} = 3,576$. Therefore the mosquito population, M, that satisfies $R_{0m} < 1 \equiv M < \frac{0.01 \times 0.1}{0.5^2 \times 0.5 \times 0.15} \times 3,576 < 191$. For $m = 113$, a similar calculation gives $M < 65$.

Both of these M thresholds are very small, that makes it very difficult to discern under the graphs scale. It implies that in order to control malaria with this combination of parameters the population needs to be lowered to boundaries close to extinction. Again, the parameter values used are prone to discussion, this quantities are very difficult to measure in the field and are setting specific. The same caveats apply with the equation chosen to calculate the malaria reproductive rate, under the Ross-Macdonald derivation, but it serves the purpose of exemplification of what can be achieved by

Table 4.4: Parameter values used for the calculation of R_{0m} and derivation of a population threshold level that interrupts malaria transmission (M).

Parameter	Value	Reference
m	38 and 113	[122, 179]
a	0.5	[147, 180]
b_1	0.5	[180]
b_2	0.15	[180]
r	0.01	[147, 180]
g	0.1	[147]

the use of this kind of model in specific settings for which there might be information available.

The graphs in Figure 4.3 were created by independently decreasing the survival probabilities of the original parameterisation, by different factors: -10, -30, -40 and -80%. The first observation is that it does not seem plausible to decrease the female adult population by targeting only the male population, in line with what would be intuitively expected.

The female adult population collapsed with a 30% reduction of the larval daily survival (from 0.94 to 0.66), achieving extinction with the smallest reduction in survival from all the scenarios shown. This parameter incorporates only the density-independent fraction of mortality that occurs at this stage, the overall mortality may vary because of adjustments due to density-dependence. The presence of competition is expected to diminish larval survival, that we accounted for by incorporating density dependence.

Density-dependence is important in order not to ignore these endogenous processes, otherwise we could overestimate predictions of impact of exogenous interventions [181]. Although we included density dependence in our model, we do not explicitly explore its impact in the simulations presented here. We could for example have started the population above the carrying capacity and checked the effects of resistance on competition, investigating the suspected fitness costs associated with resistance. What is known is that density-dependence additionally decreases survivorship and in *A. gambiae* also influences the development time, fecundity and contributes to the emergence of smaller adults [182], which can influence adult survival. A further impact of density-dependence on control strategies is that at lower densities the fewer individuals may compensate by optimising traits that maximise their reproduction and survival reducing the efficacy of the interventions [182].

This model does not account for heterogeneities between larval sites, which can contribute with different degrees of density-dependence. We also did not account for potential differences in regulation in the four larval instars, it seems that the most important are the interactions between rather than within instars [183], we do not know

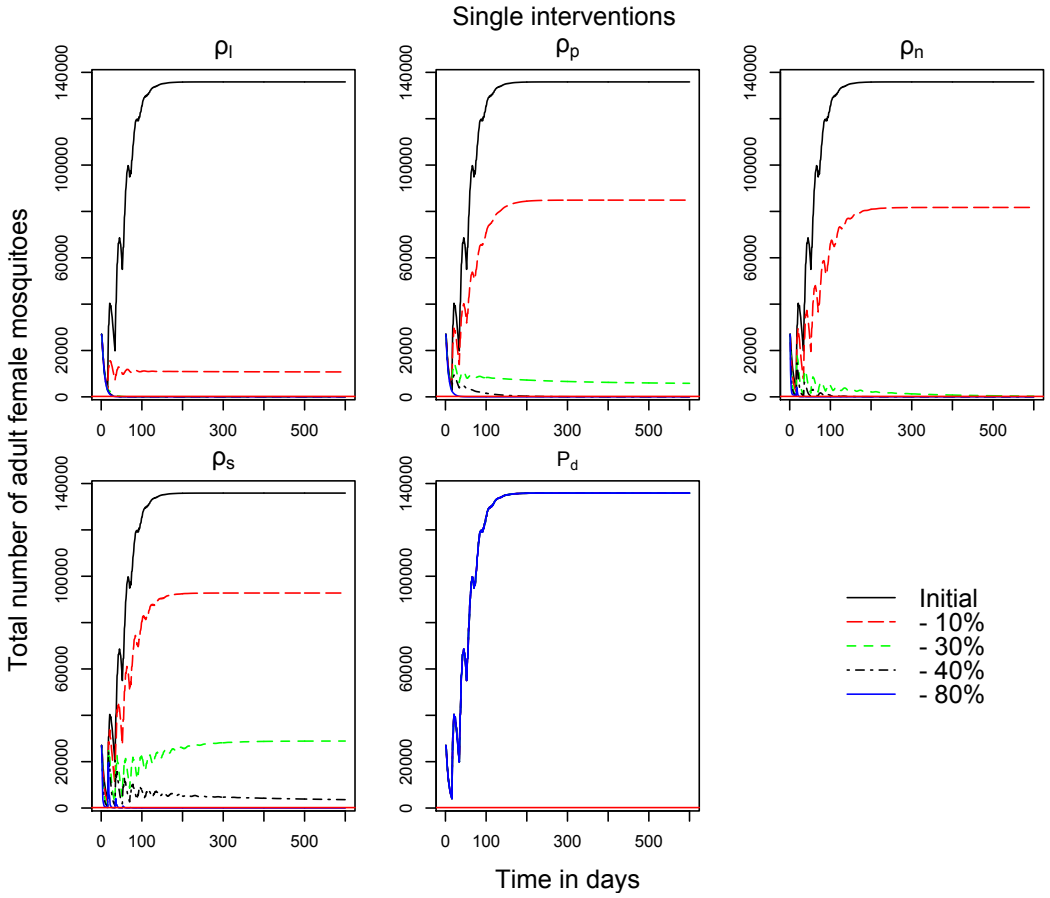

Figure 4.3: Simulations of the impact on female adult population of independently decreasing survival at different stages: larval (ρ_l), pupal (ρ_p), adult females host-seeking (ρ_s), adult females resting (ρ_n) and adult males (P_d). These parameters can potentially be affected by insecticidal interventions aimed at keeping the mosquito population under control to minimise malaria transmission. The legend shows the percentage of decreased survival imposed independently in each parameter (intervention) by different factors: -10, -30, -40 and -80%. The red horizontal line ($y = 191$) is the threshold number of adult females below which malaria transmission would be interrupted. We assume no insecticide resistance.

to what extent this could affect the breeding sites yield.

These results agree with the belief that larval survival has a great impact on the adult population density, although as pointed by [143], it does not directly kill adult mosquitoes that are potentially infectious, having a smaller impact on disease transmission. It is obvious from the point of view of disease transmission to target mosquitoes that are host-seeking, because although it makes little difference to overall mosquito survival, killing adult females before or after feeding makes a substantial difference to malaria transmission [122].

The survival proportion of pupae that was able to sustain the population in this parametrisation was 0.55, the average of all the other positive equilibrium simulations (0.76) was slightly higher. From the results on Figure 4.3 the survival at this stage would have to be lowered to 0.33 (40% of the original parameter) in order to bring the population to extinction. We modelled pupae independently from larvae, because pupae do not feed and therefore are not prone to density-dependence regulation phenomena, but in practice both stages share the same physical space. Interventions such as larviciding usually target both and pupacides alone are not used very often. The scenario of decreasing survival only at the pupal stage, that we show here is therefore very unlikely to be used in the field. It serves to show that theoretically we would have to kill more pupae than larvae to achieve the same level of reduction of individuals in the adult stage.

Adult females seeking a host for a blood meal are the main target of ITNs, one of the widespread malaria interventions, based on these results it would be necessary to decrease survival on the seeking stage (ρ_s) to values as low as 0.14 (80% of the original parameter 0.96) in order to eliminate the population (even though 40% would bring population to an equilibrium at low levels). Extinction was achieved by targeting the resting stage only (ρ_n) with a more modest decrease of 30% (0.96 to 0.67). According to these results the most effective control with the less amount of effort would be achieved by targeting the larval and the resting stage, considering only the killing effect. Another aspect worth noting is that extinction was achieved faster by targeting the larval stages.

We investigated how much the emergence and spread of resistance would impact these results on Figure 4.4. We do not further explore the exclusive targeting of males, as it does not seem relevant. Each graph shows the level of resistance needed to counterbalance the effect of the insecticide, that was previously found to be efficient at declining the mosquito population towards extinction (see Figure 4.3). We present here the worst case scenarios in terms of spread of resistance, because we assume resistance to be complete dominant, i.e, we assume the survival probabilities of the heterozygote and homozygote resistant genotypes to be equal.

There is no consensus on the impact of resistance to ITNs and IRS efficacy in the field.

There are mixed opinions about the association between *kdr* mutation and pyrethroid and/or DDT-resistance phenotype in *Anopheles gambiae* [184], and the role of additional resistance mechanisms, such as metabolic resistance. If resistance is of multigenic nature and the *kdr* genotype does not fully account for all the variance in phenotype, this could explain why attempts to infer the impact of *kdr* on the efficacy of ITNs have had contradictory results. Trials in the north of Côte d'Ivoire have failed to show decrease efficacy of personal protection offered by ITNs in the presence of 50% to 60% *kdr* frequency [185]. On the other hand, hut trials in Benin have provided evidence that high *kdr* frequencies ($> 80\%$) conferred pyrethroid resistance indeed capable of undermining control measures based on ITNs [186]. Nevertheless, at present, *kdr* screening seems to be the best molecular diagnostic tool for predicting pyrethroid and DDT efficacy [184].

We can only state based on our results that: i) even a small increase in survival (10% to 40%), that we assume here to be due to resistance, brings a population back to levels that sustain transmission; ii) the resistance level necessary to restore the mosquito population is higher in the host-seeking stage, although the required level of control is very high; iii) spread of resistance is faster in the larval stage (larvicides kill mosquitoes before they reproduce, so selection for resistance is strong [86]).

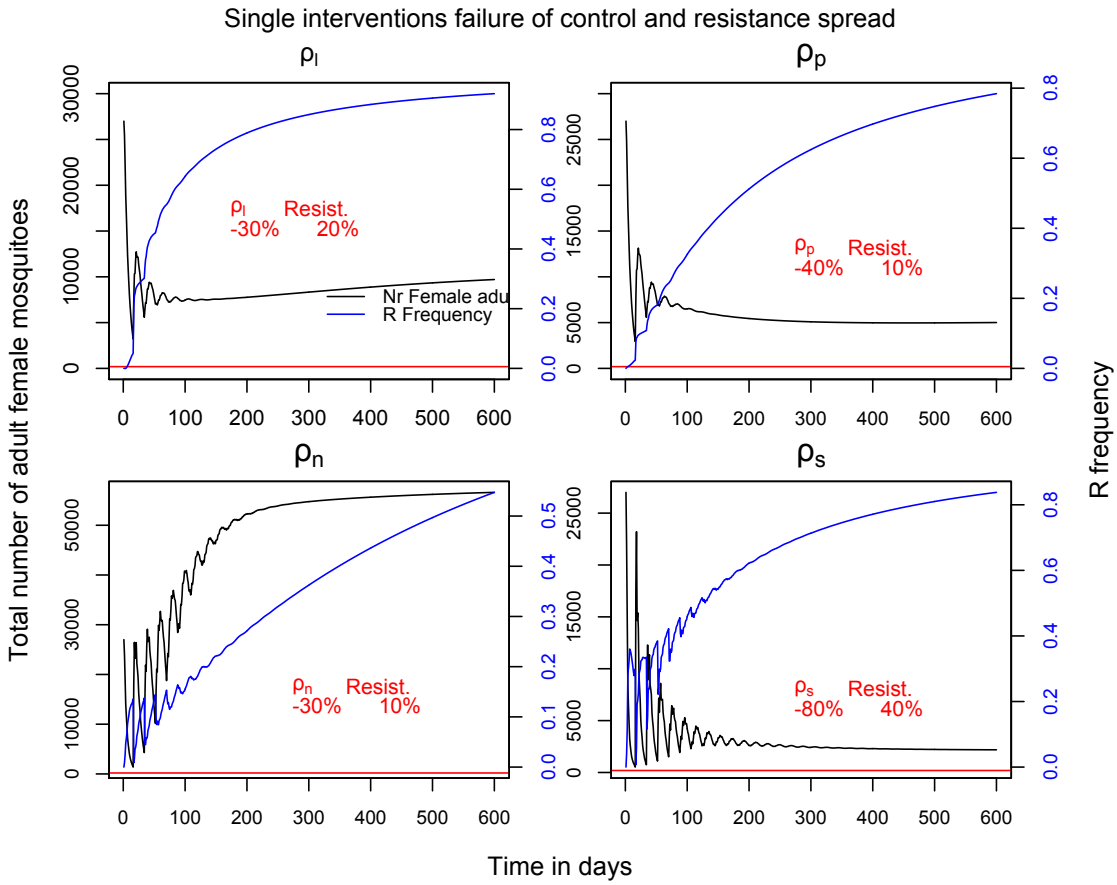

Figure 4.4: Simulations illustrating the level of resistance required to induce failure of control interventions, aimed at each of the stages independently, that would otherwise decrease the female adult population to zero (see Figure 4.3). Legend in red displays the imposed percentage of decreased survival (ρ_{stage}) and resistance (Resist.) used to generate the graphs. Blue line shows the resistant allele frequency trend in time and black line shows the number of female adults mosquitoes trend. The red horizontal line ($y = 191$) is the threshold number of total female adults below which theoretically malaria transmission would be interrupted.

4.6.2 Combined interventions

In many field settings these interventions are used in combination, we explore the effects of such strategies in Figure 4.5.

The use of IRS and ITNs in combination is thought to offer the advantage of increasing the probability of a mosquito meeting an intervention, contributing for reaching and maintaining high coverage that is usually very difficult to attain [187, 188]. Nevertheless, there is yet to be found indisputable empirical evidence that ITNs-IRS combinations can indeed offer any additional communal or personal protection, compared to using either method alone, because there have not been many studies designed specifically to test the two vector control methods in combination [189]. A recent one, conducted by Corbel *et al.* (2012) [187] was to a great extent inconclusive but discouraging on the use of combinations of interventions. Some researchers [190] discussed these results, advocating that initiatives to deploy IRS for malaria transmission control in endemic countries should continue and if anything Corbel and colleagues study draws attention to the need to improve and optimise operational factors.

We mimic the combined use of IRS and ITNs by simultaneously changing survival of parameters ρ_n and ρ_s. Our simulations reached the threshold for malaria transmission interruption with a combination of decreased survival of -10% ($\rho_n = 0.87$) in the non-seeking stage and -30% ($\rho_s = 0.5$) in the seeking stage, this survival proportion is much more feasible than the 0.14 when considering a single intervention targeting the host-seeking stage alone.

There are many confounding effects that can affect the results of this type of analysis, that should be considered: the type of insecticides used, because the killing effect is not always the most significant factor and insecticides can interfere or synergise with each other; insecticide persistence; the coverage with a second intervention being at the household or community level; the degree of endemicity; and behaviour of the vector (degree of endophily); among others (see [191] for a comprehensive review on this subject).

Our analysis does not takes in account many of these factors so what we can derive from our results is that the levels of insecticide necessary would be reduced when used in combination. This has been concluded before, when retrospectively analysing non-experimental data from Solomon Islands. It was established that house spraying (with DDT) was more effective than ITNs but that the amount of the insecticide required would be reduced if ITNs were also used [192].

The same study was not able to associate reduction in malaria cases with larviciding (with temephos) in combination with other interventions, while our results suggest that larviciding would be more effective in reducing the mosquito population in combination with IRS or ITNs than IRS-ITNs. These comparisons should be taken with caution because our results do not translate directly into malaria cases such as surveyed in

the mentioned study. Similarly to our results, White *et al.* (2011) [143], modelling the density of mosquitoes, found that larviciding when used in combination with high levels of ITNs coverage was as effective or superior as using IRS.

We finish by exploring the effects of resistance in such combination of interventions. We assume that the mode of action of the insecticides is different in each intervention (different resistance mechanisms) so that we screen another advantage of using interventions in combinations: the possibility of delaying the spread of resistance to each of the insecticides used.

Figure 4.6 shows the impact of resistance in adult females numbers and the associated spread of resistance. For each combination of interventions two graphs are shown, each displaying the impact of resistance in each parameter at a time, shown in the legend in red.

In all of our simulations, just 10% of resistance was able to bring population to numbers capable of sustaining transmission, according to our population threshold, although, the speed of spread of resistance and the population numbers changed between combinations. Resistance seems to spread much faster if it occurs in the interventions targeting the larval stages (as seen in the scenarios simulating single larviciding interventions) and slower if targeting the seeking stage. There is, as far we know, no literature that has investigated directly the impact on resistance spread of the use of different combinations of interventions.

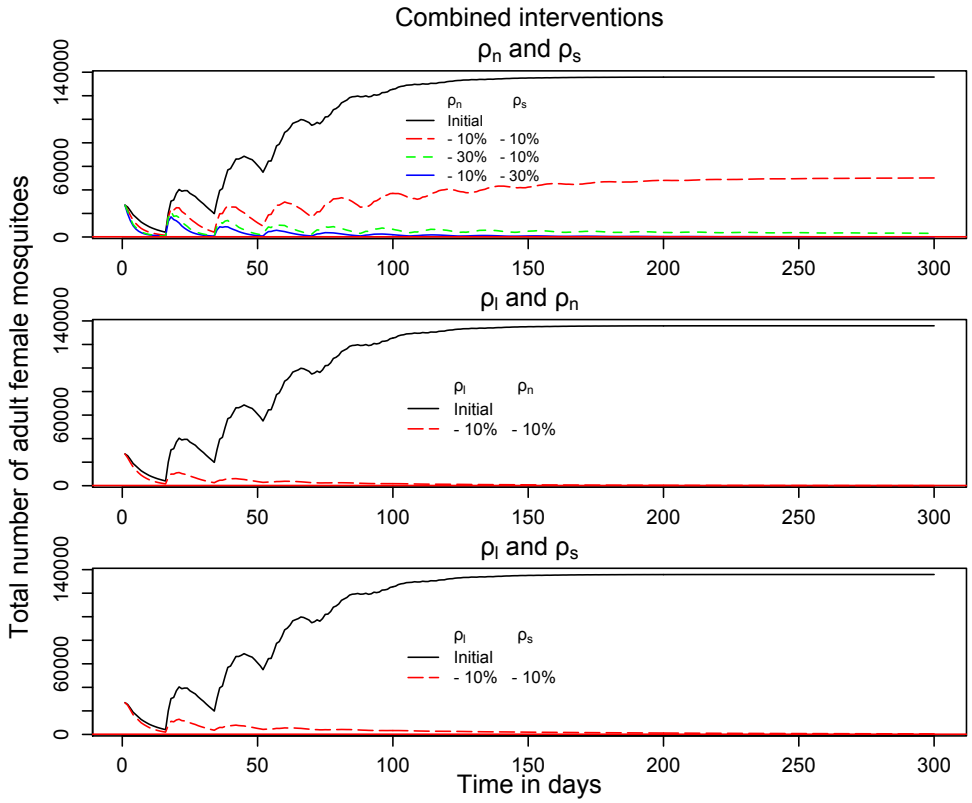

Figure 4.5: Simulations of the impact on female adult population of decreasing survival at different stages in combination: i) adult females host-seeking (ρ_s) and adult females resting (ρ_n),ITNs-IRS; ii) larval (ρ_l) and adult females resting (ρ_n),larviciding-IRS; iii) larval (ρ_l) and adult females host-seeking (ρ_n),larviciding-ITNs. The legend shows the percentage of decreased survival imposed in each parameter (intervention). The red horizontal line ($y = 191$) is the threshold number of total female adults below which theoretically malaria transmission would be interrupted. We assume no insecticide resistance.

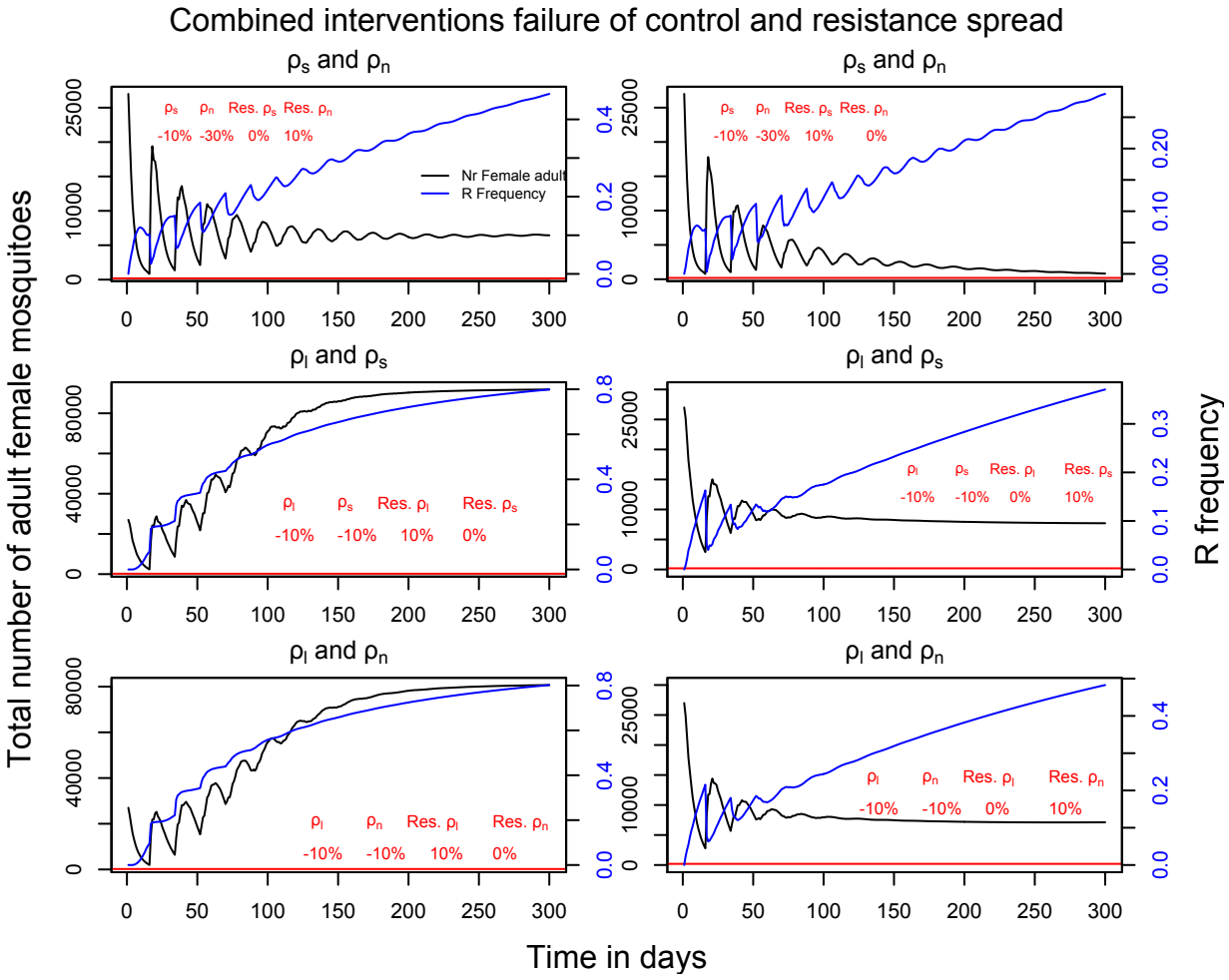

Figure 4.6: Simulations illustrating the level of resistance required to induce failure of control of interventions used in combination. The level of control (reduced survival) presented, would in the absence of resistance decrease the female adult population to zero, as shown in Figure 4.5. The legend shows the percentage of decreased survival and resistance imposed in each parameter (intervention).The red horizontal line ($y = 191$) is the threshold number of total female adults below which theoretically malaria transmission would be interrupted.

4.7 Conclusion

We present here a model for the mosquito life cycle that explores the effects of insecticide application on mosquito density, and potentially on malaria transmission, that at the same time can track the spread of insecticide resistance. It is a stage-structured model, that considers the metamorphic nature of mosquito populations and improves on other types of models by assuming that population size is not constant and depends on numbers of adults (and keeps track of the mating genotypes). The adult production is expressed in terms of not only time to adult but also numbers surviving to adulthood. We include endogenous population regulation by including a density-dependence mechanism in the larval stage. The modelling of the adult stage allows for mortality rates to differ over the feeding, digesting/ovipositing stages of the cycle, and to be significantly different between sexes. Including this compartmentalisation has lead to parameter calibration problems because, as far as we know, there are no estimates of mortality rates at these stages, much less by sex and genotype, which is not surprising considering the challenges of estimating survival rates in the field.

This new mathematical model still greatly simplifies the complex dynamics of mosquito life cycle and includes several assumptions. We do not account for differences in development times in the juvenile stages between males and females, however, it would be easy to adapt the model to include such differences by changing the θ parameters accordingly. We assume all females mate and that it occurs straight after emergence, which may contribute to some overestimation of female adult survival if they spend time searching for a mate.

We modelled the population using a deterministic model, however, real populations never exhibit deterministic population growth, i.e., that have a constant rate of growth in time, such as we presented above. A typical complicating factor is the exclusion of seasonality, by assuming constant per capita mortality for mosquitoes, ignoring important factors such as temperature-dependent mortality and spatial heterogeneities. The effects of seasonality could be incorporated by, for example, changing the resource availability (γ), the number of days of development in the immature stages (θ) and changing the survival probabilities in each stage at periodic intervals. We also did not consider age dependent effects on mortality [193].

The system of difference equations that we developed was too complex to enable analytical analysis and standard matrix populations modelling so that the approach used for the analysis of the dynamics had to rely on numerical simulation. We executed the model for a set of collections of parameters to observe the resulting change in model behaviour (number of individuals in each stage). Central to population dynamics analysis are the concepts of equilibrium and stability and we were able to verify that the model can show two fixed points, one trivial (extinction) and one positive (equilibrium

population size and structure) and that both are stable, although we were only able to show the local stability of the positive fixed point.

The vector reproductive rate that we developed, although not 100% accurate, can be use for rapid screening of the expected trend of a population with given characteristics (parameterisation) without running the full model.

Many strategies to control malaria and other vector borne disease use insecticides against mosquitoes at various stages so control is undermined by the continual evolution of resistant mosquitoes. Different scenarios of insecticide deployment were explored by changing some key parameters in the model that we developed. The analysis we performed explore the comparative impact of ITNs, IRS, larvicides and pupacides, and the benefits that can be achieved by combining these interventions.

It is also possible to check the effects of resistance in the effectiveness of interventions. We assume that the success of each intervention is measured in terms of its capacity to decrease the number of female adult mosquitoes, although interventions directed against female adult mosquitoes will have the added benefit of reducing the probability that an infected mosquito survives sporogony to become infectious, as mentioned previously.

We do not account for transient effects such as the decay of nets and insecticide and have not yet investigated the effects of migration and mutation that we incorporated in the system of equations, that would require other set of specific simulations.

We linked our demographic model with malaria transmission by calculating a threshold value for the female mosquito population size that would interrupt transmission. It is based on a theoretical quantity, the reproductive value of malaria R_{0m} derived by Ross-Macdonald. The classic derivation we used is known to have its shortcomings. It describes idealised populations, where each infectious bite lands on a different host human population, assumed to be infinite, and that humans are bitten at the same rate, which is clearly not the case [86]. It does not include the effects of acquired immunity, that reduces the infectivity of humans to mosquitoes and decreases human susceptibility to infection [193].

We recognise that it would have been useful to conduct the analysis on a set of collection of parameters (instead of only one), in this way incorporating variance between different hypothetical settings, and therefore gaining more confidence that the results provide a general picture. Additionally, given that the effectiveness of control methods depends on the context of local transmission, it is obviously necessary to calibrate the parameters against field data (on mosquito population and malaria transmission) before any of these results could be used as sound evidence for enforcing any malaria control strategy. Our results are therefore preliminary and serve the purpose of demonstrating the potential of our model.

There is renewed interest in larval mosquito dynamics and opinion is mounting that

the goals of programmes such as Roll Back Malaria can best be met by an integrated approach combining disease treatment and interventions against both adult and larval stages of the vector [194]. If we consider our results, in terms of decreasing the numbers of adult females, the use of larviciding seems a valid option either alone or in combination.

The effects of resistance on the failure of the control are severe, given that only 10% of resistance appears to be able to jeopardise control measures in all interventions combinations. The spread of resistance, however, seems to occur faster if insecticide targets the larval stages rather than the adult stages. This is expected because insecticides have a bigger impact on larvae, so selection for resistance may be higher.

So far we have focused on interventions that are currently in widespread use by national malaria-control programs but the model we constructed is flexible and can be used to compare the effects of novel interventions such as the use of entomopathogenic fungi, transgenically modified mosquitoes and the effects of behaviour avoidance by some species that change to bite earlier in the evening when people are less likely to be shielded by bed nets.

We also considered only one species while there are nearly 70 species of *Anopheles* that transmit malaria, of which some have overlapping geographic distribution. We could use this model to examine the interaction of more than one species in a transmission setting. We also make no assumptions regarding the type of host on which mosquitoes blood feed. The model can be used to explore the effects of mosquitoes feeding on alternative hosts with different survival patterns.

The fact that we can control the length of the different stages, and the survival probabilities in each of these stages allows for other types of studies, such as changes in life-history characteristics (evolutionary strategies). For example, females balance the risk of increasing probability of mortality to herself while she searches for an oviposition site with the increased probability of finding a high quality site. Such oviposition site selection may affect adult population size, because of the inherent effects of density-dependence, competition and predation. The optimal tradeoff between survival as adults and the characteristics of the breeding site could be explored using the model we developed.

We present here a stand alone model that co-investigates both mosquito demography and the genetics of resistance. We proceeded by exploring some scenarios where it can be used, with some suggestions for parallel studies. Nonetheless, it will be most useful, in the context of malaria transmission, when used in conjunction with models that simulate the dynamics of malaria in the mosquito vector and explore variations among humans in their exposure to mosquitoes, and in their responses to the parasite. This is the next step that will have to be accomplished in future work.

Chapter 5

Final remarks and conclusions

Models play a central role in science, whether they are empirical, such as animal models used for human diseases research, or formal, as the ones presented here. The word 'model' is used mostly to refer to formal modelling, a simple symbolic representation that shares structural properties with the real system. Models of this kind are sometimes confused with theories, which are explanations about some real world property, that can be use to predict events outcomes and that were thoroughly tested, and therefore are endorsed by consistent evidence. These theories provide a framework from which simplified versions, models, are constructed, that may incorporate only a limited amount of the theory components. There are many different manners by which theories and models can be expressed, in the case of the models developed here we used the language of mathematics and computation in their construction, using proven modelling and statistical tools, based on population genetics and ecology theories.

All empirical work is grounded in some abstraction of the real world and therefore constitutes a model in itself, something not always acknowledged by empiricists. Mathematical modelling is sometimes disregarded among empiricist due (in part) to the misconception that mathematical models make so many more assumptions and therefore reduce reality to a larger extent than experimental work. If anything, mathematical modelling is just more clear about assumptions made, most likely because it deals in the realm of formalised and exact mathematics.

Another aspect that contributes to mathematical modelling sometimes being considered a lesser scientific methodology is the perception that developing and analysing a model is a straightforward task, easy and fast. The process of building, studying, testing and using a model for the understanding of a biological process is not free from setbacks, there is a considerable amount of trial and error, time spent learning techniques and implementing and debugging code. Modelling has a clear advantage over experimental approaches in cases where it is not feasible to conduct experiments for financial or ethical reasons.

Even though we used mathematics and computations as tools, the emphasis of this

work was on the design of models that are useful without being unnecessarily complex, and on their interpretation in terms of biological conclusions. Within each chapter we have discussed our major findings, and their significance and limitations. Here we briefly summarise our results.

In chapter 2 the aim was to develop a methodology to quantify the strength of selection acting in the field, that could simultaneously provide an estimate of the dominance relationship and initial resistance allele frequency. We developed what is in fact an empirical model, because the cornerstone of the method are the data and we do not take into account the mechanism by which the changes in resistance allele frequencies occur.

We first showed that the maximum likelihood methodology (ML) was accurate with idealised simulated data. Subsequent analysis with suboptimal simulated data, explored the important effects of sample size and dominance on estimations of the genetic parameters. We made the case that it is not possible to establish the trend of spread and quantify strength of resistance when sampling occurs only in the early stages or close to fixation of a resistance allele. This is a common observation in field datasets, in which initially the data shows negligible or low frequencies of resistance and at later time points very high frequencies. We showed how difficult it would be to achieve estimates of dominance without information about intermediate time points, since many trajectories may provided similar fits particularly if the allele is semi-recessive.

We applied our method to a field dataset that we believe reflects the current level of information regarding insecticide resistance surveillance. In order to achieved a moderate sample size to deal with the issues described before we pooled data from different locations. However, we also demonstrated that it is erroneous to gather data from different areas using our method. The ML methodology was not able to correctly estimate the genetic parameters using simulated pooled data with spatially variable parameter values. Additionally, we explored the effects of temporal changes in the parameters, and concluded that no accurate estimations could be derived if there are, for example, modifications of the pattern of applications of insecticides, which will have a great impact on the dominance factor.

With these caveats regarding the nature of datasets, the results we obtained using the field dataset are to be considered with caution. The most relevant result points towards incomplete dominance in the field, higher than expected from laboratory works. In retrospect, overall these conclusions are not surprising, however we demonstrated by simulations of plausible datasets just how non-robust these estimates can be. These results reinforce the need of implementation of insecticide resistance surveillance systems in areas of vector control. Without it, further understanding of insecticide resistance spread is limited and measures to counteract it cannot be put in place.

In chapter 3 we examined the contribution of a prototype of a bed net, designed for use in areas with (pyrethroid) resistant malaria vectors, which has been marketed as a

tool to delay the spread of insecticide resistance. We considered it important to explore this question in the context in which these nets are actually deployed. A mosquito population is likely to encounter insecticides in more than one area (we considered only the use of one class of insecticides), and enjoy some areas free of control measures. In this way we included the important impact of agriculture in the development of insecticide resistance.

Using a genetic model we concluded that important parameters for vector control are the proportion of mosquitoes that encounter insecticides and also the fitness scaling factors that define survival of susceptible mosquitoes when meeting insecticides. These scaling factors and presence of numerous environmental niches were novel features of the model. The spread of resistance is determined by the selection and dominance coefficients. We showed that indeed the bed net is capable of delaying the spread of resistance, but to a surprisingly small extent. On the other hand, and most interestingly, we showed that in some particular circumstances the existence of a bed net with the characteristics of this new prototype can intensify the spread of resistance. It was not an aim of this work to consider how this level of delayed frequency would translate into a significant decrease in disease transmission. We additionally demonstrated the possibility of resistance alleles reaching an equilibrium frequency in such an heterogeneous environment, something not usually expected in the absence of heterozygote advantage.

In chapter 4 we developed a model that allows the exploration of the effects of control measures on the actual number of mosquitoes (and trends in the population size), since the rational behind vector control is the reduction of the mosquito population that will translate into a reduction in disease transmission. We contribute further understanding to this area of research with the inclusion of our stage-structured model of genotype differentiation. This allows the tracking of the resistance status of the population and the analysis of the tradeoff effects of a particular intervention (or combination of interventions) in reducing population size and the spread of insecticide resistance that such intervention(s) induce.

We performed a preliminary analysis based on a hypothetical setting (since field estimates were not available for most parameters) on the impact of insecticide treated nets, indoor residual spraying, larvicides and pupacides. We changed key parameters in the model to evaluate the benefits that can be achieved by using these intervention alone and in combinations and the intensity of resistance selection that arises from different amounts and patterns of insecticide usage. We showed that targeting the larval stages has the greatest effect on the adult population size followed by targeting of the non host-seeking females. Our results suggest that low levels of resistance can induce failure of interventions, and the rate of spread of resistance is higher when insecticide targets the larval stages and lower when directed to the adult female host-seeking stages. From

all the three models developed in this work, this is the one with more potential for future work. It is presented here methodologically verified but not yet validated against a real field setting, ready to be used in further explorations of strategies of control, insecticide resistance and/or other related subjects.

Final remarks

In this work we designed, implemented and analysed three models that were developed in a logical progression: we first analysed field data on insecticide resistance to obtain estimates of significant genetic parameters, followed by the study of the spread of insecticide resistance in an heterogeneous environment, ignoring changes in mosquito population, and finally combining resistance and demographic impact in the population.

The main aspect that became evident is the lack of field data for calibration of models, which reflects the scattered way in which research has been conducted and the complexity of the issue in hands. It is evident that collection of field data is not easy, and it is not our intention to criticise the community for poor field estimates, but it is necessary to conduct field trials with a view to maximise information (as an example: after decades of collecting mosquitoes inside huts, there is nothing but educated guesses about the proportion of male mosquitoes that enter human dwellings in many important anopheline species). Mathematical models are useful approximations that need to be calibrated for real settings, we could only compare our results with our perception of the real world, which can be misleading, and with results of previous research, which most of the times are not directly comparable. We expect to have a positive role in this field by having identified the type of data that needs to be collected and, as in the second chapter, how surveillance systems may be implemented in order to obtain these data.

Malaria elimination is an ambitious goal that has been brought back to the global agenda, and mathematical modelling has been identified by the malaria eradication research agenda initiative (malERA) [195] as an indispensable part of the multidisciplinary research process that needs to be put in place. Ideally, modellers, other scientists and policy makers will be working together for optimal use of resources.

Bibliography

[1] Hinman AR: **1889 to 1989: a century of health and disease**. *Public Health Rep* 1990, **105**(4):374–380.

[2] Labbé P, Alout H, Djogbénou L, Pasteur N, Weill M: *Evolution of resistance to insecticide in disease vectors* In *Genetics and evolution of infectious diseases*. Elsevier Insights 2010.

[3] WHO: **Pesticides and their application for the control of vectors and pests of public health importance**. *World Health Organization, Geneva, Switzerland* 2006.

[4] Lemon RSM, Sparling PF, Hamburg MA, Relman DA, Choffnes ER, Mack A: *Vector-borne diseases: understanding the environmental, human health, and ecological connections, workshop summary (Forum on microbial threats)*. The National Academies Press 2008.

[5] Gratz N: **Emerging and resurging vector-borne diseases**. *Annu Rev Entomol* 1999, **44**:51–75.

[6] Zell R: **Global climate change and the emergence/re-emergence of infectious diseases**. *Int J Med Microbiol* 2004, **293**(Suppl. 37):16–26.

[7] Harrus S, Baneth G: **Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases**. *Int J Parasitol* 2005, **35**:1309–1318.

[8] Morse SS: **Factors in the emergence of infectious diseases**. *Emerg Infect Dis* 1995, **1**:7–15.

[9] Sutherst RW: **Global change and human vulnerability to vector-borne diseases**. *Clin Microbiol Rev* 2004, **17**:136–173.

[10] Rogers DJ, Randolph SE: **Climate change and vector-borne diseases**. *Adv Parasitol* 2006, **62**:345–381.

[11] van Lieshout M, Kovats R, Livermore M, Martens P: **Climate change and malaria: analysis of the SRES climate and socio-economic scenarios**. *Global Environ Chang* 2004, **14**:87–99.

[12] Taubes G: **Apocalypse Not**. *Science* 1997, **278**(5340):1004–1006.

[13] Gubler DJ: **Resurgent vector-borne diseases as a global health problem**. *Emerg Infect Dis* 1998, **4**(3):442–450.

[14] Tolle MA: **Mosquito-borne diseases**. *Curr Probl Pediatr Adolesc Health Care* 2009, **39**:97–140.

[15] Hemingway J, Ranson H: **Insecticide resistance in insect vectors of human disease**. *Annu Rev Entomol* 2000, **45**:371–391.

[16] WHO: **Vector control: Methods for use by individuals and communities / prepared by Jan A. Rozendaal.** *World Health Organization, Geneva, Switzerland* 1997.

[17] Ault SK: **Environmental managment: a re-emerging vector control strategy**. *Am J Trop Med Hg* 1994, **50**(Suppl. 6):35–49.

[18] Brewster D: **Environmental management for vector control. Is it worth a dam if it worsens malaria?** *BMJ* 1999, **319**(7211):651–652.

[19] Nam VS, Yen NT, Holynska M, Reid JW, Kay BH: **National progress in dengue vector control in Vietnam: survey for *Mesocyclops* (Copepoda), *Micronecta* (Corixidae), and fish as biological control agents**. *Am J Trop Med Hyg* 2000, **62**:5–10.

[20] Lacey LA: ***Bacillus thuringiensis* serovariety *israelensis* and *Bacillus sphaericus* for mosquito control**. *J Am Mosq Control Assoc* 2007, **23**(Suppl. 2):133–163.

[21] Dyck V, Hendrichs J, Robinson AS: *Sterile insect technique. Principles and practice in area-wide integrated pest management*. Springer 2005.

[22] Benedict MQ, Robinson AS: **The first releases of transgenic mosquitoes: an argument for the sterile insect technique**. *Trends Parasitol* 2003, **19**(8):349–355.

[23] Knols BGJ, Bossin HC, Mukabana WR, Robinson AS: **Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world**. *Am J Trop Med Hyg* 2007, **77**(Suppl. 6):232–242.

[24] Harris AF, Nimmo D, Mckemey AR, Kelly N, Scaife S, Donnelly CA, Beech C, Petrie WD, Alphey L: **Field performance of engineered male mosquitoes**. *Nat Biotechnol* 2011, **29**(11):1034–1037.

[25] Alphey L: **Malaria control with genetically manipulated insect vectors**. *Science* 2002, **298**(5591):119–121.

[26] Smith AE, Secoy DM: **Forerunners of pesticides in classical Greece and Rome**. *J Agric Food Chem* 1975, **23**(6):1050–1055.

[27] Casida J, Quistad G: **Golden age of insecticide research: past, present, or future?** *Annu Rev Entomol* 1998, **43**:1–16.

[28] Potter C: **The future of chemical control of insects**. *Ann Appl Biol* 1961, **49**:391–402.

[29] WHO: **Monitoring of insecticide resistance in malaria vectors**. *World Health Organization, Geneva, Switzerland* 2009.

[30] Bloomquist J: **Toxicology, mode of action and target site-mediated resistance to insecticides acting on chloride channels**. *Comp Biochem Physiol* 1993, **106C**(2):301–314.

[31] Aldridge W: **An assessment of the toxicological properties of pyrethroids and their neurotoxicity**. *Crit Rev Toxicol* 1990, **21**(2):89–104.

[32] Vijverberg HPM, van der Zalm JM, van den Bercken J: **Similar mode of action of pyrethroids and DDT on sodium channel gating in myelinated nerves**. *Nature* 1982, **295**:601–603.

[33] Vijverberg H, vanden Bercken J: **Neurotoxicological effects and the mode of action of pyrethroid insecticides**. *Crit Rev Toxicol* 1990, **21**(2):105–126.

[34] WHO: **World Malaria Report: 2011**. *World Health Organization, Geneva, Switzerland* 2011.

[35] Devine G, Perea E, Killeen G, Stancil J, Clark S, Morrison A: **Using adult mosquitoes to transfer insecticides to *Aedes aegypti* larval habitats**. *Proc Natl Acad Sci USA* 2009, **106**(28):11530–11534.

[36] WHO: **Global insecticide use for vector-borne disease control: a 10 -year assessment (2000-2009)**. *World Health Organization, Geneva, Switzerland* 2011.

[37] WHO: **Insecticide-treated mosquito nets: a WHO position statement**. *World Health Organization, Geneva, Switzerland* 2007.

[38] WHO: **Expert Committee on Insecticides**. *World Health Organization, Geneva, Switzerland* 1957, (125).

[39] Whalon M, Mota-Sanchez D, Hollingworth R: *Global pesticide resistance in arthropods*. CAB International, Cambridge, MA 2008.

[40] Li X, Schuler M, Berenbaum M: **Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics**. *Annu Rev Entomol* 2007, **52**:231–253.

[41] Zhu F: **Molecular mechanisms of cytochrome P450 monooxygenase-mediated pyrethroid resistance in the house fly, *Musca Domestica* (L.)**. Ph.D. Thesis. Auburn University 2007.

[42] Nauen R: **Insecticide resistance in disease vectors of public health importance**. *Pest Manag Sci* 2007, **63**(7):628–633.

[43] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, Williamson MS, Bass C: **Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid Insecticides in the aphid *Myzus persicae***. *PLoS Genetics* 2010, **6**(6).

[44] Wood O, Hanrahan S, Coetzee M, Koekemoer L, Brooke B: **Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus***. *Parasit Vectors* 2010, **3**:67.

[45] Lines JD: **Do agricultural insecticides select for insecticide resistance in mosquitoes? A look at the evidence**. *Parasitol Today* 1988, **4**(7):S17–S20.

[46] Mouchet J: **Agriculture and vector resistance**. *Insect Sci Appl* 1988, **9**(3):1–6.

[47] Overgaard H: *Malaria mosquito resistance to agricultural insecticides: risk area mapping in Thailand*. IWMI Research Report 103, International Water Management Institute. Colombo, Sri Lanka: 2006.

[48] Roush RT, Tabashnik BE: *Pesticide resistance in arthropods*. Chapman and Hall 1990.

[49] Otto SP, Day T: *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press 2007.

[50] Hoy M: **Myths, models and mitigation of resistance to pesticides**. *Philos Trans R Soc Lond B Biol Sci* 1998, **353**:1787–1795.

[51] Lord CC: **Modeling and biological control of mosquitoes**. *J Am Mosq Control Assoc* 2007, **23**(Suppl. 2):252–264.

[52] Committee on Strategies for the Management of Pesticide Resistant Pest Populations National Research Council. *Pesticide resistance: strategies and tactics for management*. The National Academies Press 1986.

[53] Georghiou GP: **Principles of insecticide resistance management**. *Phytoprotection* 1994, **75**(4):51–59.

[54] Georghiou G: **The evolution of resistance to pesticides**. *Annu Rev Ecol Evol Syst* 1972, **3**:133–168.

[55] Georghiou G, Taylor C: **Genetic and biological influences in the evolution of insecticide resistance**. *J Econ Entomol* 1977, **70**(3):319–323.

[56] Comins H: **The development of insecticide resistance in the presence of migration**. *J Theor Biol* 1977, **64**:177–197.

[57] Taylor C, Georghiou G: **Suppression of insecticide resistance by alteration of gene dominance and migration**. *J Econ Entomol* 1979, **72**:105–109.

[58] Caprio M, Tabashnik B: **Gene flow accelerates local adaptation among finite populations: simulating the evolution of insecticide resistance**. *J Econ Entomol* 1992, **85**(3):611–620.

[59] Pasteur N, Raymond M: **Insecticide resistance genes in mosquitoes: their mutations, migration, and selection in field populations**. *J Hered* 1996, **87**:444–449.

[60] Peck S: **Antibiotic and insecticide resistance modeling–is it time to start talking?** *Trends Microbiol* 2001, **9**(6):286–292.

[61] Caprio M: **Source-sink dynamics between transgenic and non-transgenic habitats and their role in the evolution of resistance**. *J Econ Entomol* 2001, **94**(3):698–705.

[62] Rosenheimi J, Tabashnik B: **Evolution of pesticide resistance: interactions between generation time and genetic, ecological, and operational factors**. *J Econ Entomol* 1990, **83**(4):1184–1193.

[63] McKenzie J, Batterham P: **Predicting insecticide resistance: mutagenesis, selection and response**. *Phil Trans R Soc B* 1998, **353**(1376):1729–1734.

[64] Mallet J: **The evolution of insecticide resistance: Have the insects won?** *Trends Ecol Evolut* 1989, **4**(11):336–340.

[65] Groeters F, Tabashnik B: **Comment: Insecticide resistance**. *Trends Ecol Evolut* 1995, **10**(4):164.

[66] Jain P, Bhargava M: *Entomology: Novel Approaches*. New India Publishing Agency 2007.

[67] Roush R, McKenzie J: **Ecological genetics of insecticide and acaricide resistance**. *Annu Rev Entomol* 1987, **32**:361–380.

[68] Ffrench-Constant R, Daborn P, Goff G: **The genetics and genomics of insecticide resistance**. *Trends in Genetics* 2004, **20**(3):163–170.

[69] McKenzie J, Batterham P: **The genetic, molecular and phenotypic consequences of selection for insecticide resistance**. *Trends Ecol Evolut* 1994, **9**(5):166–169.

[70] Groeters FR, Tabashnik BE: **Roles of selection intensity, major genes, and minor genes in evolution of insecticide resistance**. *J Econ Entomol* 2000, **93**(6):1580–1587.

[71] Firko M, Haynes J: **Quantitative genetic tools for insecticide resistance risk assessment: estimating the heritability of resistance**. *J Econ Entomol* 1990, **83**(3):647–654.

[72] Klerks P, Xie L, Levinton J: **Quantitative genetics approaches to study evolutionary processes in ecotoxicology; a perspective from research on the evolution of resistance**. *Ecotoxicology* 2011, **20**:513–523.

[73] Rosenheim J: **Realized heritability estimation for pesticide resistance traits**. *Entomol Exp Appl* 1991, **58**:93–97.

[74] Tabashnik B: **Implications of gene amplification for evolution and management of insecticide resistance**. *J Econ Entomol* 1990, **83**(4):1170–1176.

[75] Georghiou G, Taylor C: **Operational Influences in the Evolution of Insecticide Resistance**. *J Econ Entomol* 1977, **70**(5):653–658.

[76] Carriere Y, Tabashnik B: **Reversing insect adaptation to transgenic insecticidal plants**. *Proc R Soc Lond B* 2001, **268**(1475):1475–1480.

[77] Curtis C: **Theoretical models of the use of insecticide mixtures for the management of resistance**. *Bull Entomol Res* 1985, **75**(2):259–265.

[78] Tabashnik B: **Managing resistance with multiple pesticide tactics: theory, evidence, and recommendations**. *J Econ Entomol* 1989, **82**(5):1263–1269.

[79] Mani G: **Evolution of resistance in the presence of two insecticides**. *Genetics* 1985, **109**:761–783.

[80] Birch C, Shaw M: **When can reduced doses and pesticide mixtures delay the build-up of pesticide resistance? a mathematical model**. *J Appl Ecol* 1997, **34**(4):1032–1042.

[81] Mani GS: **Evolution of Resistance with Sequential Application of Insecticides in Time and Space**. *Proc R Soc Lond B* 1989, **238**(1292):245–276.

[82] Lenormand T, Raymond M: **Resistance management: the stable zone strategy**. *Proc R Soc Lond B* 1998, **265**(1409):1985–1990.

[83] WHO: **The technical basis for coordinated action against insecticide resistance: preserving the effectiveness of modern malaria vector control**. *World Health Organization, Geneva, Switzerland* 2011.

[84] Insecticide resistance action committee (IRAC). *Prevention and management of insecticide resistance in vectors of public health importance.* 2010.

[85] Read A, Lynch P, Thomas M: **How to make evolution-proof insecticides for malaria control**. *PLoS Biol* 2009, **7**(4).

[86] Koella JC, Lynch PA, Thomas MB, Read AF: **Towards evolution-proof malaria control with insecticides**. *Evol Appl* 2009, **2**(4):469–480.

[87] Gourley SA, Liu R, Wu J: **Slowing the evolution of insecticide resistance in mosquitoes: a mathematical model**. *Proc R Soc A* 2011, **467**(2132):2127–2148.

[88] Glunt K, Thomas M, Read A: **The Effects of Age, Exposure History and Malaria Infection on the Susceptibility of Anopheles Mosquitoes to Low Concentrations of Pyrethroid**. *PLoS ONE* 2011, **6**(9):e24968.

[89] Curtis C, Hill N, Kasim S: **Are there effective resistance management strategies for vectors of human disease?** *Biol J Linn Soc* 1993, **48**:3–18.

[90] Beier JC, Keating J, Githure JI, Macdonald MB, Impoinvil DE, Novak RJ: **Integrated vector management for malaria control**. *Malar J* 2008, **7**(Suppl. 1):S4.

[91] Karlin S: **Some mathematical models of population genetics**. *AMM* 1972, **79**:699–739.

[92] Hartl DL, Clark AG: *Principles of population genetics*. Sinauer Associates 2007.

[93] DuMouchel W, Anderson W: **The analysis of selection in experimental populations**. *Genetics* 1968, **58**(3):435–449.

[94] Hastings I, Watkins W: **Tolerance is the key to understanding antimalarial drug resistance**. *Trends Parasitol* 2006, **22**(2):71–77.

[95] García GP, Flores AE, Fernández-Salas I, Saavedra-Rodríguez K, Reyes-Solis G, Lozano-Fuentes S, Bond JG, Casas-Martínez M, Ramsey JM, García-Rejón J, Domínguez-Galera M, Ranson H, Hemingway J, Eisen L, Iv WCB: **Recent rapid rise of a permethrin knock down resistance allele in *Aedes aegypti* in México**. *PLoS Negl Trop Dis* 2009, **3**(10):e531.

[96] Fisher R, Bennett J: *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press 1999.

[97] Smith JM: *Evolutionary Genetics*. Oxford University Press 1998.

[98] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2010, `[http://www.R-project.org]`. [ISBN 3-900051-07-0].

[99] Myung I: **Tutorial on maximum likelihood estimation**. *J Math Biol* 2003, **47**:90–100.

[100] Scott M, Diwell K, McKenzie J: **Dieldrin resistance in Lucilia cuprina (the Australian sheep blowfly): chance, selection and response**. *Heredity* 2000, **84**:599–604.

[101] McKenzie J, Yen J: **Genotype, environment and the assymetry phenotype. Dieldrin-resistance in *Lucilia cuprina* (the Australian sheep blowfly)**. *Heredity* 1995, **75**:181–187.

[102] Endler J: *Natural selection in the wild*. Princeton University Press 1986.

[103] Linnen CR, Hoekstra HE: **Measuring natural selection on genotypes and phenotypes in the wild**. *Cold Spring Harb Symp Quant Biol* 2009, **74**:155–168.

[104] Curtis C, Cook L, Wood R: **Selection for and against insecticide resistance and possible methods of inhibiting the evolution of resistance in mosquitoes**. *Ecol Entomol* 1978, **3**(4):273–287.

[105] Wood R, Cook L: **A note on estimating selection pressures on insecticide-resistance genes**. *Bull World Health Organ* 1983, **61**:129–134.

[106] Livingston M, Carlson G, Fackler P: **Use of mathematical models to estimate characteristics of pyrethroid resistance in tobacco budworm and**

bollworm (Lepidoptera : Noctuidae) field populations. *J Econ Entomol* 2002, **95**(5):1008–1017.

[107] Labbé P, Sidos N, Raymond M, Lenormand T: **Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long-term cline series**. *Genetics* 2009, **182**:303–312.

[108] Lenormand T, Bourguet D, Guillemaud T, Raymond M: **Tracking the evolution of insecticide resistance in the mosquito Culex pipiens**. *Nature* 1999, **400**(6747):861–864.

[109] Templeton A: **Analysis of selection in populations observed over a sequence of consecutive generations**. *Theor Appl Genet* 1974, **45**:179–191.

[110] Ohta T: **Slightly deleterious mutant substitutions in evolution**. *Nature* 1973, **246**(5428):96–98.

[111] Gorrochotegui-Escalante N, Gomez-Machorro C, Lozano-Fuentes S, Fernandez-Salas I, Munoz M, Farfan-Ale J, Garcia-Rejon J, Beaty B, Iv WCB: **Breeding structure of Aedes aegypti populations in Mexico varies by region**. *Am J Trop Med Hyg* 2002, **66**(2):213–222.

[112] Lehmann T, Hawley WA, Grebert H, Collins FH: **The effective population size of Anopheles gambiae in Kenya: implications for population structure**. *Mol Biol Evol* 1998, **15**(3):264–276.

[113] Pinto J, Donnelly MJ, Sousa CA, Malta-Vacas J, Gil V, Ferreira C, Petrarca V, Rosário VED, Charlwood JD: **An island within an island: genetic differentiation of Anopheles gambiae in São Tomé, West Africa, and its relevance to malaria vector control**. *Heredity* 2003, **91**(4):407–414.

[114] Solano P, Ravel S, Bouyer J, Camara M, Kagbadouno MS, Dyer N, Gardes L, Herault D, Donnelly MJ, Meeûs TD: **The population structure of Glossina palpalis gambiensis from island and continental locations in Coastal Guinea**. *PLoS Negl Trop Dis* 2009, **3**(3):e392.

[115] Ranson H, Abdallah H, Badolo A, Guelbeogo WM, Kerah-Hinzoumbe C, Yangalbe-Kalnone E, Sagnon N, Simard F, Coetzee M: **Insecticide resistance in *Anopheles gambiae*: data from the first year of a multi-country study highlight the extent of the problem**. *Malar J* 2009, **8**(299).

[116] Flores AE, Grajales JS, Salas IF, Garcia GP, Becerra MHL, Lozano S, Brogdon WG, Black WC, Beaty B: **Mechanisms of insecticide resistance in field populations of *Aedes aegypti* (L.) from Quintana Roo, Southern Mexico**. *J Am Mosq Control Assoc* 2006, **22**(4):672–677.

[117] Hastings I, Donnelly M: **The impact of antimalarial drug resistance mutations on parasite fitness, and its implications for the evolution of resistance**. *Drug Resist Update* 2005, **8**:43–50.

[118] Saavedra-Rodriguez K, Urdaneta-Marquez L, Rajatileka S, Moulton M, Flores A, Fernandez-Salas I, Bisset J, Rodriguez M, Mccall P, Donnelly M: **A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American Aedes aegypti**. *Insect Mol Biol* 2007, **16**(6):785–798.

[119] Saavedra-Rodriguez K, Strode C, Suarez AF, Salas IF, Ranson H, Hemingway J, Black WC: **Quantitative Trait Loci Mapping of Genome Regions Controlling Permethrin Resistance in the Mosquito Aedes aegypti**. *Genetics* 2008, **180**(2):1137–1152.

[120] Clements AN: *The biology of mosquitoes. Development, nutrition and reproduction*, *Volume 1*. Chapman and Hall 1992.

[121] Murray CJL, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, Lopez AD: **Global malaria mortality between 1980 and 2010: a systematic analysis**. *Lancet* 2012, **379**:413–431.

[122] Chitnis N, Smith T, Steketee R: **A mathematical model for the dynamics of malaria in mosquitoes feeding on a heterogeneous host population**. *J Biol Dyn* 2008, **2**(3):259–285.

[123] Hermsen R, Hwa T: **Sources and Sinks: A Stochastic Model of Evolution in Heterogeneous Environments**. *Phys Rev Lett* 2010, **105**(24):248104.

[124] Metcalf RL: **Mode of action of insecticide synergists**. *Annu Rev Entomol* 1967, **12**:229–256.

[125] B-Bernard C, Philogène B: **Insecticide synergists: role, importance, and perspectives**. *J Toxicol Environ Health A* 1993, **38**(2):199–223.

[126] Guillet P, N'Guessan R, Darriet F, Traore-Lamizana M, Chandre F, Carnevale P: **Combined pyrethroid and carbamate 'two-in-one' treated mosquito nets: field efficacy against pyrethroid-resistant *Anopheles gambiae* and *Culex quinquefasciatus***. *Med Vet Entomol* 2001, **15**:105–112.

[127] Oxborough RM, Mosha FW, Matowo J, Mndeme R, Feston E, Hemingway J, Rowland M: **Mosquitoes and bednets: testing the spatial positioning of insecticide on nets and the rationale behind combination insecticide treatments**. *Ann Trop Med Parasitol* 2008, **102**(8):717–727.

[128] **Vestergaard-Frandsen PermaNet® 3.0 - Features** [http://www.vestergaard-frandsen.com/permanet/permanet-3/features].

[129] Bourguet D, Genissel A, Raymond M: **Insecticide resistance and dominance levels**. *J Econ Entomol* 2000, **93**(6):1588–1595.

[130] Bourguet D, Prout M, Raymond M: **Dominance of insecticide resistance presents a plastic response**. *Genetics* 1996, **143**:407–416.

[131] Zhivotovsky LA, Feldman MW, Bergman A: **Fitness patterns and phenotypic plasticity in a spatially heterogeneous environment**. *Genet Res* 1996, **68**(3):241–248.

[132] Kidwell JF, Clegg MT, Stewart FM, Prout T: **Regions of stable equilibria for models of differential selection in the two sexes under random mating**. *Genetics* 1977, **85**:171–183.

[133] Sanchez M, Blower S: **Uncertainty and sensitivity analysis of the basic reproductive rate: tuberculosis as an example**. *Am J Epidemiol* 1997, **145**(12):1127–1137.

[134] Gu W, Novak RJ: **Predicting the impact of insecticide-treated bed nets on malaria transmission: the devil is in the detail**. *Malar J* 2009, **8**(256).

[135] Breiman L: *Classification and regression trees*. Wadsworth International Group 1984.

[136] Therneau T, Atkinson EJ: *An introduction to recursive partitioning using the RPART routines*. Mayo Foundation 2011.

[137] Miller J, Franklin J: **Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence**. *Ecol Model* 2002, **157**(2-3):227–247.

[138] Austin PC: **A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality**. *Stat Med* 2007, **26**(15):2937–2957.

[139] Tungu P, Magesa S, Maxwell C, Malima R, Masue D, Sudi W, Myamba J, Pigeon O, Rowland M: **Evaluation of PermaNet 3.0 a deltamethrin-PBO combination net against *Anopheles gambiae* and pyrethroid resistant *Culex quinquefasciatus* mosquitoes: an experimental hut trial in Tanzania**. *Malar J* 2010, **9**(21).

[140] Yakob L, Dunning R, Yan G: **Indoor residual spray and insecticide-treated bednets for malaria control: theoretical synergisms and antagonisms**. *J R Soc Interface* 2010.

[141] N'guessan R, Asidi A, Boko P, Odjo A, Akogbeto M, Pigeon O, Rowland M: **An experimental hut evaluation of PermaNet(®) 3.0, a deltamethrin-piperonyl butoxide combination net, against pyrethroid-resistant *Anopheles gambiae* and *Culex quinquefasciatus* mosquitoes in southern Benin**. *Trans R Soc Trop Med Hyg* 2010, **104**(12):758–765.

[142] Killeen G, Okumu F, N'guessan R, Coosemans M, Adeogun A, Awolola S, Etang J, Dabiré R, Corbel V: **The importance of considering community-level effects when selecting insecticidal malaria vector products**. *Parasit Vectors* 2011, **4**:160.

[143] White MT, Griffin JT, Churcher TS, Ferguson NM, Basáñez MG, Ghani AC: **Modelling the impact of vector control interventions on *Anopheles gambiae* population dynamics**. *Parasit Vectors* 2011, **4**:153.

[144] N'Guessan R, Darriet F, Guillet P, Carnevale P, Traore-Lamizana M, Corbel V, Koffi AA, Chandre F: **Resistance to carbosulfan in *Anopheles gambiae* from Ivory Coast, based on reduced sensitivity of acetylcholinesterase**. *Med Vet Entomol* 2003, **17**:19–25.

[145] Ross R: *The prevention of malaria*. New York, E.P. Dutton & company 1911.

[146] Macdonald G: *The epidemiology and control of malaria*. Oxford University Press 1957.

[147] Smith D, McKenzie E: **Statics and dynamics of malaria infection in *Anopheles* mosquitoes**. *Malar J* 2004, **3**(13).

[148] Kermack WO, McKendrick AG: **A contribution to the mathematical theory of epidemics**. *Proc R Soc Lond A* 1927, **115**:700–721.

[149] Mandal S, Sarkar R, Sinha S: **Mathematical models of malaria-a review**. *Malar J* 2011, **10**(202).

[150] Koella JC: **On the use of mathematical models of malaria transmission**. *Acta Trop* 1991, **49**:1–25.

[151] Ross R: **The logical basis of the sanitary policy of mosquito reduction**. *Science* 1905, **22**:689–699.

[152] Fillinger U, Lindsay S: **Larval source management for malaria control in Africa: myths and reality**. *Malar J* 2011, **10**(353).

[153] Gu W, Novak R: **Habitat-based modeling of impacts of mosquito larval interventions on entomological inoculation rates, incidence, and prevalence of malaria**. *Am J Trop Med Hyg* 2005, **73**(3):546–552.

[154] Killeen G, Fillinger U, Knols B: **Advantages of larval control for African malaria vectors: Low mobility and behavioural responsiveness of immature mosquito stages allow high effective coverage**. *Malar J* 2002, **1**(8).

[155] Griffin J, Hollingsworth T, Okell L, Churcher T, White M, Hinsley W, Bousema T, Drakeley C, Ferguson N, Basáñez M: **Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies**. *PLoS Med* 2010, **7**(8):e1000324.

[156] Saul A, Graves P, Kay B: **A cyclical feeding model for pathogen transmission and its application to determine vectorial capacity for vector infection rates**. *J Appl Ecol* 1990, :123–133.

[157] Menach AL, Takala S, McKenzie FE, Perisse A, Harris A, Flahault A, Smith DL: **An elaborated feeding cycle model for reductions in vectorial capacity of night-biting mosquitoes by insecticide-treated nets**. *Malar J* 2007, **6**(10).

[158] Killeen G, Smith T: **Exploring the contributions of bed nets, cattle, insecticides and excitorepellency to malaria control: a deterministic model of mosquito host-seeking behaviour and mortality**. *Trans R Soc Trop Med Hyg* 2007, **101**:867–880.

[159] Chitnis N, Schapira A, Smith T, Steketee R: **Comparing the effectiveness of malaria vector-control interventions through a mathematical model**. *Am J Trop Med Hyg* 2010, **83**(2):230–240.

[160] Chitnis N, Hardy D, Smith T: **A periodically-forced mathematical model for the seasonal dynamics of malaria in mosquitoes**. *Bull Math Biol* 2012, **74**(5):1098–1124.

[161] Yakob L, Yan G: **Modeling the effects of integrating larval habitat source reduction and insecticide treated nets for malaria control**. *PLoS ONE* 2009, **4**(9):e6921.

[162] Saul A: **Zooprophylaxis or zoopotentiation: the outcome of introducing animals on vector transmission is highly dependent on the mosquito mortality while searching**. *Malar J* 2003, **2**(32).

[163] Boëte C, Koella JC: **A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control**. *Malar J* 2002, **1**(3).

[164] Hancock PA, Thomas MB, Godfray HCJ: **An age-structured model to evaluate the potential of novel malaria-control interventions: a case study of fungal biopesticide sprays**. *Proc Biol Sci* 2009, **276**(1654):71–80.

[165] Depinay JMO, Mbogo CM, Killeen G, Knols B, Beier J, Carlson J, Dushoff J, Billingsley P, Mwambi H, Githure J, Toure AM, McKenzie FE: **A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission**. *Malar J* 2004, **3**(29).

[166] Eckhoff P: **A malaria transmission-directed model of mosquito life cycle and ecology**. *Malar J* 2011, **10**(303).

[167] Otero M, Solari HG, Schweigmann N: **A stochastic population dynamics model for *Aedes aegypti*: formulation and application to a city with temperate climate**. *Bull Math Biol* 2006, **68**(8):1945–1974.

[168] Cooke K, van den Driessche P, Zou X: **Interaction of maturation delay and nonlinear birth in population and epidemic models**. *J Math Biol* 1999, **39**(4):332–352.

[169] Hancock PA, Godfray HCJ: **Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions**. *Malar J* 2007, **6**(98).

[170] Ahumada JA, Lapointe D, Samuel MD: **Modeling the population dynamics of *Culex quinquefasciatus* (Diptera: Culicidae), along an elevational gradient in Hawaii**. *J Med Entomol* 2004, **41**(6):1157–1170.

[171] Focks DA, Haile DG, Daniels E, Mount GA: **Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development**. *J Med Entomol* 1993, **30**(6):1003–1017.

[172] Dye C: **Models for the population dynamics of the yellow fever mosquito, *Aedes aegypti***. *J Anim Ecol* 1984, **53**:247–268.

[173] Yakob L, Alphey L, Bonsall MB: ***Aedes aegypti* control: the concomitant role of competition, space and transgenic technologies**. *J Appl Ecol* 2008, **45**(4):1258–1265.

[174] Lu J, Li J: **Dynamics of stage-structured discrete mosquito population models**. *JAAC* 2011, **1**:53–67.

[175] Vandermeer J, Goldberg D: *Population Ecology: First Principles*. Princeton paperbacks, Princeton University Press 2003.

[176] Anderson R, May R: *Infectious diseases of humans: dynamics and control*. Oxford University Press 1992.

[177] Olayemi IK, Ande AT: **Life table analysis of *Anopheles gambiae* (Diptera: Culicidae) in relation to malaria transmission**. *J Vector Dis* 2009, **46**(4):295–298.

[178] Frey HC, Patil S: **Identification and review of sensitivity analysis methods**. *Risk Anal* 2002, **22**(3):553–578.

[179] Takken W, Charlwood J, Billingsley P, Gort G: **Dispersal and survival of *Anopheles funestus* and A. gambiae sl (Diptera: Culicidae) during the rainy season in southeast Tanzania**. *B Entomol Res* 1998, **88**(5):561–566.

[180] Menach A, McKenzie F, Flahault A, Smith D: **The unexpected importance of mosquito oviposition behaviour for malaria: non-productive larval habitats can be sources for malaria transmission**. *Malar J* 2005, **4**(23).

[181] Russell T, Lwetoijera D, Knols B, Takken W, Killeen G, Ferguson H: **Linking individual phenotype to density-dependent population growth: the influence of body size on the population dynamics of malaria vectors**. *Proc R Soc Lond B* 2011, **278**(1721):3142–3151.

[182] Gimnig J, Ombok M, Otieno S, Kaufman M, Vulule J, Walker E: **Density-dependent development of *Anopheles gambiae* (Diptera: Culicidae) larvae in artificial habitats**. *J Med Entomol* 2002, **39**:162–172.

[183] Dye C: **Models for the Population Dynamics of the Yellow Fever Mosquito, *Aedes aegypti***. *J Anim Ecol* 1984, **53**:247–268.

[184] Donnelly MJ, Corbel V, Weetman D, Wilding CS, Williamson MS, Iv WCB: **Does kdr genotype predict insecticide-resistance phenotype in mosquitoes?** *Trends Parasitol* 2009, **25**(5):213–219.

[185] Ranson H, N'guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V: **Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control?** *Trends Parasitol* 2011, **27**(2):91–98.

[186] N'Guessan R, Corbel V, Akogbéto M, Rowland M: **Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin**. *Emer Infect Dis* 2007, **13**(2):199–206.

[187] Corbel V, Akogbeto M, Damien G, Djenontin A, Chandre F, Rogier C, Moiroux N, Chabi J, Banganna B, Padonou GG, Henry MC: **Combination of malaria vector control interventions in pyrethroid resistance area in Benin: a cluster randomised controlled trial**. *Lancet Infect Dis* 2012, **12**:617–626.

[188] Hamel MJ, Otieno P, Bayoh N, Kariuki S, Were V, Marwanga D, Laserson KF, Williamson J, Slutsker L, Gimnig J: **The combination of indoor residual spraying and insecticide-treated nets provides added protection against malaria compared with Insecticide-treated nets alone**. *Am J Trop Med Hyg* 2011, **85**(6):1080–1086.

[189] Kleinschmidt I, Schwabe C, Shiva M, Segura JL, Sima V, Mabunda SJA, Coleman M: **Combining indoor residual spraying and insecticide-treated net interventions**. *Am J Trop Med Hyg* 2009, **81**(3):519–524.

[190] NGuessan R, Rowland M: **Indoor residual spraying for prevention of malaria**. *Lancet Infect Dis* 2012, **12**:581–582.

[191] Okumu F, Moore S: **Combining indoor residual spraying and insecticide-treated nets for malaria control in Africa: a review of possible outcomes and an outline of suggestions for the future**. *Malar J* 2011, **10**(208).

[192] Over M, Bakote'e B, Velayudhan R, Wilikai P, Graves PM: **Impregnated nets or ddt residual spraying? Field effectiveness of malaria prevention techniques in Solomon islands, 1993-1999**. *Am J Trop Med Hyg* 2004, **71**(Suppl. 2):214–223.

[193] Smith DL, McKenzie FE, Snow RW, Hay SI: **Revisiting the basic reproductive number for malaria and its implications for malaria control**. *PLoS Biol* 2007, **5**(3):e42.

[194] Hancock P, Godfray H: **Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions**. *Malar J* 2007, **6**(98).

[195] malERA: **A Research Agenda for Malaria Eradication: Modeling**. *PLoS Med* 2011, **8**:e1000403.

Appendix A

Comment

Comment posted online on PLoS ONE website (24^{th} of January 2012) regarding the paper:

Glunt KD, Thomas MB, Read AF: **The Effects of Age, Exposure History and Malaria Infection on the Susceptibility of Anopheles Mosquitoes to Low Concentrations of Pyrethroid**. PLoS ONE 2011, **6**(9): e24968.

We read with interest this paper by Glunt *et al* (2011). It contains some interesting ideas on developing a methodology to differentially kill young and old mosquitoes, following the reasoning that killing old mosquitoes, and ideally old malaria-infected mosquitoes, could provide effective malaria control while only weakly selecting for resistance (Read *et al.*, 2009). However, we would like to quantify and draw attention to what is, in our opinion, the largest danger inherent in this approach, namely increasing the dominance levels of resistant mutations. The dominance of a mutation is not an inherent property, but is determined by the environment within which selection takes place. Denote the resistance mutation as 'R' and the original susceptible form as 'S'. High concentrations may kill both RR and RS genotypes making the 'R' mutation recessive, while low concentration may allow the RS genotype to survive, making it dominant (see Figure 1 of Barbosa et al). This is vitally important because dominance relationships between susceptible and resistance alleles hugely affect the rate of spread of resistance. Increasing dominance greatly increases the rate at which resistance evolves (Figure 2 of Barbosa *et al.* (2011)). As a specific example, if initial frequency of resistance is 0.1% and insecticide deployment makes the resistant homozygote 20% fitter than the wild type (s=0.2), we can use the standard population genetic equation (Equation 1 in Barbosa *et al.* (2011)) to predict the time for insecticide resistance to spread. If the resistant mutation is near recessive (h=0.1) it takes 248 generations (around 20 years assuming 12 generations per year) to reach an overall frequency of 50%. In contrast, if low insecticide concentrations make it semi-dominant (h=0.5) then

it takes 73 generations (around 6 years) and if it is near dominant (h=0.9) it only takes around 47 generations (4 years) to reach 50%. These differences far exceed the differences likely to be generated by their proposal to deploy insecticides at low concentrations. They note these concerns about altered dominance levels stating that it represents 'conventional wisdom'. We would prefer the more objective term 'elementary genetic theory' which, at least in our opinion, in this case appears to be extremely robust. Glunt *et al.* (2011) discussed a number of practical difficulties in translating their approach into policy, to which we would add the following:

(1) Insecticides applied at low concentrations will decay over time to nearly ineffective levels. The application on surfaces such as walls may also be patchy. This temporal and spatial heterogeneity may well result in mosquitoes being exposed to a mosaic of ineffective, 'low' and 'high' concentrations and it is not clear how this heterogeneity will affect their conclusions.

(2) It would be difficult to accurately calibrate this strategy because we have no real idea of how exposure in the lab correlates with that in the field. For example Glunt *et al.*(2011) used a modified WHO assay where mosquitoes are continuously exposed for 1 hour and it is hard to understand how this will translate into killing in the field where mosquitoes exposure to insecticides on walls may be very prolonged (if they rest on the walls) or extremely brief when mosquitoes may make contact with bednets for only a few seconds.

(3) The experiments were conducted using a single susceptible strain, reared under controlled laboratory conditions, which does not mimic the genetic and environmental variation that occurs in nature. The patterns of survival would most likely change if resistance was already present and the pattern would depend also on the type of resistance (target site or detoxification) (Rajatileka *et al.*, 2010).

In summary, we would note that the paper makes some interesting points but, for the sake of policy makers, would stress that such strategies carry huge dangers in altering the dominance/recessively of insecticide resistance that, at least in our opinion, preclude its practical application in the present form.

References:

Barbosa S, Black WC IV, Hastings I: **Challenges in Estimating Insecticide Selection Pressures from Mosquito Field Data**. PLoS Negl Trop Dis 2011, **5**(11): e1387.

Rajatileka S, Burhani J, Ranson H: Mosquito age and susceptibility to insecticides. T Roy Soc Trop Med H 2011, **105**(5).

Read AF, Lynch PA, Thomas MB: **How to make evolution-proof insecticides for malaria control**. PLoS Biol 2009, **7**(4): e1000058.

Appendix B

Co-authorship paper

Field, Genetic, and Modeling Approaches Show Strong Positive Selection Acting upon an Insecticide Resistance Mutation in *Anopheles gambiae s.s.*

Amy Lynd,[1] David Weetman,[1] Susana Barbosa,[2] Alexander Egyir Yawson,[3] Sara Mitchell,[1] Joao Pinto,[4] Ian Hastings,[2] and Martin J. Donnelly*[,1]

[1]Vector Group, Liverpool Tropical School of Medicine, Liverpool, United Kingdom

[2]Molecular and Biochemical Parasitology Group, Liverpool Tropical School of Medicine, Liverpool, United Kingdom

[3]Biotechnology and Nuclear Agriculture Research Institute, Ghana Atomic Energy Commission, Accra, Ghana

[4]Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisbon, Lisbon, Portugal

***Corresponding author:** E-mail: m.j.donnelly@liv.ac.uk.

Associate editor: John H McDonald

Research article

Abstract

Alleles subject to strong, recent positive selection will be swept toward fixation together with contiguous sections of the genome. Whether the genomic signatures of such selection will be readily detectable in outbred wild populations is unclear. In this study, we employ haplotype diversity analysis to examine evidence for selective sweeps around knockdown resistance (*kdr*) mutations associated with resistance to dichlorodiphenyltrichloroethane and pyrethroid insecticides in the mosquito *Anopheles gambiae*. Both *kdr* mutations have significantly lower haplotype diversity than the wild-type (nonresistant) allele, with *kdr L1014F* showing the most pronounced footprint of selection. We complement these data with a time series of collections showing that the *L1014F* allele has increased in frequency from 0.05 to 0.54 in 5 years, consistent with a maximum likelihood–fitted selection coefficient of 0.16 and a dominance coefficient of 0.25. Our data show that strong, recent positive selective events, such as those caused by insecticide resistance, can be identified in wild insect populations.

Key words: *Anopheles gambiae*, malaria, insecticide resistance, selection.

Introduction

The sequencing of the *Anopheles gambiae* genome has opened up the possibility for genome-wide single nucleotide polymorphism (SNP)–based association mapping studies that have been successful in identifying positively selected loci in the human genome (Sabeti et al. 2002, 2007; Bersaglieri et al. 2004). The resolution of the association mapping approach is defined by the probability that recombination will have broken down the association between markers and a trait-associated functional polymorphism. Data from extensive resequencing of (primarily) detoxification genes in samples from wild populations of *A. gambiae* revealed a very high frequency of segregating sites (Wilding et al. 2009), consistent with high rates of recombination (Begun and Aquadro 1992; Begun et al. 2007) and/or a long history of outbreeding. In isofemale lab strains of *Drosophila spp.*, it has been possible to observe selective sweeps around insecticide resistance–associated loci (Schlenke and Begun 2004; Aminetzach et al. 2005), but how long these signatures persist in wild populations is unknown. In this paper, we use linkage disequilibrium (LD)–based haplotype diversity analysis (Sabeti et al. 2006) to investigate the pattern of molecular genetic variation associated with insecticide resistance mutations at the pyrethroid and dichlorodiphenyltrichloroethane (DDT) knockdown resistance locus, *kdr*, in the African malaria mosquito *A. gambiae s.s.* Furthermore, as a corollary of this indirect genetic approach we demonstrate, using a series of temporal collections, a dramatic increase in *kdr* frequency in a population of *A. gambiae s.s.* over a period of approximately 72 generations. Data from these temporal collections are used to estimate the selection and dominance coefficients operating on *kdr* in the field to illustrate the potential levels of selection necessary to produce the patterns of LD we observe.

Insecticide-treated bed nets are the principal method for preventing malaria in sub-Saharan Africa. Currently, pyrethroids are the only class of insecticides licensed for use on nets, and there is concern that resistance will compromise control programs. To date the most commonly recorded resistance mechanism is termed "knockdown resistance" and results from single–base pair mutations in the voltage-gated sodium channel. The sodium channel gene, located within division 20C near the centromere of chromosome 2L, codes for a protein that is the target site of pyrethroid insecticides. Two alternative single–base pair mutations have been found in *A. gambiae*, and these *kdr* mutations can cause target-site insensitivity to pyrethroids as well as cross-resistance to DDT. The substitutions cause amino acid changes at codon 1014 within the transmembrane structure of segment 6 in domain II of the voltage-gated sodium channel (numbering according to the

Table 1. Origin and *kdr* Genotype of Specimens Used in the Study.

Population	Year Collected	Total *N*	Form	Number of Each *kdr* Genotype					
				L1014S/ *L1014S*	*L1014F*/ *L1014F*	*L1014F*/ *L1014S*	*L1014S*/ wt	*L1014F*/ wt	Wt/wt
Asembo Bay, Kenya, 00°10′S, 34°22′E	2005[1]	48	S	11	—	—	17	—	20
Dienga, Gabon, 01°52′S, 12°40′E	1999–2000[2]	30	S	—	—	—	4	2	24
Bakoumba, Gabon, 01°49′S, 13°01′E	1999–2000[2]	42	S	—	5	8	5	7	17
Libreville, Gabon, 00°22′N, 09°26′E	1999–2000[2]	73	S	34	8	31	—	—	—
Okyereko and Accra area, Ghana, 05°24.9′N, 00°36.6′W, 05°38′N, 00°15′E	2002[3]	35	S	—	33	—	—	2	—
Okyereko, Ghana, 05°24.9′N, 00°36.6′W	2002[3]	30	M	—	—	—	—	2	28

NOTE.—The population name and total numbers of each DNA sample utilized. Molecular form is indicated, and the numbers of each *kdr* genotype are shown. Additional information on the collection sites may be obtained from the publications where the specimens are originally described: [1]Müller et al. (2008), [2]Pinto et al. (2006), and [3]Yawson et al. (2004); wt, wild type.

housefly *para* sequence, GenBank X96668). The *L1014F* mutation, a leucine to phenylalanine change, was first observed in West Africa (Martinez-Torres et al. 1998), and the same substitution has been observed in a diverse array of insects (Davies et al. 2007a). A second substitution, *L1014S*, was observed more recently in East African *A. gambiae* (Ranson et al. 2000) and involves the adjacent base of the same codon, resulting in a leucine to serine change.

There are two incipient species within the nominal taxon *A. gambiae s.s.* that are characterized by mutations on the X chromosome and are termed M and S form. The distribution of the *kdr* mutation is not uniform either within or between forms, although in general *kdr* alleles have been found at much higher frequencies in *A. gambiae s.s.* S-form samples compared with M-form samples (reviewed in Santolamazza et al. 2008). The reasons for the differences in distribution remain unclear because little is known about the origins of the *kdr* mutations and the selection pressures acting upon them in wild populations. In a sample from Benin, the *L1014F* was found in tight LD with two upstream intronic polymorphisms in both M- and S-form individuals. The two upstream polymorphisms associated with the *L1014F* variant were not found in wild-type M-form individuals but were common in wild-type S-form individuals, suggestive of an introgression event from S-form to M-form populations (Weill et al. 2000). This linkage between *kdr* and the intronic polymorphisms was not seen in M-form individuals from Bioko Island and was thought to indicate de novo mutation (Reimer et al. 2008). More recently, a study of S-form specimens from 15 countries suggested that the *L1014F* and *L1014S* mutations have both arisen independently on at least two separate occasions (Pinto et al. 2007).

Samples were obtained from three regions in sub-Saharan Africa; Kenya (East Africa) *A. gambiae*, S molecular form, *kdr L1014S* allele present; Ghana (West Africa) both M and S molecular form, *kdr L1014F* allele present; Gabon (Central Africa) S molecular form, both *L1014S* and *L1014F kdr* alleles present.

These population samples allow us to address a number of questions.

1. Available evidence suggests that the *L1014S* mutation has high penetrance for a DDT-resistant phenotype but lower penetrance for a pyrethroid-resistant phenotype than the *L1014F* mutation (Ranson et al. 2000). DDT was banned in Kenya in 1990, and we can investigate the signature of positive selection associated with weaker selection or recombination and relaxed selection.
2. The populations from central Africa are some of the few locations where both *L1014F* and *L1014S* alleles are observed sympatrically (Santolamazza et al. 2008). Indeed, in an earlier study, a significant, albeit marginal, *L1014F*/*L1014S* heterozygote excess was observed in samples from Libreville, Gabon (Pinto et al. 2006). By comparing patterns of LD around the three alleles, we investigate whether the unusually high frequency of the *L1014S* allele in these populations (63%; Pinto et al. 2006) is a result of a recent selective sweep.
3. In many S-form populations in West Africa, including our collections from Ghana, the *L1014F* allele is close to fixation. In the absence of wild-type alleles, we are unable to control for local variation in recombination rates (Sabeti et al. 2007), and it is therefore impossible to ascribe patterns of LD to a positive selection event. Recently developed approaches such as cross-population extended haplotype homozygosity (EHH) have been developed to allow interpopulation comparisons in instances where alleles proceed to near fixation in some populations (Sabeti et al. 2007), but in our system resistance alleles may have multiple origins, presenting a confounding variable (Pinto et al. 2007). However, the presence of sympatric M-form populations in southern Ghana (Yawson et al. 2004, 2007) allows us to both document the increase in frequency of the same *L1014F* haplotype, following an introgression event, over a period of 5 years and estimate the selection and dominance coefficients associated with the signatures of positive selection.

Materials and Methods

Sample Sites, DNA Extraction, and Species Identification

Adult female *A. gambiae s.s.* mosquitoes used in this study were obtained from aspirator and pyrethroid knockdown collections from the field in various geographic locations (table 1). DNA was extracted from single female *A. gambiae* using either a modified Livak method or a phenol–chloroform method (Livak 1984; Ballinger-Crabtree et al. 1992). Species identification polymerase

chain reaction (PCR) was carried out on *A. gambiae s.l.* according to the protocol (Scott et al. 1993). Reactions were then digested with *Cfo*I restriction enzyme for 24 h at 37 °C in order to type *A. gambiae s.s* mosquitoes to M and S form (Fanello et al. 2002), and products visualized under UV light after electrophoresis on a 2% agarose Tris/borate/EDTA (TBE) gel with ethidium bromide. *Kdr* genotypes were determined by allele-specific PCR, heated oligonucleotide ligation assay (Lynd et al. 2005), or Taqman assay (Bass et al. 2007) depending upon year of collection.

Sodium Channel SNP Identification

The voltage-gated sodium channel gene is nearly 74 kbp in length and is composed of 35 exons including two duplicate exons (Davies et al. 2007a). Ten regions of the sodium channel were amplified by PCR for direct sequencing. Where possible, primers were designed to bind within exons to produce amplicons that spanned an intron with a maximum size of 1.5 kbp. Exons (numbering as Davies et al. 2007a) 1–2, 3, 4, 7–9, 13–14, 15–17, 20c, 23–24, 28–30, and 32–33 were selected as targets for sequencing. Primer and amplification details are provided (supplementary table 1, Supplementary Material online). Sequencing for SNP detection was carried out on up to 12 individuals of known *kdr* genotype from Ghana, São Tomé, Gabon, Angola, Mozambique, Malawi, and Kenya, from a susceptible laboratory strain (KISUMU), and from a permethrin tolerant resistant laboratory strain (reduced susceptibility to permethrin), both originating from Kenya. PCR products were cleaned using a Mini Elute PCR Purification kit (Qiagen) and then sequenced in both directions. Sequences were aligned using Bioedit software version 7.0.5.2 (Hall 1999) and then manually annotated for polymorphisms and ambiguities.

In addition, seven M-form individuals from Accra, Ghana, homozygous for the *L1014F* allele were bidirectionally sequenced across PCR amplicons 13–14, 15–17, and 21 to determine the associated haplotype of the *kdr* allele in this population.

SNP Screening

SNPs discovered through resequencing were screened in the large-scale SNP detection study using the SNPStart Primer Extension Kit on the Beckman CEQ 8000 Genetic Analysis System. Details of SNPs both included and excluded from the SNP screening are given in supplementary table 2, Supplementary Material online. Multiplex PCR was carried out to amplify the regions of DNA containing SNPs of interest, including a region of exon 20 and the preceding intron to allow high-throughput detection of the *kdr* mutation and three other well-characterized SNPs (Weill et al. 2000; Diabate et al. 2004; Pinto et al. 2006) (primers and reaction conditions detailed in supplementary table 3, Supplementary Material online). Products were visualized on a 2% TBE agarose gel. Successfully multiplexed samples were prepared for subsequent SNP extension by *Exo*I/shrimp alkaline phosphatase (SAP) enzymatic digestion. Interrogation primers were then designed for each individual SNP chosen for investigation according to the manufacturers' recommendations (supplementary table 4, Supplementary Material online). Single-base extension to the 3′ end of the interrogation primer by a dye terminator molecule, corresponding to the nucleotide found at the SNP location, was carried out using a GenomeLab SNPStart Primer Extension Kit (Beckman Coulter, Amersham, UK). The SAP-digested product was then scored on the Beckman CEQ 8000 Genetic Analysis System.

Data Analysis

As reviewed exhaustively by Sabeti et al. (2006), there are numerous statistical tests of positive selection which differ in their ability to detect selection events on different timescales. For the present SNP data set, it is not possible to use the suite of sequence-based tests that compare synonymous/nonsynonymous differences or detect an excess of rare alleles. We are therefore fortunate that on the timescales in which the emergence, and selection, of insecticide resistance is likely to occur, estimates of interpopulation divergence (e.g., based on *F* statistics) and screens of LD around selected versus wild-type alleles are likely to be the two most powerful analytical approaches. With the sample sizes available in our study, single-marker analyses based on *F*-statistic estimates would perform better as indicators of selection when markers can be typed at a more coarse scale, with consequently enhanced signal:noise ratio. However, with sample size constraints the signal would be difficult to localize. By contrast, long-range haplotype analyses, such as EHH (Sabeti et al. 2002) analysis, perform very well at a fine physical scale in identifying narrow candidate regions (Sabeti et al. 2006).

EHH analysis was carried out to assess the patterns of LD associated with wild type and the two *kdr* alleles. EHH can be defined as the probability that two random chosen chromosomes carrying the core (e.g., the wild-type or *kdr* allele) haplotype of interest are identical by descent. This approach first identifies core haplotypes surrounding the locus of interest and then examines the decay in LD from these core haplotypes to the surrounding loci. The resulting EHH can be used as evidence of recent positive selection at a locus in haplotypes that have high frequency and high EHH (Sabeti et al. 2002). EHH analysis requires haplotype information that cannot be empirically determined from the genotype data gathered by the methods used in this study. Therefore, haplotypes were inferred using PHASE software version 2.1.1 using default parameters (Stephens et al. 2001; Stephens and Scheet 2005). PHASE utilizes a Bayesian coalescent–based approach to determine phase and allows for varying rates of recombination at each SNP interval. The method is based on the idea that an unresolved haplotype is more likely to be the same or be similar to a previous haplotype. This approach was found to outperform other methods available for autosomal human data sets (Stephens et al. 2001; Stephens and Scheet 2005). Data were analyzed together rather than as separate subpopulations because 1) previous studies found this to be more accurate and 2) haplotype determination methods of this nature are relatively insensitive to departures

from Hardy–Weinberg equilibrium so are fairly robust to population substructuring. This approach is also more conservative than determining haplotypes for individual populations because the latter is liable to lead to an underestimation in differences in haplotype frequencies (Stephens and Scheet 2005). Phase reconstruction was executed ten times upon the total data set, and differences in counts of best haplotypes were noted.

The estimated haplotypes obtained from PHASE were used as input for EHH analysis implemented by SWEEP version 2.1.1 (Sabeti et al. 2002). Core haplotypes were selected manually to include only the two adjacent *kdr*-causing loci. Significance of EHH values is usually assigned through comparison to an empirically generated null distribution from other regions of the genome. However, given that we had already identified the causal mutations of interest, we were able to make a comparison of patterns of LD around wild-type and resistant cores. The primary advantage of this approach is that it is not subject to the genome-wide variations in recombination rate which can affect the null distribution approach in species lacking detailed recombination maps. Significant differences in EHH values were determined in two ways: 1) Within country samples, at individual SNP positions with nonoverlapping 95% confidence intervals (CIs). These CIs were calculated at each SNP position using a bootstrapping procedure, carried out in SAS version 9 software. Resampling was carried out 1,000 times. 2) Across all SNPs within and among country samples, the diversity of the different *kdr* allele–bearing haplotypes was compared using sign tests, implemented by SPSS 14. Where exact sign test probabilities could not be calculated, a Monte Carlo procedure with 10,000 permutations was performed. The sequential Bonferroni procedure was applied to determine statistical significance following correction for multiple testing (Holm 1979). Although our data—EHH values at each SNP position—are not independent, it is this nonindependence caused by LD that will cause departure from the null hypothesis of equality of median EHH values. Therefore, the null hypothesis remains that there is no difference in median EHH between *kdr* and wild-type alleles. Bifurcation plots were also created using the SWEEP software. In a bifurcation plot, the core haplotype is represented as a black circle. Each SNP, moving out from the core both upstream and downstream of the *kdr* locus, is a potential site for a bifurcation that would result from the presence of two segregating alleles. Therefore, the diagram provides a means of displaying the breakdown in LD at increasing distance from the core haplotypes. The radius of the circle at each node is proportional to the number of individuals with that haplotype.

Calculation of Selection and Dominance Coefficients

The spread of the *L1014F* allele was modeled using the standard recursive population genetic formula:

$$p' = \frac{p^2(1+s) + p(1-p)(1+hs)}{\bar{W}}, \qquad (1)$$

where p is the frequency of the *L1014F* allele, p' is the frequency in the next generation, s is the selective coefficient of the resistance mutation, h is the dominance coefficient (1 = complete dominance, 0 = complete recessivity), and W is the normalizing factor (Maynard-Smith 1998).

Tracking allele frequencies over time requires three input parameters: initial allele frequency at time zero, s, and h. Estimates of all three unknown parameters were obtained by maximum likelihood assuming a binomial distribution of observed allele frequencies around the predicted frequency. The analysis was performed in R (http://www.r-project.org) using maximum likelihood functions and optimizing routines. The generation time was set at the standard of one generation per calendar month (Lehmann et al. 1998).

Results

SNP Discovery and Screening

Ten genomic regions of a combined length of ≈6.5 kb of DNA, spanning a region of ≈73 kb of the voltage-gated sodium channel, were amplified and sequenced in *A. gambiae s.s.* individuals from seven countries across sub-Saharan Africa. A total of 62 potential SNPs were found, of which 14 were exonic (supplementary table 1, Supplementary Material online). Six intronic indels were observed, usually in poly-A or tandem AT repeats (supplementary table 1, Supplementary Material online). On average, there was one SNP every 106 bp, which represents a low SNP frequency for *A. gambiae*, but similar to other genes in the same genomic locality (chromosome 2L division 20; Wilding et al. 2009). Thirty-two SNPs, including the two *kdr* mutations, were selected for screening in 258 individuals. In S-form individuals, the SNP adjacent to the core in the upstream (centromeric) direction was excluded from further analysis as it was found to be monomorphic. Details of the populations and associated *kdr* genotypes are given in table 1. The genotypic data were resolved into haplotypes with ten runs of the analysis. In only one instance, did the replicate runs resolve a novel estimated haplotype, which in a subsequent comparative analysis was found to exert no qualitative effect on the results. Therefore, all analyses reported here are based upon the haplotypes resolved in the vast majority of the phasing runs.

EHH analysis was carried out to assess the patterns of LD associated with the wild-type and the two *kdr* alleles. The intronic SNPs that have been used to identify the origin of the *kdr* mutations were the proximate SNPs in the centromeric direction (Weill et al. 2000; Pinto et al. 2007). LD decay was examined between these core haplotypes and the remaining 29 or 30 SNP loci (for S or M forms, respectively).

Only two core-alleles were present in the western Kenyan sample: wild type and *L1014S*. In the downstream telomeric direction, EHH decays at a similar rate for both wild type and *L1014S*, but there was a marked contrast between alleles in the centromeric direction, with entirely nonoverlapping confidence limits from just a few kilobases away from the core (figs. 1 and 2*A*). In the Gabonese collection, the difference between resistance-associated alleles

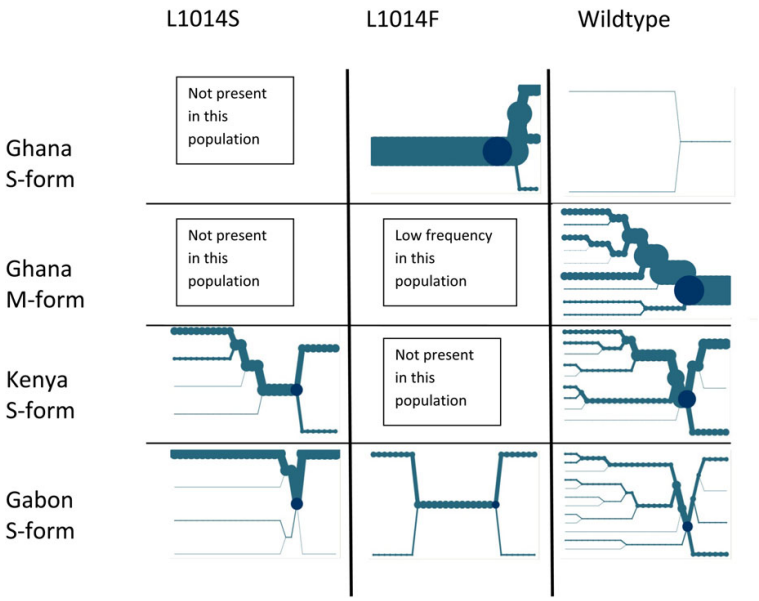

Fig. 1. Bifurcation plots showing patterns of recombination in the centromeric (5′ toward the left) and telomeric (3′ toward the right) directions. The core is marked by the dark circle, and each of the 29/30 SNPs is represented by a node and a recombination event is represented by a bifurcation. The diameter of the circle at each SNP node is proportional to the numbers of individuals with the same long-range haplotype at that position. No bifurcation plot is shown for the *L1014F* core in Ghanaian M-form populations as only a single haplotype was observed (see Results).

and wild type was even more marked with significantly lower EHH in the wild type in both centromeric and telomeric directions less than 5 kb from the core (fig. 2*B*). Indeed, both the *L1014F* and the *L1014S* resistance mutations showed little haplotype bifurcation in the Gabon samples over the length of the sodium channel (fig. 1), suggesting a relatively recent origin for both these mutations accompanied by a strong selective sweep. The patterns of LD are most marked around the resistant *L1014F* haplotype in Ghanaian S-form samples in which the *L1014F kdr* allele was at very high frequency (figs. 1 and 2*C*), as would be expected given the near fixation of this allele in southern Ghana in the S molecular form (mean frequency = 0.96; 95% CI 0.95–0.97) (Yawson et al. 2004). The presence of only two wild-type haplotypes in the sample prevent any meaningful comparison of LD decay, but it should be noted that there was complete LD over the entire 64-kb length of the sodium channel in the centromeric direction. The wild-type allele, observed in the Ghanaian M-form populations (figs. 1 and 2*C*), showed marked LD, only in the telomeric direction, between exons 20 and 32, the opposite directional asymmetry to the *L1014F* mutation in Ghana S-form populations. Although simulation studies have shown that LD decay may be asymmetric even when rates of mutation and recombination are constant (Kim and Stephan 2002), it is possible that the LD observed in these samples may reflect the presence of one or more hitherto overlooked selectively advantageous mutants, although we cannot rule out recombination with unsampled haplotypes (supplementary table 5, Supplementary Material online). Davies et al. (2007b) have summarized that there are a number of additional nonsynonymous changes observed in a variety of taxa, and detailed association mapping studies are presently underway to investigate this phenomenon. Comparing overall levels of EHH for the whole 72.6-kb regions typed, it is interesting to note that median EHH values are statistically indistinguishable for the same allele typed in different populations (table 2) and that a clear hierarchy of evidence for selective sweeps emerged. Median EHH levels were highest for the *L1014F* resistance mutation, followed by those for the *L1014S* mutations, with the lowest for the wild-type allele (table 2). The only exception to this pattern was within the Gabonese sample, the only one in which both resistance alleles were present, where median EHH was equal for the two resistance alleles. Nevertheless, despite the possibilities of different origins of the same allele, and local variation in recombination rates, EHH levels across the genomic region investigated suggest some degree of commonality in selection across populations for each allele, although the actual rate of change in LD with distance can be quite complex and dependent on direction from the core (figs. 1 and 2).

We investigated temporal change in the frequency of the *L1014F* allele and associated haplotype in sympatric populations of M-form individuals in a subset of the populations previously described by Yawson et al. (2004). Using the data reported in Yawson et al. (2004), we estimated the

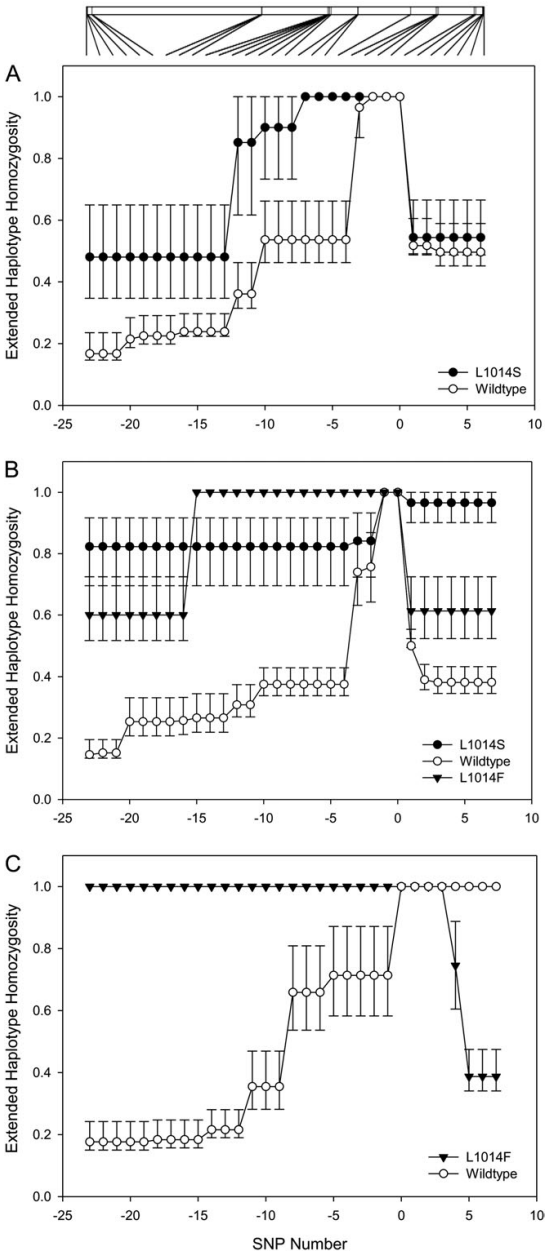

FIG. 2. EHH analysis showing LD decay with increasing distance from the core (marked as the origin on the *x* axis). The 95% CIs were estimated by bootstrapping (see Materials and Methods). The *x* axis is ordinal, negative numbers are in the centromeric direction and positive numbers in the telomeric direction. The scale bar at the top of the figure is 72.6 kb in length and shows the physical distance between the SNPs. (*A*) Kenya data for *L1014S* and wild-type alleles; (*B*) Gabon data for *L1014S*, *L1014F*, and wild-type alleles, and (*C*) Ghana data for *L1014F* (S form) and wild type (M form).

L1014F allele frequency in M-form populations from around Accra, southern Ghana (≈30 km diameter collection area), during 2002 ($freq_{L1014F}$ = 0.03; 95% CI 0.01–0.05). Additional screening in 2007 and 2008 from the same greater Accra regions revealed that within 5 years, this frequency had reached $freq_{L1014F}$ = 0.54 (95% CI 0.49–0.60; fig. 3). The data from years 2007 and 2008 are reported here for the first time. Phasing of the SNP genotypes of two M-form individuals with a wild-type/*L1014F* genotype showed that the *L1014F*-associated haplotype was identical to that found in the S form. This was confirmed by sequences obtained from seven M-form individuals collected from Accra, Ghana in 2008, which were homozygous for the *L1014F* allele (supplementary table 6, Supplementary Material online). Therefore, the *L1014F* allele, which has increased in frequency in M-form populations, is the same that has been putatively swept toward fixation in sympatric S-form populations. Introgression of *kdr* alleles between forms has been documented previously (Weill et al. 2000) and is unsurprising given that in southern Ghana there is a low but temporally stable level of interform matings (Yawson et al. 2004, 2007).

Using a maximum likelihood estimation procedure with random starting values for selection coefficient (s), dominance (h), and initial allele frequency (p_0), the parameter estimates converged to s = 0.163 (standard deviation [SD] = 0.052), h = 0.249 (SD = 0.142), and initial frequency p_0 = 0.025 (SD = 0.008) (fig. 3).

Discussion

These data show that there is marked LD around *kdr* mutations, loci exhibiting high penetrance, and, for *L1014F* at least, subject to strong recent positive selection. Despite similar median EHH levels, there were differences in the patterns of LD associated with the *L1014S* mutation in Kenya and Gabon. In Kenyan samples, the rate of dissipation of LD around the *L1014S* core was quite rapid suggesting that the mutation has not been subject to as recent or as strong a selective sweep as the same mutation in Gabon (or indeed as the *L1014F* mutation in Ghana). This is as predicted if the serine resistance allele was primarily selected by the use of DDT in the latter part of the 20th century rather than by the more recent use of pyrethroids in agriculture and insecticide control programs. In *Culex* mosquitoes, the equivalent *L1014S* mutation gives low levels of *kdr* to pyrethroids compared with the *L1014F* mutation but confers high levels of DDT resistance (Martinez-Torres et al. 1998; Ranson et al. 2000). Stump et al. (2004) investigated the change in allele frequency of the *L1014S* allele before and after the commencement of a large-scale ITN project in Asembo Bay, Western Kenya, the site of our collections (Stump et al. 2004). The frequency of the *L1014S* allele in the region approximately 10 years before bed net introduction was approximately 0.04 (95% CI 0.02–0.08). In 2002, 15 years after this initial survey and 5 years after the introduction of nets, the frequency of the *L1014S* allele had increased, nonsignificantly to only 0.075 (95% CI 0.05–0.12). This suggests that there is little selective advantage for this

Table 2. Comparison of Median EHH Levels between Alleles at the *kdr* Loci.

	Kenya *L1014S* (S form)	Kenya Wild Type (S form)	Gabon *L1014S* (S form)	Gabon Wild Type (S form)	Gabon *L1014F* (S form)	Ghana Wild Type (M form)
Kenya wild type (S form)	0.0001					
Gabon *L1014S* (S form)	**0.26 NS**	**0.0001**				
Gabon wild type (S form)	0.0001	**0.026NS**	0.0001			
Gabon *L1014F* (S form)	**0.0001**	**0.0001**	**1.00NS**	**0.0001**		
Ghana wild type (M form)	0.005	**1.00NS**	0.005	**0.86NS**	0.005	
Ghana *L1014F* (S form)	**0.0005**	**0.0001**	**0.0001**	**0.0001**	**0.04NS**	**0.0003**

NOTE.—Probabilities from sign tests are shown. The values followed by NS were not significant after sequential Bonferroni corrections. Values that are underlined indicate that the EHH values were significantly higher for the sample given in the column heading; values that are in bold indicate that the EHH values were significantly higher for the sample given in the row heading.

mutation in the present environment, although it should be noted that in a neighboring district in Uganda, a recent study reported that the *L1014S* mutation was at a frequency of 0.85 (95% CI 0.83–0.87) (Ramphul et al. 2009). An alternative explanation would be that in Uganda there is an epistatic interaction between *L1014S* and some, as yet unidentified locus, which may affect the selection, and indeed dominance coefficients, and thereby result in a higher *L1014S* frequency.

The high frequency and marked LD associated with *L1014S* in Gabon may be a result of the co-occurrence in genotypes, though not haplotypes (supplementary table 5, Supplementary Material online), with *L1014F*. A recent study from Cameroon showed that although *L1014F*/*L1014S* heterozygotes were significantly less resistant to permethrin than *L1014F* homozygotes, *L1014F*/*L1014S* heterozygotes were significantly more resistant to all insecticides tested than *L1014F*/*L1014*-wild type heterozygotes (Reimer et al. 2008). Repetitive mutation at the *1014* locus could, at least in part, be responsible for the patterns of LD

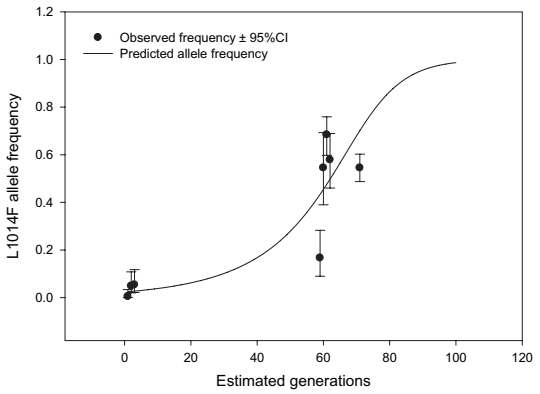

FIG. 3. Observed and predicted changes in *L1014F* allele frequency in the *Anopheles gambiae* M-form populations from southern Ghana. Observed data obtained from surveys conducted in 2002, 2006, and 2007. First collection point (Generation 1) was June 2002. Data from 2002, first three data points, are taken from Yawson et al. (2004); all other data are novel. One generation per month is assumed following Lehmann et al. (1998). The 95% CIs for each observed data point were calculated according to Newcombe (1998). Expected data generated from simultaneous maximum likelihood estimates of initial frequency and selection and dominance coefficients (see Materials and Methods).

around the *kdr* locus in the Gabonese data. Indeed, there is evidence for repeated mutations of *kdr* alleles across the species range of *A. gambiae* (Pinto et al. 2007). However, we argue that on the recent timescales on which *kdr* has arisen and spread it is more parsimonious to assume that recombination is the dominant influence on patterns of LD rather than high rates of repetitive mutations.

Although *kdr* is the best-documented resistance mechanism in *A. gambiae*, there are many other resistance-associated loci. Microarray and recombinant protein expression work has shown that resistant mosquitoes over express a small number of enzymes that catalyze insecticide degradation (Ortelli et al. 2003; Müller et al. 2007; Chiu et al. 2008; Müller et al. 2008). LD-based screens could be a powerful way of identifying regions of the genome carrying the scars of recent selection that regulate such overexpression. However, whether association mapping approaches will effectively identify genes subject to much older and comparatively weaker selection is currently unclear. The bounded estimate of the selection coefficient reported here is at the upper limit of estimates generated to date and of a similar magnitude to estimates generated for resistance alleles in the mosquito *Culex pipiens* (Labbe et al. 2009). In human populations, mutations associated with resistance to malaria infection such as G6PD and sickle cell trait have coefficients of selection of 0.02–0.05 (Tishkoff and Williams 2002) and 0.05–0.18 (Li 1975), respectively. In the third actor in the malaria transmission cycle of *Plasmodium falciparum*, a selection coefficient of 0.1 has been obtained for the locus *dhfr* that confers resistance to the chemotherapeutic agent, pyrimethamine (Nair et al. 2003).

Together with strong and recent positive selection, the major determinant of LD around selected loci will be the rate of recombination. Indications of dramatic variation in the recombination rate across the *A. gambiae* genome have already been reported (Pombi et al. 2006; Black et al. 2008), and it is possible that, being close to the centromere of chromosome 2L, the sodium channel locus is in an area of reduced recombination. However, our Kenyan data are consistent with rates of recombination sufficient to reduce the region hitchhiked with a selectively advantageous locus in a relatively short period of time. Indeed, detection of the signatures of selection for loci with low selection coefficients will be more logistically challenging in *A. gambiae* than humans because of much lower background levels of LD (Weetman D, Wilding CS, Steen K, Donnelly MJ,

unpublished data). We attempted to amplify microsatellites from around the sodium channel to fully define the extent of the swept region as has been done for drug resistance loci in *P. falciparum* (Wootton et al. 2002; Nair et al. 2003). However, the sodium channel is situated in a region with an abundance of repetitive sequences and it was not possible to identify unique locus-specific microsatellite primer pairs.

Given the apparently high selection pressure on the *L1014F* mutation, it is curious that there are no studies, with adequate sample size, that have observed either of the *kdr* alleles at fixation (Santolamazza et al. 2008). One explanation would be that of overdominance; however, insecticide bioassays studies suggest that this is unlikely to be the case (Chandre et al. 2000; Reimer et al. 2008), and our estimate of the dominance coefficient shows the *kdr L1014F* allele to be partially recessive. Therefore, it is likely that there is some fitness cost to the *L1014F* allele and that this could be attributable to heterogeneity in exposure to pyrethroids in the environment or a consequence of an Hill–Robertson effect where selection at a *kdr* locus can interfere with the selection at nearby beneficial mutations (Hill and Robertson 1966).

The data presented herein show that it is possible to detect genomic signatures of strong positive selection in pest species with large effective population size and generally low levels of LD. We suggest that such approaches are likely to extremely powerful in many nonmodel taxa subject to similar selective events.

Supplementary Material

Supplementary tables 1–6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Nelson Cuamba, Eveline Klinkenberg, and Pie Muller for providing wild caught specimens and the associate editor for constructive comments. This work was supported by the National Institutes for Health (U01 AI58542 to Dr Edward Walker, Michigan State University), the Innovative Vector Control Consortium, and the Royal Society.

References

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309:764–767.

Ballinger-Crabtree ME, Black WC IV, Miller BR. 1992. Use of genetic polymorphisms detected by the random-amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) for differentiation and identification of *Aedes aegypti* subspecies and populations. *Am J Trop Med Hyg*. 47:893–901.

Bass C, Nikou D, Donnelly MJ, Williamson MS, Ranson H, Ball A, Vontas J, Field LM. 2007. Detection of knockdown resistance (*kdr*) mutations in *Anopheles gambiae*: a comparison of two new high-throughput assays with existing methods. *Malar J*. 6e:111.

Begun DJ, Aquadro C. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:519–520.

Begun DJ, Holloway AK, Stevens K, et al. (13 co-authors). 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol*. 5:e310.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet*. 74:1111–1120.

Black WC IV, Gorrochetegui-Escalante N, Randle NP, Donnelly MJ. 2008. The yin and yang of linkage disequilibrium: mapping of genes and nucleotides conferring insecticide resistance in insect disease vectors. *Transgenesis and the Management of Vector-Borne Disease*. 627:71–83.

Chandre F, Darriet F, Duchon S, Finot L, Manguin S, Carnevale P, Guillet P. 2000. Modifications of pyrethroid effects associated with *kdr* mutation in *Anopheles gambiae*. *Med Vet Entomol*. 14:81–88.

Chiu T, Wen Z, Rupasinghe S, Schuler M. 2008. Comparative molecular modeling of *Anopheles gambiae* CYP6Z1, a mosquito P450 capable of metabolizing DDT. *Proc Natl Acad Sci U S A*. 105:8855–8860.

Davies TGE, Field LM, Usherwood PNR, Williamson MS. 2007a. A comparative study of voltage-gated sodium channels in the Insecta: implications for pyrethroid resistance in Anopheline and other Neopteran species. *Insect Mol Biol*. 16:361–375.

Davies TGE, Field LM, Usherwood PNR, Williamson MS. 2007b. DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life*. 59:151–162.

Diabate A, Brengues C, Baldet T, et al. (11 co-authors). 2004. The spread of the Leu-Phe *kdr* mutation through *Anopheles gambiae* complex in Burkina Faso: genetic introgression and *de novo* phenomena. *Trop Med Int Health*. 9:1267–1273.

Fanello C, Santolamazza F, della Torre A. 2002. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol*. 16:461–464.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res*. 8:269–294.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 6:65–70.

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160: 765–777.

Labbe P, Sidos N, Raymond M, Lenormand T. 2009. Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long-term cline series. *Genetics* 182:303–312.

Lehmann T, Hawley WA, Grebert H, Collins FH. 1998. The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Mol Biol Evol*. 15:264–276.

Li W. 1975. The first arrival time and mean age of a deleterious mutant gene in a finite population. *Am J Hum Genet*. 27: 274–286.

Livak KJ. 1984. Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* 107:611–634.

Lynd A, Ranson H, McCall PJ, Randle NP, Black WC, Walker ED, Donnelly MJ. 2005. A simplified high-throughput method for pyrethroid knock-down resistance (*kdr*) detection in *Anopheles gambiae*. *Malar J*. 4:16.

Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Berge JB, Devonshire AL, Guillet P, Pasteur N, Pauron D. 1998. Molecular

characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae s.s. Insect Mol Biol*. 7:179–184.

Maynard-Smith J. 1998. Evolutionary genetics. Oxford University Press, Oxford.

Müller P, Chouaibou M, Pignatelli P, Etang J, Walker ED, Donnelly MJ, Simard F, Ranson HW. 2008. Pyrethroid tolerance is associated with elevated expression of antioxidants and agricultural practice in *Anopheles arabiensis* sampled from an area of cotton fields in Northern Cameroon. *Mol Ecol*. 17: 1145–1155.

Müller P, Donnelly MJ, Ranson H. 2007. Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae*-strain from Ghana. *BMC Genomics*. 8:e36.

Müller P, Warr E, Stevenson BJ, et al. (12 co-authors). 2008. Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genet*. 4:e1000286.

Nair S, Williams JT, Brockman A, et al. (12 co-authors). 2003. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol*. 20:1526–1536.

Newcombe RG. 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 17: 857–872.

Ortelli F, Rossiter LC, Vontas J, Ranson H, Hemingway J. 2003. Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*. *Biochem J*. 373: 957–963.

Pinto J, Lynd A, Elissa N, Donnelly MJ, Costa C, Gentile G, Caccone A, Do Rosario VEI. 2006. Co-occurrence of East and West African *kdr* mutations suggests high levels of resistance to pyrethroid insecticides in *Anopheles gambiae* from Libreville, Gabon. *Med Vet Entomol*. 20:27–32.

Pinto J, Lynd A, Vicente JL, et al. (13 co-authors). 2007. Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. *PLoS ONE*. 2:e1243.

Pombi M, Stump AD, Della Torre A, Besansky NJ. 2006. Variation in recombination rate across the X chromosome of *Anopheles gambiae*. *Am J Trop Med Hyg*. 75:901–903.

Ramphul U, Boase T, Bass C, Okedi LM, Donnelly MJ, Müller P. Forthcoming. 2009. Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae*from eastern Uganda. *Trans R Soc Trop Med Hyg*. 103: 1121–1126.

Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FHI. 2000. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol*. 9: 491–497.

Reimer L, Fondjo E, Patchoke S, et al. (11 co-authors). 2008. Relationship between *kdr* mutation and resistance to pyrethroid and DDT insecticides in natural populations of *Anopheles gambiae*. *J Med Entomol*. 45:260–266.

Sabeti PC, Reich DE, Higgins JM, et al. (17 co-authors). 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ESW. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.

Sabeti PC, Varilly P, Fry B, et al. (12 co-authors). 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.

Santolamazza F, Calzetta M, Etang J, et al. (11 co-authors). 2008. Distribution of knock-down resistance mutations in *Anopheles gambiae*molecular forms in west and west-central Africa. *Malar J*. 7:e74.

Schlenke TA, Begun DJ. 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc Natl Acad Sci U S A*. 101:1626–1631.

Scott JA, Brogdon WG, Collins FH. 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg*. 49:520–529.

Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet*. 76:449–462.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68:978–989.

Stump AD, Atieli FK, Vulule JM, Besansky NJ. 2004. Dynamics of the pyrethroid knockdown resistance allele in western Kenyan populations of *Anopheles gambiae* in response to insecticide-treated bed net trials. *Am J Trop Med Hyg*. 70:591–596.

Tishkoff SA, Williams SM. 2002. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. 3:611–621.

Weill M, Chandre F, Brengues C, Manguin S, Akogbeto M, Pasteur N, Guillet P, Raymond M. 2000. The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Mol Biol*. 9:451–455.

Wilding CS, Weetman D, Steen K, Donnelly MJ. 2009. High, clustered, nucleotide diversity in the genome of *Anopheles gambiae*revealed by SNP discovery through pooled-template sequencing: implications for high-throughput genotyping protocols. *BMC Genomics*. 10:e320.

Wootton JC, Feng XR, Ferdig MT, Cooper RA, Mu JB, Baruch DI, Magill AJ, Su XZI. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 418: 320–323.

Yawson AE, McCall PJ, Wilson MD, Donnelly MJ. 2004. Species abundance and insecticide resistance of *Anopheles gambiae* in selected areas of Ghana and Burkina Faso. *Med Vet Entomol*. 18:372–377.

Yawson AE, Weetman D, Wilson MD, Donnelly MJ. 2007. Ecological zones rather than molecular forms predict genetic differentiation in the malaria vector *Anopheles gambiae s.s.* in Ghana. *Genetics* 175:751–761.