An individual-based model of tsetse fly populations dynamics: modelling an extensive mark-release-recapture experiment

by

Roux-Cil Ferreira

Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce in Mathematical Statistics in the Faculty of Economic and Management Sciences at Stellenbosch University

Department of Statistics and Actuarial Science,
University of Stellenbosch.

Supervisors:

Prof. J. W. Hargrove
Prof. S. J. Steel

March 2015

Declaration

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

Signature: .....
R. Ferreira

Date: 2014/11/01 .....

Abstract

An individual-based model of tsetse fly populations dynamics: modelling an extensive mark-release-recapture experiment

R. Ferreira

Department of Statistics and Actuarial Science,
University of Stellenbosch.

Thesis: MCom

March 2015

Tsetse flies (*Glossina* spp), native to mid-continental Africa, are the vectors of trypanosomes that causes human (sleeping sickness) and animal (nagana) trypanosomiasis. Vector control plays a major role in alleviating the burden of the disease. Mathematical models of tsetse population dynamics provide insights into how best to manage these control efforts.

A major mark-recapture experiment, carried out in Zimbabwe, provided valuable information on tsetse population dynamics, but the analyses so far published could be improved on because not all of the information available on the marking procedure was used.

We have constructed an individual-based model that follows the life of individual tsetse flies, their progeny and, in particular, the sequence of occasions on which individual flies were captured and given distinctive marks. We have access to comprehensive data from the tsetse fly mark-release-recapture experiment carried out on Antelope Island, Lake Kariba, Zimbabwe. In order to calibrate or validate the model, we model both the growth of the introduced tsetse population and the mark-recapture process. We have compared the model outputs to the original data and recommend processes that may be followed for model calibration.

It is possible to construct an individual-based model that adequately models tsetse fly populations. Whereas the focus of this study has been on modelling the mark-recapture study, the individual-based model could also be used in more general settings to model the growth, and reduction in fly numbers, changes in age structure, species and gender ratios and the acquisition of trypanosome infections by individual flies. This model can thus be used to inves-

tigate the effect of various factors on tsetse fly and trypanosome, population dynamics as well as on the performance of various control techniques effecting fly mortality and disease transmission.

Uittreksel

'n Individu-gebaseerde model van die tsetsevlieg bevolkingsdinamika: modellering van 'n uitgebreide merk-vrylaat-hervang eksperiment.

("An individual-based model of tsetse fly populations dynamics: modelling an extensive mark-release-recapture experiment")

R. Ferreira

Departement Statistiek en Aktuariële Wetenskap,
Universiteit Stellenbosch.

Tesis: MCom

Maart 2015

Tsetsevlieë (*Glossina* spp), inheems aan sentraalkontinentale Afrika, is die draers van trypanosomen wat trypanosomiasis by die mens (slaapsiekte) en by diere (nagana) veroorsaak. Die beheer van draers speel 'n belangrike rol om die las wat die siekte veroorsaak, te verlig. Wiskundige modelle van tsetse bevolkingsdinamika bied insigte oor hoe om beheerpogings die beste te bestuur.

'n Belangrike merk-hervang eksperiment, wat in Zimbabwe uitgevoer is, bevat waardevolle inligting oor tsetse bevolkingsdinamika. Die ontleding daarvan, wat tot dusver gepubliseer is, kan egter verbeter word aangesien nie al die inligting beskikbaar in die merkprosedure, gebruik is nie.

Ons het 'n individu-gebaseerde model saamgestel wat die lewens van individuele tsetsevlieë en hul nageslagte volg, in besonder die volgorde waarop individuele vlieë gevang en herkenbaar gemerk is. Ons het toegang tot omvattende data van die tsetsevlieg merk-vrylaat-hervang eksperiment wat uitgevoer is op Antelope Eiland, Karibadam, Zimbabwe. Ten einde die model te kalibreer of om die model se geldigheid te bevestig, modelleer ons beide die groei van die ingevoerde tsetse bevolking en die merk-hervangs metode. Ons vergelyk die modeluitsette met die oorspronklike data en beveel prosesse aan wat gevolg kan word om die model te kalibreer.

Dit is moontlik om 'n individu-gebaseerde model saam te stel wat tsetsevliegbevolkings voldoende moduleer. Terwyl hierdie studie die modellering van die merk-hervang data bestudeer, kan die individueel-gebaseerde model

ook gebruik word in meer algemene gevalle vir die modellering van die vermeerdering en vermindering in vlieë getalle, veranderinge in die ouderdomstruktuur, spesies en geslagverhoudings en die verwerwing van trypanosomen infeksies deur individuele vlieë. Hierdie model kan dus gebruik word om die effek te ondesoek van verskeie faktore op die tsetsevlieg en trypanosomen, populasiedinamiek sowel as die prestasie van verskillende beheertegnieke rakende vliegsterftes en siekte-oordrag.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations. SACEMA for providing funding and general support for this research. My supervisors Prof. J. Hargrove and Prof. S. Steel for their guidance and Mr. P. Labuschagne for performing the code reviews. This investigation received financial support from the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR).

Contents

List of Figures

List of Tables

Glossary of entomological terms used in this thesis

Oocyte An immature egg.

Spermathecae A female organ which receives and stores sperm from the male.

Ovarian category The number of times that the female has ovulated.

Instar Denotes the developmental stage of the larval.

Larviposition The act of depositing a larva by an adult female tsetse..

Pupa The life stage during which the larva undergoes the transformation into the adult stage.

Teneral The period when the adult insect is newly emerged from the pupal case and has not yet taken its first bloodmeal. During the teneral period, the insect's exoskeleton has not yet hardened, the flight muscles are not fully developed and it has low fat reserves.

Inter-larval period The length of time between the production of two consecutive larvae by the same female.

Individual-based model The term used for individual-based models in fields other than ecology is agent-based models.

Chapter 1

Introduction

In late 1979 Glyn Vale introduced two species of tsetse flies (*Glossina morsitans morsitans* Westwood and *G. pallidipes* Austen) onto Antelope Island, Lake Kariba, Zimbabwe. He thereby set in train in Zimbabwe what was to be one of the most intensive mark-recapture exercises ever carried out on any insect. The idea was to grow large populations of the two tsetse species in the island environment, which was shown to be effectively closed to in and out migration. The populations could then be used to test the efficacy of odour-baited trapping devices and "targets" (insecticide panels of fabric) as candidates for the cost-effective control, and even eradication, of tsetse flies.

In order to monitor the quantitative effects of the various control techniques it was necessary to be able to follow temporal changes in the number, and mortality and birth rates, of both genders of both tsetse species. This was achieved using multiple mark-release-recapture techniques following the experimental and analytical methodologies developed, independently, by Jolly (1965) and Seber (1965). Briefly, tsetse captured from bait-oxen were marked to indicate the week in which the fly had been captured. If the fly was captured again in a later week it was marked with a different colour. The results were originally described by Vale *et al.* (1986) and the temporal changes in the various population parameters for *G. m. morsitans* have been modelled by Hargrove and Williams (1998) and the data have also been used to investigate the factors affecting adult tsetse mortality (Hargrove, 2001*a*,*b*,*c*).

While the data have thus already been used to good effect, it is evident that the analysis of the original mark-release-recapture data, and the modelling of the resulting population parameter estimates can be improved upon. The Jolly-Seber (J-S) estimates (Section 2.3.2) of population parameters involved the recording and use of only the most recent mark on any recaptured fly. It is intuitively clear that, since some flies bore marks of six or more different colours, much information about time-related changes in mortality and capture probability was lost.

The J-S analytical procedure in fact involves making a number of explicit simplifying assumptions two of which are:

(i) Every animal in the population - old, young, marked and unmarked - has the same probability of being captured in a sample, given that it is alive and in the population when the sample is taken.

(ii) Every marked animal has the same probability of surviving between samples, given that it is alive and in the population immediately after release.

When the Antelope Island data was first analysed there was no unequivocal evidence to reject these assumptions. Since then, however, it has become clear that both survival and capture probability change with fly age (Hargrove *et al.*, 2011; Hargrove, 1990, 1991, 1992*a*). Taking these changes into account complicates the data analysis. Nonetheless, to gain a complete understanding of tsetse fly populations we need to make sense of the effect of these changes with age on the population dynamics. The structural richness provided by an individual-based model is the only way to combine the changes that are taking place in the tsetse fly biology and the full recapture history of flies. The aim of this work is to develop an individual-based stochastic micro-simulation model, to simulate the capture-recapture experimental protocol in all of its detail.

In particular, the aim is to follow the progress of individual flies through their entire lives, recording the occasions when they are captured and marked and building up, thereby, a picture of the way in which the distribution of colour markings changes as a function of time in the simulated population. The general idea is then to compare the distribution of colour markings among samples of flies in the simulated population with the colour distribution observed in the actual samples from the population in the real experiment. The cost function ("goodness-of-fit") used to estimate the unknown parameters could be calculated using a measure of "distance" between the weekly colour distributions of the trapped flies in the experiment and the weekly colour distributions of the sampled flies from the simulation.

There are several advantages to this approach compared to the Hargrove and Williams (1998) approach of fitting the time evolution of the size of the simulated population to the time evolution of the size of the population estimated using the capture-recapture formula:

1. All of the available experimental data are used to estimate the parameters. This may help to estimate/identify certain parameters (for example, effect of humidity) that were not identified in the Hargrove and Williams (1998) paper.

2. Assumptions implicit in the capture-recapture formulation are dropped.

3. The capture-recapture formulation has significant noise, as it is based on several random variables. A series of capture-recapture estimates on a single population should have significant serial correlation which could be used to refine the population size estimates, but the mathematics likely

gets rather complicated, since an underlying model of the population dynamics would need to be included. The micro-simulation approach bypasses the mathematics directly.

4. The experimental protocol can be represented directly without approximation.

In the present study the following objectives are met:

1. Understanding of the requirements to build an individual-based model for a tsetse fly population (Chapter 2).

2. Description of the mark-release-recapture data, the model and the software reliability tests (Chapter 3).

3. Running the individual-based model using parameters from relevant literature (Chapter 4).

4. Describing further research in individual-based models for tsetse fly populations (Chapter 5).

Chapter 2

Review

2.1 Introduction

The individual-based approach is a bottom-up approach to the modelling of biological systems which starts with the components of a system and then tries to understand how the properties of the system emerge from the interaction of the components. Finding the right level of complexity to describe the components of an individual-based model is a major task and difficulty when building the model (Grimm and Railsback, 2005), which involves determining the variables, processes and emergent processes that should be included in the model. For example, in the model of the mark-recapture experiment, carried out on Antelope Island, the tsetse flies (components) make-up a population (system) and one of the properties of the system is the distribution of coloured marks observed in the mark-recapture samples. In this particular model the right level of complexity to describe the life processes of the tsetse fly needs to be considered.

Describing the components as realistically as possible is one possible approach. However, it soon becomes clear that "realism" is a poor guideline for modelling since taking into account all the elements of the real world which influence a system will result in unnecessarily complex models and the model will never be completed because there are always more details to add. Moreover, we do not know everything about the system. A better approach is to include only the details that are essential to solving the problem or question at hand, but no more than that (Grimm, 1999).

The matter was put succinctly by Albert Einstein who said:

> It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

The problem to be solved provides a filter which should be passed only by those elements of the real system considered essential to understanding the

problem (Grimm, 1999). A clear definition of the purpose of the model is, therefore, crucial. In the particular case of the tsetse mark-recapture model the aim is to identify functions that describe the processes of survival, birth and capture. These functions can then be used to explain distributions of coloured marks observed in mark-recapture samples of tsetse captured in the field. To identify these functions familiarity with the tsetse life cycle (processes) is necessary.

2.2 Tsetse Biology

2.2.1 Introduction

Tsetse include all the species in the genus *Glossina*, which are placed in their own family, Glossinidae. There are about twenty-nine species and sub-species depending on the classification used. Tsetse flies are a native of mid-continental Africa. Both genders feed exclusively on blood which provides both nutrients and water to fulfil the needs of the fly. Tsetse are the vectors of trypanosomes, which affect both human health and agriculture through the transmission of human (sleeping sickness) and animal (nagana) trypanosomiasis.

Tsetse have been extensively studied because of their importance in disease transmission. Because flies can be raised in a laboratory, and because they have been much studied in the field, their life cycle is well understood by biologists.

Tsetse reproduce by way of a highly specialised process termed adenotrophic viviparity. That is to say, tsetse give birth to live young. The female tsetse has paired ovaries with two ovarioles in each ovary. Oocytes (immature eggs) mature singly in the ovarioles in the strict order: right inner, left inner, right outer, left outer (Figure 2.1). The mature oocyte is ovulated and then retained as an egg in the uterus, having been fertilized during ovulation by sperm stored in the spermathecae. Female tsetse generally only mate once. At about five days post-ovulation the egg hatches in the uterus to produce a first-instar (stage) larva which is fed a mixture of fat and protein from a modified adrenal gland, the so-called "milk gland" in the uterus.

Over the next 4-5 days, depending on temperature, the larva completes the development of three instar stages. The late third instar larva is deposited (larviposition), generally on a substrate of loose soil, sand or vegetable litter. Having deposited the larva, females ovulate the next oocyte and start the incubation of the next larva. Ovulations, therefore, occur at regular intervals (inter-larval period), of about 9 days at 27°C. The newly independent tsetse larva burrows into the substrate to a depth of a few centimetres, and then forms a hard outer shell called the puparial case, in which it completes its morphological transformation into an adult fly.

The duration of the puparial phase varies in a non-linear fashion with temperature: at 30°C the puparial duration is about 20 days, at 25°C it is 26

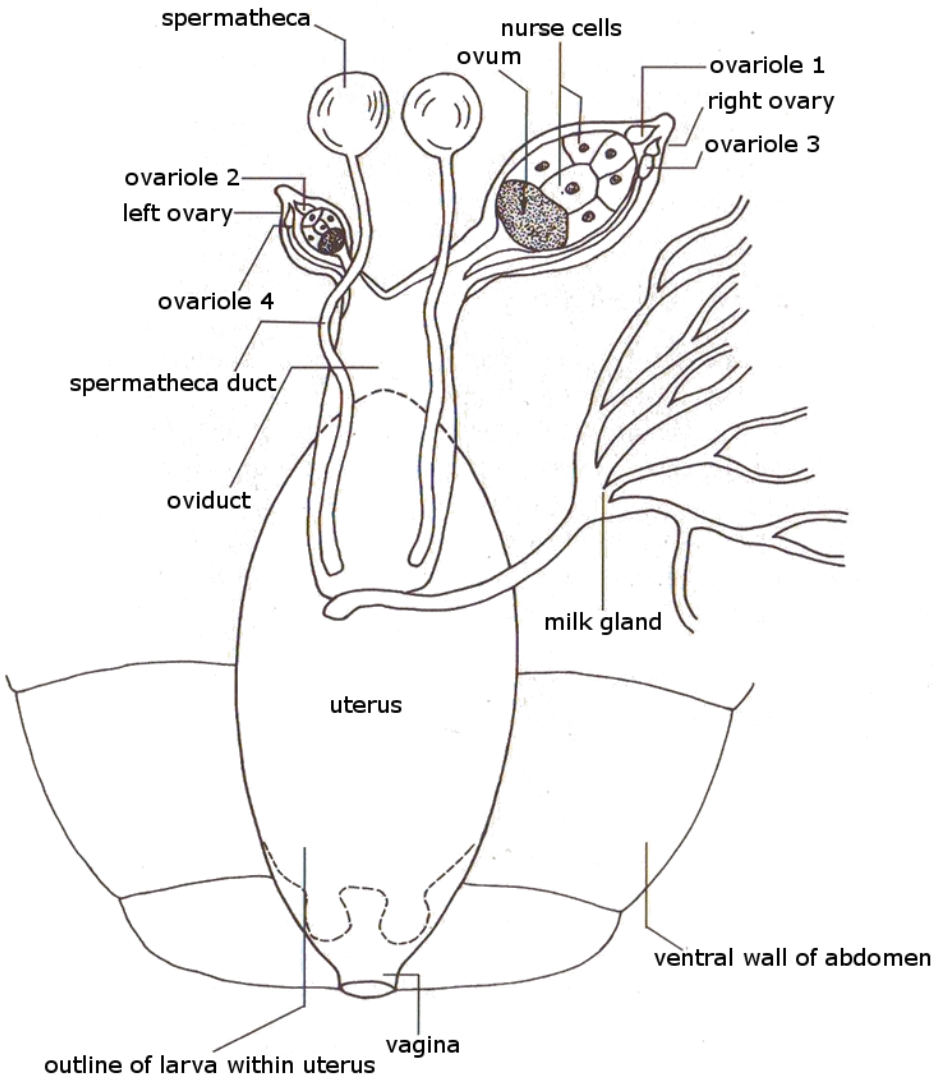


Figure 2.1: The reproductive system of the female tsetse (*Glossina spp*).

days and at 20°C it is 47 days. The pupa does not feed during this time, relying entirely on stored resources. Since the adult that emerges from the puparium has the full linear dimensions of the mature adult, it is not surprising that the fully developed third-instar larva weighs almost as much as its mother.

While the nutritional contents of the pupa is sufficient to produce an adult, the young unfed fly (teneral) which emerges has smaller fat reserves and less well developed flight musculature than the mature fly. Before the emerging adult female embarks on reproduction she uses the first three or so blood meals to rectify this situation. As a consequence the time to the production of the first larva is nearly twice as long as the time between the production of subsequent larvae. Moreover, the massive inputs of raw material required by the larva means that only one larva can be produced at a time.

The dependence of the different life processes on temperature have been studied and structural equations for predicting the duration of various stages at different temperatures are available. The results of these studies are used to inform the survival, birth and capture functions used in the individual-based model developed here to explain distributions of coloured marks observed in mark-recapture samples of tsetse captured in the field.

2.2.2 Reproduction rates

The reproductive rate in tsetse depends on the rate of production of larvae/pupae and on the rate at which those pupae develop into adults. Both rates are dependent on temperature. Key studies focused on larval development and puparial duration are summarised below.

Rates of larva production

Larval production technically comprises oocyte formation, ovulation and fertilization, development of the egg and three larval stages. Hargrove (1994, 1995) studied these processes in detail, specifically focusing on the inter-larval period (the length of time between the production of two consecutive larvae). The oocytes grow in a regular way during pregnancy as has been established through the dissection of laboratory flies of known age or, more importantly, from experiments using flies released into the field bearing marks indicating their day of emergence from the puparium.

The studies were carried out in Zimbabwe on Redcliff Island, Lake Kariba, and at Rekomitjie Research Station in the Zambezi Valley. Marked *G. pallidipes* and *G. m. morsitans* females were released within 12h of their emergence as adults. Ovarian dissections were performed on recaptured flies. The dissector determined the ovarian category (the number of times that the female had ovulated) - using the method described by Challier (1965) - as well as the length of the two largest oocytes and any uterine inclusions, and the contents of the spermathecae.

The lengths of the oocytes grew approximately exponentially for *G. pallidipes* and *G. m. morsitans* (Hargrove, 1995). The largest oocyte grew to full size during the first 5-6 days of adult life and then stopped growing. The second largest oocyte grew at a low rate during the first ovarian cycle: after the largest oocyte was ovulated, the growth rate increased, but was still lower than the growth rate of the largest oocyte during the first ovarian cycle. The length of the larva increased approximately exponentially during the last half of pregnancy.

The lengths of the two largest oocytes, and/or the length of a larva in the uterus, if a larva was present, were used to estimate the proportion of the current pregnancy completed and the ovarian category defined the number of ovulations (and pregnancies) completed. Given this information and the known age of the recaptured marked fly it was possible to estimate the inter-larval period (I days). The curve given by Equation 2.1 was fitted to the estimates of inter-larval period over different temperatures, independently for the first larva and subsequent larvae. The parameter k_1 represents the rate at which larva would develop if a constant temperature of 24°C was applicable. The parameter k_2 adjusts the rate of larva development for temperatures other than 24°C.

$$I = \frac{1}{k_1 + k_2(T - 24)} \tag{2.1}$$

Estimates of the insemination rate were obtained from ovarian category zero (have not ovulated) wild flies, sampled at Rekomitjie Research Station (Hargrove, 2012). The ages of these flies were estimated from the lengths of the largest and second largest oocytes. Insemination rates increased exponentially with age in both species; 90% of females were inseminated after 5 days post-emergence for *G. m. morsitans* and 7 days for *G. pallidipes*, varying little with season. More than 95% of both species had ovulated by the age of 8 days and 99% by 12 days. The Hargrove (1994, 1995, 2012) studies raise questions about the effect of temperature on the rates of various metabolic processes in tsetse flies. They show that it is clearly dangerous to infer, from laboratory measures, the effect of temperature on metabolic rates in the field.

Table 2.1: Parameter estimates resulting from fitting the model given in Equation 2.1 to the data of inter-larva periods obtained by Hargrove (1994).

Larvae category	Parameter	Value	Standard Error
First	k_1	0.061	0.002
	k_2	0.002	0.0009
Subsequent	k_1	0.1046	0.0004
	k_2	0.0052	0.0001

This complete description of larvae production is not used in the individual-based model developed here. Only the simple formulation in Equation 2.1 is used to describe oocyte formation, ovulation and development of the eggs into a third-instar larvae. Since the female is generally inseminated only once and insemination rates are high, we assume in our modelling that all female flies are inseminated and do not explicitly model the timing of this. The estimated base-line parameters used in the individual-based model are presented in Table 2.1. The parameters are fixed at these values for all females, hence, there is no variation between females.

Puparial duration

Phelps and Burrows (1969*b*) produced the most complete and precise measure of the effect of temperature on the duration of the pupal phase. A logistic curve was fitted to the rate of development of pupae kept at constant temperatures between 8 and 32°C, independently for males and females. The logistic curve is given by:

$$r = \frac{k}{1 + exp(a + bT)} \tag{2.2}$$

where r is the rate of development (r per day), which is the reciprocal of the puparial duration (I_p days). The prediction of the puparial duration - reciprocal of the fitted curve - is therefore exponential with a non-zero asymptotic duration ($1/k$) at high temperatures. The most important source of variation in their model was between individual pupae kept at the same constant temperature. The variation was between 3 and 4 days.

Phelps and Burrows (1969*b*) noted a peculiarity that suggests conditioning during the larva stage within the female which effects puparial development. The effect was seen when the pupal duration of two cohorts of pupae were compared. The only difference between the cohorts was the temperatures experienced by the mothers during gestation. Pupa deposited by females held at 20°C had puparial periods that were about 2 days longer than those deposited by females kept at 25°C when both batches of pupae were maintained at 20°C. Curiously there was no significant difference between the pupal durations when both batches were maintained at 25°C.

In a related study Phelps and Burrows (1969*a*) used the average of maximum and minimum Stevenson screen temperatures in Zimbabwe to produce predictions of I_p for pupae deposited in the field. These predictions generally followed the seasonal changes in ambient temperature in the field. The observed variation in I_p was higher due to the variation in mean temperatures between different types of deposition sites.

This complete description of the puparial duration is not used in the individual-based model developed here. Only the simple formulation given in Equation 2.2 is used. The estimated base-line parameters used in the model are presented in Table 2.2. The parameters are fixed at these values for all

Table 2.2: Parameter estimates resulting from fitting the model given in Equation 2.2 to the data of puparial duration obtained by Phelps and Burrows (1969*b*) over the temperature range 16 and 32°C.

Gender	Parameter	Value	Standard Error
Males	k	0.053	0.001
	a	5.3	0.2
	b	-0.24	0.001
Females	k	0.057	0.001
	a	5.5	0.2
	b	-0.25	0.001

male pupae and female pupae, hence, there is no variation between pupae of the same gender.

2.2.3 Mortality

Tsetse fly mortality is not as easily estimated as the reproduction rates and it varies with the development stage. To identify functions for mortality the different stages are considered separately.

In utero mortality

Hargrove (1999) investigated abortion rates by performing ovarian dissections on *G. m. morsitans* and *G. pallidipes* captured at Rekomitjie Research Station, Zambezi Valley, Zimbabwe. Estimated abortion rates varied between 0.8-4.5% in *G. m. morsitans* and 0.3-2.8% in *G. pallidipes*. The abortion rates increase to >2% when mean temperatures exceeded 27°C. Ovarian category had little effect on the abortion rate, but the frequency of empty uteri declined markedly with age - with a suggestion, however, that it might increase again in the oldest flies.

The Hargrove (1999) study indicates that *in utero* losses are not a major source of loss. Abortions are, therefore, ignored in the individual-based model developed here.

Pupal mortality

Pupae are generally not easily found in the field and estimation of their numbers in a population is a problem which has so far defeated tsetse biologists. There are, consequently, few estimates of the natural death rate in pupae but parasitism and predation may nonetheless be severe.

The loss rate in the individual-based model developed here is taken as 1% per day. The longer the puparial period the lower the number of pupae emerging as adults.

Fly mortality

One would expect newly emerged tsetse to have a greater risk of starvation than fully mature adults since they have not taken their first bloodmeal, have very low fat reserves and have undeveloped flight musculature. The greater risk of starvation results in a higher mortality among very young flies. On the other hand, in certain seasons of the year, flies live long enough to die of old age. The relationship between age and mortality is therefore complicated. Hargrove (1990) provided support for the expectation that tsetse have a U-shaped age-mortality curve. Males showed a higher loss rate than females at every age. The pattern in male mortality was similar to that of females, but with the time scale compressed.

Hargrove *et al.* (2011) fit a three-parameter curve to the same data used by Hargrove (1990). The survival curve is given in Equation 2.3 and was independently fitted for males and females:

$$\phi(A) = exp[k_1\big(exp(-k_2A) - exp(k_3A)\big)] \tag{2.3}$$

where $\phi(A)$ estimates the probability of a fly surviving to age A days. The parameter k_1 determines the overall level of mortality, k_2 the rate of the losses in roughly the first week of the fly's life and k_3 the losses due to ageing.

The equation presented in Equation 2.3 does not consider the change in mortality with change in the environment. In studies that do consider environmental factors, it is not generally agreed which factors are most important in determining density-independent mortality.

Hargrove (2001*a*) pointed out that mark-release-recapture, and ovarian dissection, estimates of mortality among adult flies found that temperature was the most important determinant of mortality. By contrast, when Rogers (1979) used the Moran curve method to estimate losses between successive generations, measures of dryness such as saturation deficit seemed to be more important. These two lines of evidence can be reconciled if the mortality among pupal and teneral adult stages are primarily affected by dryness.

Table 2.3: Parameter estimates resulting from fitting the model given in Equation 2.3 to the data obtained by Hargrove (1990) as shown in Hargrove *et al.* (2011).

Gender	Parameter	Value	Confidence Interval	Base-line Value
Males	k_1	0.389	0.295 to 0.518	0.01945
	k_2	0.395	-0.0130 to 0.802	0.01975
	k_3	0.0583	0.0425 to 0.0741	0.002915
Females	k_1	0.605	0.553 to 0.656	0.03025
	k_2	0.201	0.141 to 0.262	0.01005
	k_3	0.0119	0.0106 to 0.0132	0.000595

In the individual-based model developed here both ageing effects and temperature effects are considered. The proposed model is an adaptation of the Hargrove *et al.* (2011) equation; allowing the fixed parameters k_1, k_2 and k_3 to vary linearly with temperature. The parameter estimates obtained by Hargrove *et al.* (2011) are presented in Table 2.3. We see that the estimated values for k_1 and k_3 are larger for males than for females and the estimated value for k_2 is smaller for males than for females meaning that the mortality for males is higher over all age ranges compared to females. This tendency is observed in laboratory-bred tsetse flies (Maudlin *et al.*, 1998), but it is amplified in field tsetse due to the differences in behaviour between the genders, which puts males at additional risk. Thus, while female tsetse only approach animals to feed, males approach animal hosts both for purposes of feeding and mating with virgin female who approach the animal to feed, thereby putting themselves at increased risk of predation.

The estimated base-line parameters used during simulation were then obtained by dividing these estimates of k_1, k_2 and k_3 by 20, because the average daily temperature during the Hargrove (1990) experiment was 20°C and in the present study it is assumed that these parameters vary linearly with temperature. The base-line parameters are also presented in Table 2.3.

2.2.4 Capture

Using bait-oxen as a method to capture samples of tsetse flies relies on flies being attracted to the oxen before they can be part of the samples. If sampling is random all the tsetse in the population would have the same probability of being included in the sample. We examine the assumption that the fly-round method of sampling produces a random sample of tsetse.

In examining the number of *G. m. morsitans* captured during the Vale *et al.* (1986) experiment (Figure 2.2), one notices that there is a gender bias in the samples. The samples contain more males than females, especially for *G. m. morsitans*. Hargrove (1990) investigated whether there is a bias within the females and male groups. He found that the age of a tsetse fly influences its capture probability. The influence is nonlinear.

In the present study gender and age are the only factors that affect capture probability. The capture probability is the same for tsetse of the same gender and age. The estimated starting capture probabilities are those presented in Hargrove (1990) shown in Figure 2.3.

2.3 Tsetse Population Models

2.3.1 The Hargrove and Williams (1998) model

Hargrove and Williams (1998) described a method for optimizing a simulation

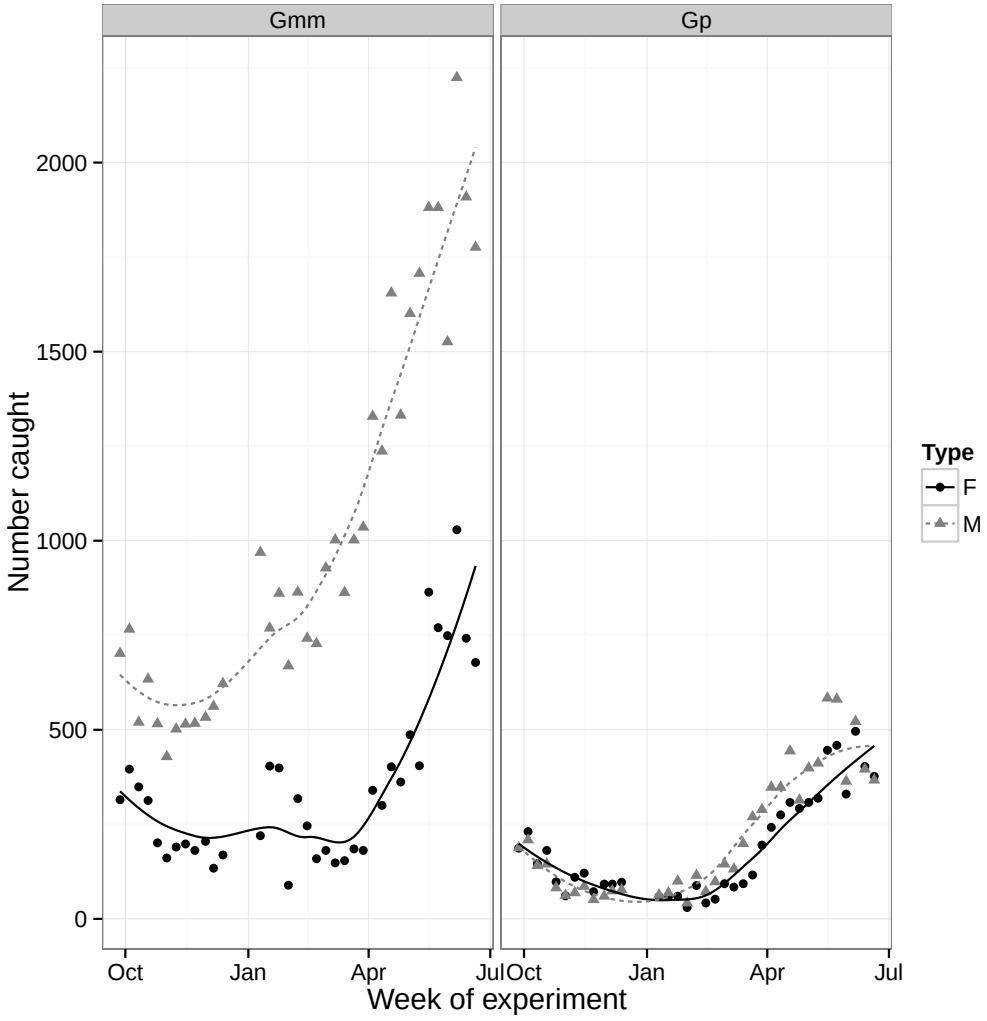


Figure 2.2: The number of male (M) and female (F) *G. m. morsitans* (Gmm) and *G. pallidipes* (Gp) caught per week in the field on Antelope Island. Samples were carried out between 24 September 1980 - 23 June 1981. The fitted curve is a local linear regression (loess) curve (see A.1).

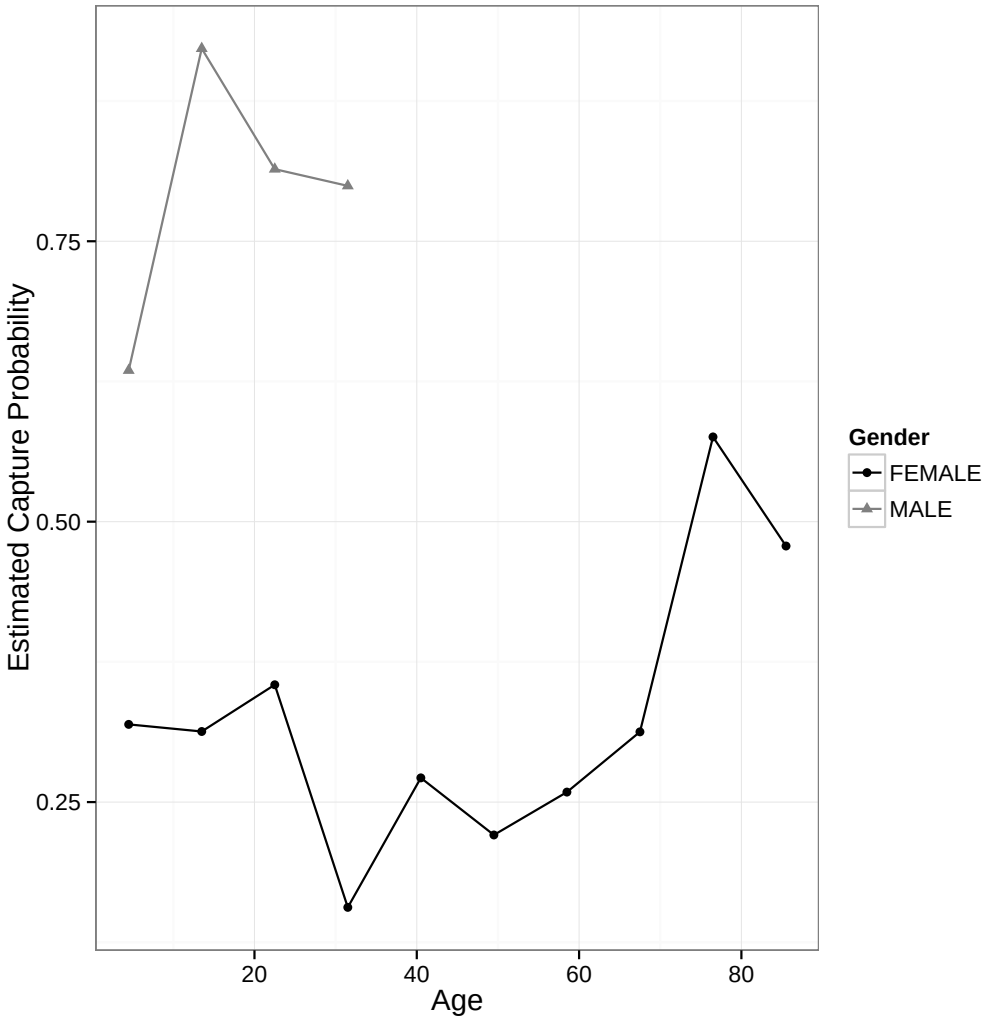


Figure 2.3: The estimated probability that a male (triangles) or female (dots) *G. m. morsitans* was captured at least once on an ox fly-round during successive 9-day periods. As in Hargrove (1990).

model by linking the simulation procedures to an iterative minimization routine. The optimized simulation technique automates the process of estimating the parameter values - which give the best fit to a data set - given a particular model formulation. The technique is use to fit a tsetse fly simulation model to the Vale *et al.* (1986) series of weekly Jolly-Seber estimates of total numbers, births and survival probabilities.

Simulation involved following the lives of cohorts of flies, and of all their progeny, from larviposition until all flies in the cohort died. The puparial duration and larva production were regarded as fixed functions of temperature. The best fits for various formulations of the mortality functions in both pupae and adult flies were obtained using this technique. The best fits for the different formulations were compared to obtain the final model. The final model formulation included the effect of maximum temperature on adult survival, the various modes of trapping, and an annual cycle.

It was clear to Hargrove and Williams (1998) that their model was deficient in several important aspects - in particular, they were unable to detect any density-dependent effects on mortality or birth and they could not find any effect of relative humidity (or saturation deficit) on adult or immature mortality, whereas on general grounds and even from other analyses of the data one might have expected both effects to occur.

2.3.2 Jolly-Seber model

The technical details that follow for mark-release-recapture analysis is based on Jolly (1965) and Seber (1965).

Consider a sequence of s samples of sizes $n_1, n_2, \cdots, n_s$. For each sample, only R_i of the n_i are marked and returned to the population. This allows for accidental deaths due to marking and handling or in commercial exploitation where the sample is permanently removed from the population.

Model assumptions:

1. Every animal in the population, whether marked or unmarked, has the same probability $p_i(= 1 - q_i)$ of being caught in the ith sample, given that it is alive and in the population when the sample is taken.

2. Every animal has the same probability ϕ_i of surviving from the ith to the $(i + 1)$th sample and of being in the population at the time of the $(i+1)$th sample, given that it is alive and in the population immediately after the ith release $(i = 1, 2, \ldots, s - 1)$.

3. Marked animals do not lose their marks and all marks are reported on recovery.

4. All samples are instantaneous, *i.e.* sampling time is negligible, and each release is made immediately after the sample.

Model notation:

t_i = time when the ith sample is taken

N_i = total number of animals in the population just before time t_i

M_i = total number of marked animals in the population just before time t_i

U_i = $N_i - M_i$ total number of unmarked animals in the population just before time t_i

n_i = number caught in the ith sample

m_i = number of marked animals caught in the ith sample

u_i = $n_i - m_i$ number of unmarked animals caught in the ith sample

m_{hi} = number caught in the ith sample last captured in the hth sample $(1 \leq h \leq i-1)$

R_i = number of marked animals released after the ith sample

r_i = number of marked animals from the release of R_i animals which are subsequently recaptured

z_i = number of different animals caught before the ith sample which are not caught in the ith sample but are caught subsequently

B_i = number of new animals joining the population in the interval from time t_i to time t_{i+1}, which are still alive and in the population at time t_{i+1}

The following intermediary parameters are used:

$$\begin{aligned}
&\alpha_i = \phi_i q_{i+1}, \beta_i = \phi_i p_{i+1} \text{ and} \\
&\chi_i = (1-\phi_i) + \phi_i q_{i+1}(1-\phi_{i+1}) + \cdots + \phi_i q_{i+1} \ldots \phi_{s-1} q_s \\
&\quad = 1 - \phi_i p_{i+1} - \phi_i q_{i+1} \phi_{i+1} p_{i+2} - \cdots - \phi_i q_{i+1} \ldots \phi_{s-2} q_{s-1} \phi_{s-1} p_s \\
&\quad = 1 - \beta_i - \alpha_i \beta_{i+1} - \cdots - \alpha_i \alpha_{i+1} \ldots \alpha_{s-1} \beta_s,
\end{aligned}$$

χ_i being the conditional probability that a marked animal released in the ith release of R_i animals is not caught again.
Let ω be a non-empty subset of the set of integers $\{1, 2, \ldots, s\}$ and let a_ω be the number of marked animals with the capture history ω. Suppose that on the last occasion when the group of a_ω animals is captured, d_ω are not returned to the population. Let $b_\omega = a_\omega - d_\omega$, then the conditional distribution of the random variables $\{b_\omega, d_\omega\}$ conditional on the $\{u_i\}$ is given by

$$f(\{b_\omega, d_\omega\}|\{u_i\}) = \frac{\prod_{i=1}^{s} u_i!}{\prod_\omega \{b_\omega! d_\omega!\}} \prod_{i=1}^{s-1} \left\{ \chi_i^{R_i - r_i} \alpha_i^{z_{i+1}} \beta_i^{m_{i+1}} \right\} \times \prod_{i=1}^{s} \left\{ \nu_i^{R_i} (1-\nu_i)^{n_i - R_i} \right\}. \tag{2.4}$$

Since sampling of unmarked animals in the population is binomial, we have

$$f(\{u_i\}) = \prod_{i=1}^{s} \left\{ \binom{U_i}{u_i} p_i^{u_i} q_i^{U_i - u_i} \right\}. \tag{2.5}$$

Finally, the unconditional distribution is given by

$$f(\{b_\omega, d_\omega\}) = f(\{b_\omega, d_\omega\}|\{u_i\}) f(\{u_i\}) \tag{2.6}$$

Differentiating the logarithm of Equation 2.6 results in the maximum likelihood estimates of the parameters (Seber (1965) for the special case of no losses on capture). However, the same set of estimates can also be obtained by an intuitive argument (Jolly, 1965). The estimates are given by:

$$\begin{aligned}
\hat{p}_i &= \frac{n_i}{\hat{N}_i} = \frac{m_i}{\hat{M}_i} \\
\hat{\phi}_i &= \frac{\hat{M}_{i+1}}{\hat{M}_i - m_i + R_i} \\
\hat{v}_i &= \frac{R_i}{n_i} \\
\hat{N}_i &= \frac{\hat{M}_i n_i}{m_i} \\
\hat{M}_i &= \frac{R_i z_i}{r_i} \\
\hat{B}_i &= \hat{N}_{i+1} - \hat{\phi}_i(\hat{N}_i - n_i + R_i) \\
\hat{U}_i &= \hat{N}_i - \hat{M}_i
\end{aligned}$$

To calculate the above, estimates of m_i, r_i and z_i are required. Let

$$c_{ij} = \sum_{h=1}^{i} m_{hj} \text{ for } i < j, \tag{2.7}$$

where c_{ij} is the number of animals in the jth sample last caught in the ith or before. Then

$$m_i = c_{i-1,i} \tag{2.8}$$

$$r_i = \sum_{j=i+1}^{s} m_{ij} \tag{2.9}$$

$$z_i = \sum_{j=i+1}^{s} c_{i-1,j} \tag{2.10}$$

These can now easily be calculated from arrays of the m_{hi} and c_{hi} and used to estimate the population parameters.

Chapter 3

Materials and Methods

3.1 Data

3.1.1 Data collection

At the start of the Vale *et al.* (1986) experiment (August 1979) cattle were introduced onto Antelope Island (area ≈ 5.4 km^2) in Lake Kariba, Zimbabwe, to sustain the tsetse population that was introduced between August and October 1979. The founder individuals were 2000 pupae of *G. m. morsitans* and *G. pallidipes* that were collected from natural breeding sites near Rekomitjie Research Station, about 100 km from Lake Kariba. The pupae were placed on the sandy floors of disused burrows of the antbear (*Orycteropusafer*), or aardvark, on the island. Antbear burrows are natural larviposition sites for tsetse. The exact ages of the founder individuals were unknown and have to be estimated during the modelling process. The weekly mark-release-recapturing was started in September 1980, approximately 400 days after the introduction of the founder individuals.

The individual-based model replicates the mark-release-recapture data for male and female *G. m. morsitans* sampled on the island. The flies were captured, and the marking data thereby gathered, by monitoring teams conducting fly rounds. A monitoring team consisted of two or three men leading an ox along a 3 km section, of the 9 km, of road on the island. The section of road they travelled along was chosen such that each section was travelled evenly each week. During each fly round the monitoring team made stops at marking stations that were 100 m apart. Tsetse attracted to the ox were captured with hand nets while the team was moving between, or stopped at, a station. The gender, species and marks on individual flies were noted each time they returned to feed on the bait oxen. The tsetse were released after a mark had been placed on the fly. The mark was placed at one of 10 distinct positions on the dorsal thoracic surface (Figure 3.1). The colour and position of each paint dot denoted the week of marking. Flies that were caught that already bore the mark of the current week, were ignored. Two fly rounds were

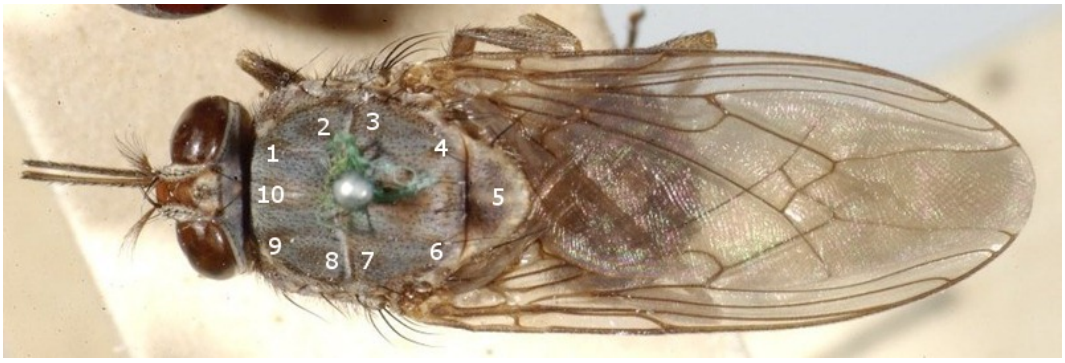


Figure 3.1: Ten mark locations on the Tsetse Fly torso

conducted each day for ≈ 3 hours in the morning and late afternoon. Marking continued, with some breaks, until April 1983, by which time a trapping-out exercise had been used to eliminate the tsetse populations and no tsetse could be captured.

The tsetse population was ultimately eliminated due to the additional mortality imposed by the traps and targets placed on the island. Either odour-baited traps or "targets" were tested as possible tsetse control methods from April 1981 until the eradication of the population. The data used here is a subset of the mark-release-recapture data collected during late 1980 and 1981, when little or no trapping was carried out. It includes the period when sterilizers were deployed since these devices did not retain the flies and there is no evidence that they caused a decrease in survival probability of *G. m. morsitans* (Hargrove, 2001*a*).

3.1.2 Data base

Prior to the present study, researchers only utilised the most recent mark identified on any recaptured fly, hence only the most recent mark was recorded in the electronic data bases and used in the subsequent analysis of mark-recapture data (Vale *et al.*, 1986; Hargrove and Williams, 1998). The aim of the present study was to gain greater insight into the details of tsetse population dynamics through the utilisation of all of the information encapsulated in all marks on every fly. It was, therefore, necessary to generate a new database containing the full mark-recapture histories for each fly. The data are stored on Google drive, and are accessed through downloading the spread sheet files. The design allows easy data entry, access and analysis since all parties involved can access the latest version, and version control is fully maintained by Google drive. The tsetse fly data base records the information from the field report forms. There is a row for each mark-recapture history considered. Apart from the column containing the history there are seven other columns (see Table 3.1).

Table 3.1: Data extract

Column name	Type	Descriptive name	Valid values	Description
Volume.Number	integer	Volume Number	3-5	This field links the records to the volume of the field report.
Mark	char(2)	Mark	R1, W1, B1, O1, Y1, G1, R2, W2, B2, O2, Y2, G2	A one letter, one number code for the colour and position of the mark place on the fly before it was released.
Flyround	integer	Flyround location	1-3	The section of road travelled during the flyround
ampm	char(2)	Flyround time	am, pm	Indication of whether the flyround was in the morning or afternoon
Date	date	Date	1981-01-01 to 1983-01-01	The date of the flyround
Species	char(2)	Species	MO, PA	A two letter code for the species of the individual. MO - *G. m. morsitans* PA - *G. pallidipes*
Type	char(2)	Gender and young or old	OM, OF, YM, YF	A two letter code for the type of the individual. OM - Old male OF - Old female YM - Young male YF - Young female
History	string	Individual capture history	Combinations of R1, W1, B1, O1, Y1, G1, R2, W2, B2, O2, Y2, G2	A string of multiple marks indicating the capture history of the individual

Although there were ten marking positions the weekly marking only utilised two of these positions. Given that six colours were used, there are thus only twelve combinations of colour and position. This led to some confusion between weeks: the field reports were accordingly organised into volumes with each volume being associated with a series of 12 weeks. The start of a new volume corresponds to a full rotation of colours and positions used to indicate the week of marking. This presents a limitation in the study since it assumes that all the flies captured at the start of a volume are dead by the end of the volume. This was not always the case, since some flies survived for about 150 days on the island. Moreover, to identify the week that the mark is pointing to, the mark needs to be amended with the volume it is referring to, to avoid confusion between the same marks in different volumes. Due to the reuse of

marks it was necessary to make the assumption that all the marks that appear on flies are the latest application of the mark.

On the field report the gender is considered in a more detailed column, namely the Type column. Monitoring teams considered four different types of individuals (young vs old, and male vs female) and assessed the age category of a captured individual as follows. The captured fly's exoskeleton was felt by gently pressing it. If the exoskeleton was hard, the fly was considered to be old; if it was soft, then it was classified as a young fly. The fly's gender was determined by examining the external genitalia. The type then followed as either young male/female or old male/female. The measuring technique of old and young is error prone: accordingly, only the gender is considered.

Each of the columns in the data base was tested to ensure that the populated values were valid according to the valid values listed in Table 3.1. Illicit values found by these checks were corrected by consulting the field report. When the field report was consulted all the histories on the report were also compared to the field report to remove any other discrepancies found. Spot checks were also performed in this manner.

3.1.3 Distribution of colour markings

Now that the data gathering process and the electronic data base have been described, attention is focused on the distribution of colour markings found on male and female *G. m. morsitans*. To understand the distribution of colour markings in the case of multiple markings, first consider the hypothetical example of a case where three samples are taken. Suppose that in the first sample captured flies are marked with a red dot and that in the second sample captured flies are marked with an orange dot. In the third sample, some flies will be unmarked, some will bear a red mark, some an orange mark, and some will bear both a red and an orange mark. The proportion of red marks is calculated by counting the number of flies bearing a red mark and dividing by the total number of marks on all the flies in the sample. The same is done for calculating the proportion of orange marks. The longer the time interval between the first two samples and the third sample the smaller these proportions will be. The decrease is due to flies exiting the catchable population by dying. The same logic holds when considering the distribution of colour markings in the multiple marking cases.

In the multiple marking case the samples are taken daily, but a new mark is only applied each week. The daily samples are, therefore, aggregated to form a weekly sample. The time between samples is also in weeks. The present study only reports the distribution of colour markings in a single volume - volume five. The distribution of colour markings are shown in Figure 3.2 and Figure 3.3.

The interpretation of Figures 3.2 and 3.3 requires some explanation. For each value on the abscissa $(1, 2, 3, \ldots, 11)$ there are 12 points (where fewer

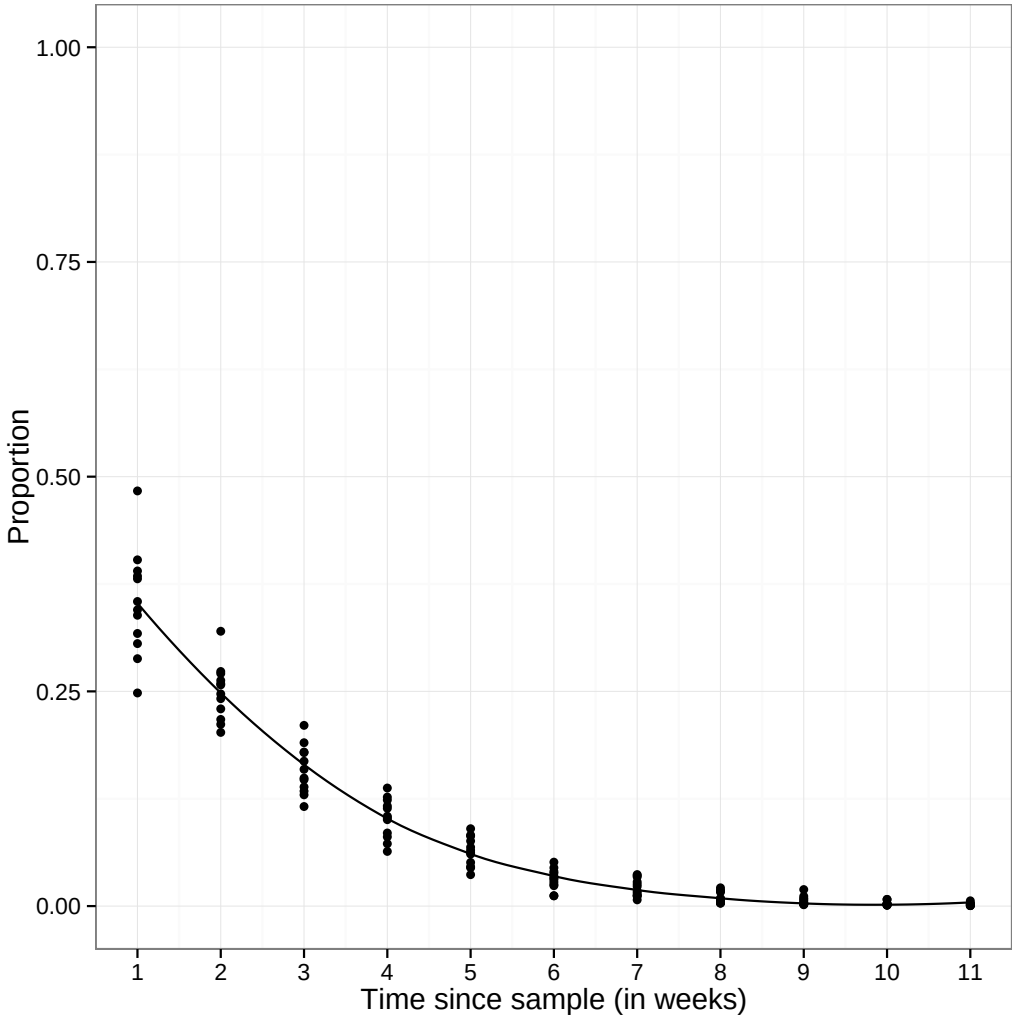


Figure 3.2: The proportion of total number of marks that were applied one, two, *etc.* weeks ago on male *G. m morsitans*, caught during the weeks 1 April 1981 - 23 June 1981 in the field experiment on Antelope Island. The fitted curve is a local linear regression (loess) curve (see A.1).

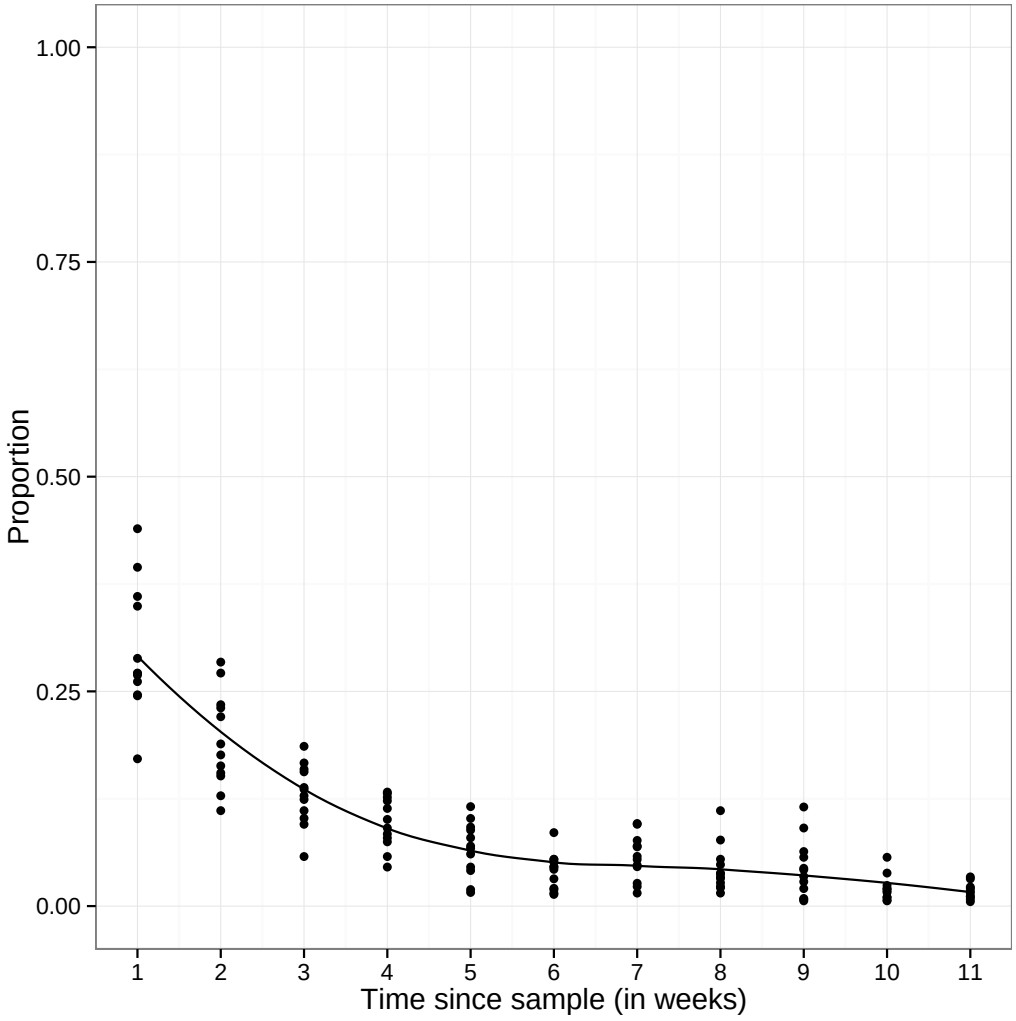


Figure 3.3: The proportion of total number of marks that were applied one, two, *etc.* weeks ago on female *G. m morsitans*, caught during the weeks 1 April 1981 - 23 June 1981 in the field experiment on Antelope Island. The fitted curve is a local linear regression (loess) curve (see A.1).

than 12 points are visible this simply indicates that some points lie on top of each other). Each point refers to the recaptures obtained in one of the weeks 1 April 1981 through 23 June 1981. For example, the dots with $x = 1$ measure the proportion of flies bearing a mark associated with the previous week - measured independently on each of 12 weeks in the volume. Similarly the dots with x=11 measures the proportion of flies bearing a mark associated with the mark applied 11 weeks ago - measured independently on each of the 12 weeks in the volume.

Consider the distribution of colour markings on the males. As the length of time (in weeks) increases, between the sample being considered and the application of the particular mark, the proportion of flies bearing the mark decreases. This is as expected, since marked flies will be dying as time passes. Moreover, the variation amongst the proportions in the weekly samples decreases when 'older' marks are considered. Again this is to be expected: if the mark was applied around 11 weeks ago there are almost no flies bearing this mark in the considered sample. The fitted local linear regression curve (loess) shows that the decline in the proportion is gradual, smooth and is close to zero at around 10 weeks between mark application and the sample being considered.

The colour distribution for females is more interesting than that of the males. Although the proportion decreases as the length of time (in weeks) increases, between the considered sample and the application of the particular mark, there is a brief increase in the proportion at around seven weeks between the sample and the application of the mark. Moreover the variance seems relatively constant for all the marks except those applied six, ten and eleven weeks ago. The fitted local linear regression curve (loess) shows that the decline starts gradually; stabilises and remains constant; then decreases again.

The observed differences between the distribution of colour markings on male and female tsetse flies is consistent with the observations of Hargrove (1990) that females have longer life spans than males and that older females have a higher capture probability than middle-aged females.

3.1.4 Data Errors

Table 3.2: Data Errors

Volume Number	Percentage of histories with errors (%)
3	1.08
4	1.31
5	5.17

Apart from ensuring that the values in the data base are correct according to the valid values column in Table 3.1, it is also necessary to assess the reliability of the data by using more advanced methods of error detection. The method used here is to notice that during mark-release-recapture experiments, marks accumulate in a particular order in the population. For example, consider the history "v5O1v5G1v5W2" that was observed in a sample taken during week thirty-three of the weekly marking experiment. The last mark in this history "v5W2" corresponds to week thirty-two (see Table A.1). We would expect to find the history "v5O1v5G1" in one of the samples taken during week thirty-two - since that is the (only) week when a fly could have acquired the mark "v5W2". If there is such a history we are confident that "v5O1v5G1v5W2" is a viable history to find in the data base. This test was performed for every history in the data base and the proportion of the histories that do not match to a previous history are shown in Table 3.2. Having found a match with a previous history it is necessary to ensure that other histories are not matching to the same past history. This is also checked in the data base. The occurrence of this error is relatively small in comparison to the previous error since only 20/16872(0.12%) of histories in volume five link back to the same past history.

3.2 The Model

The model was built using the ModGen platform, which is a superset of the C++ programming language, developed at Statistics Canada to facilitate the construction and implementation of individual-based models (Statistics Canada, 2014). The data generated by the individual-based model was subsequently analysed using R (R Core Team, 2014).

The model description follows the "Overview, Design Concepts and Details" standard protocol for documenting individual-based models proposed by Grimm *et al.* (2006). The "Overview" gives the purpose and structure of the model and consists of: Purpose, State variables and scales, and Process overview and scheduling. "Design Concepts" describe the general concepts underlying the design of the model. Lastly, the "Details" sections describe the sub-models and the baseline parameters for the simulation and consists of: Initialization, Input and Sub-models. In the description it is important to note that "Individual" is a modelling term, not a biological one. We refer to pupa individuals and adult fly individuals as the two types of individuals modelled.

3.2.1 Overview

Purpose

An individual-based model was built to simulate the distributions of colour markings emerging from samples taken from a simulated population that was:

1. Grown from 2000 founder pupa individuals of unknown age
2. Subjected to particular formulations of the mortality, reproduction and capture functions
3. Subjected to average temperatures on Antelope Island, Lake Kariba, Zimbabwe, during the actual experiment

State variables and scale

In the individual-based model, each individual and its state variables are followed over successive time intervals. The number of individuals in the simulation is variable. The individual simulated in the model is the tsetse fly specifically *G. m. morsitans*. Two categories of individuals are modelled: (i) pupa and (ii) adult fly. Both pupa individuals and adult fly individuals are characterised by the state variables: age and gender.

Female adult fly individuals and pupa individuals have additional state variables. The additional state variables for female adult fly individuals affect reproduction in the model and are: cumulative degree days since last larviposition or emergence and first larviposition or subsequent larviposition (section 3.2.3). The additional state variable for the pupa individuals affects

the time until a pupa individual is converted into an adult fly individual (*i.e.* equivalent in the real world to the emergence of the adult from the puparium) and is: cumulative degree days since larviposition (section 3.2.3). The simulation runs involve 400 days before marking started and 273 days while marking was in progress. This coincides with the first phase of the actual experiment performed on Antelope Island, Lake Kariba, Zimbabwe. The time step in the model is one day. The tsetse fly population on Antelope Island, Lake Kariba, Zimbabwe, was considered to be closed (Vale *et al.*, 1986), so immigration and emigration are not considered. The model is not spatially explicit.

Process overview and scheduling

The model is a continuous time model. Figure 3.4 is a schematic of how events are processed by the model. The tsetse life processes included in the model, as events, are ageing, death, larviposition, marking and emergence. Each event has two components: the time to event (rounded-rectangles) and event implementation (diamonds). Event times are calculated at the start of each individual's life using the appropriate sub-model (Section 3.2.3). The event times are then added to the event queue. All the events for all individual are ordered according to the time-to-event and the event with time closest to the current model time is implemented. Event implementation updates the relevant states of the individual and calculates the next timing of the event that was just implemented for that individual if the event can occur more than once in an individual's life. The death event removes all subsequent individual events, *i.e* events still in the event queue for that individual.

Apart from tsetse fly life processes there is a simulation day event. This event is implemented when the next calendar day is reached by the model time and it updates the temperature in the model.

3.2.2 Design Concepts

The distribution of colour markings in mark-release-recapture samples emerges in the model from the interaction between mortality and capture. Mortality and capture do not emerge in the model since they are represented entirely by empirical rules describing them as probabilities. Fitness (success of an individual based on some criteria) and decision making are not modelled explicitly, but included in the empirical rules for survival and capture probability.

Since decision making of individuals are not modelled explicitly, no explicit assumptions about what the individual and actual tsetse fly "knows" are made. However, it is assumed implicitly that individuals know their age and their gender to apply their age and gender specific capture probabilities. Individuals do not express aim, cognitive action or adaptation and do not directly interact together.

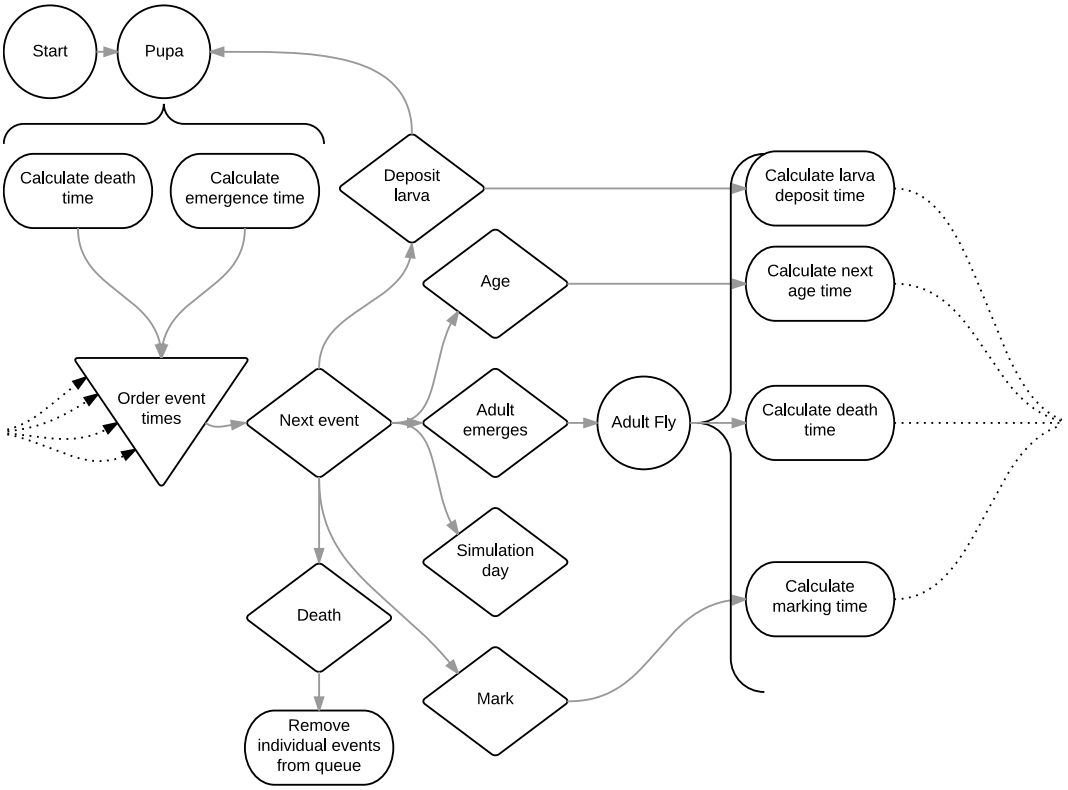


Figure 3.4: Schema of the schedule of the model of tsetse population dynamics.

The model is intrinsically stochastic: adult fly individual survival and capture are stochastic processes. This was done to include noise in mark-colour distributions and because the focus of the model is on the distribution of colour markings and not on actual tsetse fly behaviour.

At the end of the simulation, the model outputs the number of male and female adult fly individuals in the population for each simulation day and the distribution of colour markings among flies captured on all sampling occasions in a given week.

Since tsetse are poikilothermic animals, most of their life processes occur at temperature dependent rates. Temperatures used in all the modelling carried out here are taken to be exact, with no stochastic component, as given by the temperature records gathered by the weather station at Kariba Airport on the mainland about 5 km from Antelope Island. The mean daily temperature is taken as the average of the daily minimum and maximum temperatures (Figure A.1).

3.2.3 Details

Initialization

Since information on the exact age distribution of the pupae introduced onto the island is unavailable, the model simulates a random initial population by assigning a random gender and emergence time to each founder individual. The emergence of the introduced cohort is calculated by first considering the theoretical times to emergence that would have been observed if the founder pupa individuals had all been deposited the day before they were placed on the island. The synchronised timing of emergence of such a cohort, calculated from the pattern of changes in daily mean temperature on days following day zero, provides an estimate of the maximum time between the introduction of a pupa and its subsequent emergence. The "emergence" times of the founder individuals were then considered to be uniformly distributed between time zero and the estimated maximum. Simulations thus contained a random initial population representing a possible initial population in the actual experiment.

Input

The inputs are the daily average temperature and the parameters of the following sub-models derived from the literature.

Submodels

Reproduction

In the model, reproduction involves two equations that express the relationship between the timing of larviposition events and ambient temperatures. The equations represent the time lag (I_0 in days) between the time of emergence of a female adult fly individual and the time when she deposits her first larva, and the time lag (I in days) between two consecutive larviposition events (*i.e., the inter-larval period*) by the same female adult fly individual after the first larva. Hargrove (1994) suggest the equations:

$$I_0 = 1/(R1 + R2(T - 24)) \tag{3.1a}$$

$$I = 1/(R3 + R4(T - 24)) \tag{3.1b}$$

where T is the temperature and $R1$ to $R4$ are constant.

The model calculates (from the reciprocal of Equations 3.1a and 3.1b) the proportion of the inter-larval period that has been completed during day t (where t indicates the time at the start of the day t) and adds it to the running total for each female adult fly individual. On the day when the running total for a given female adult fly individual exceeds a value of 1.0, a new pupa individual is created during model time t and $t+1$. In practice, females tsetse flies deposit their larva during the afternoon, rather than at random times

during the day (Hargrove, unpublished). We accordingly assign a larviposition time of 1500h. A given female adult fly individual will thus create a pupa individual at time $t + 0.625$ if the exact time of larviposition as predicted by Equation 3.1 is at or before $t + 0.625$. If the exact time of the larviposition is after $t + 0.625$, the event is postponed to the next day, thus creating the pupa individual at time $(t + 1) + 0.625$. The exact time of the larviposition is calculated assuming the rate of development is constant during day t. Once a pupa individual is created the female adult fly individual's running total is set to zero and the model starts accumulating immediately for the next larva. The pupa individual draws a gender with probability 0.5. Equations 3.1a and 3.1b defining larviposition events do not contain a stochastic component.

Pupal development

In the model, pupal development involves two equations that express the relationship between puparial duration and ambient temperature. The equations represent the time lag (I_{fp} for female pupa individuals and I_{mp} for male pupa individuals, with units of days) between larviposition and adult emergence. Phelps and Burrows (1969*b*) suggest the equations:

$$I_{mp} = 1/(P1/(1 + exp(P2 - P3T))) \quad (3.2a)$$

$$I_{fp} = 1/(P4/(1 + exp(P5 - P6T))) \quad (3.2b)$$

where T is the temperature and $P1$ to $P6$ are constant.

The model calculates (from the reciprocal of the Equations 3.2a and 3.2b) the proportion of the puparial period that has been completed during day t (where t indicates the time at the start of the day t) and adds it to the running total for each pupa individual. On the day when the running total for a given pupa individual exceeds a value of 1.0, a new adult fly individual is created during model time t and $t + 1$. As with larviposition, the timing of adult emergence is not randomly distributed during the day. Instead adults emerge in the afternoon (Hargrove, unpublished). We assign an arbitrary emergence time of 1500h. A pupa individual thus creates a new adult fly individual that inherits its gender at time $t + 0.625$ if the exact time of emergence as predicted by Equation 3.2 is on or before $t + 0.625$. If the exact time of the emergence is after $t + 0.625$, the event is postponed to the next day, thus emerging at time $(t + 1) + 0.625$. The exact time of emergence is calculated assuming the rate of development is constant during any given day t. On the creation of the adult fly individual the pupa individual is removed from the simulation. Emerging female adult individuals start accumulating for a larva immediately after emergence. Equations 3.2a and 3.2b do not contain stochastic components.

Survival

Adult mortality: In the model, adult survival involves two equations. The equations represent the survival probabilities, that are dependent on the age of a fly and the temperature, $\varphi(A)_f$ for females and $\varphi(A)_m$ for males. Hargrove *et al.* (2011) suggest these equations for survival probabilities in adult flies. However, the equations in Hargrove *et al.* (2011) do not take account of the variation in survival probability with changing temperature. The proposed model to consider both age and temperature is an adaptation of the Hargrove *et al.* (2011) equations, allowing the parameters to vary linearly with temperature. Evidence for adult tsetse fly mortality being dependent on temperature is presented in Hargrove (2001*a*). The proposed models for male and female survival probabilities are:

$$\varphi(A)_m = exp[S1(T)(exp(-S2(T)A) - exp(S3(T)A))] \tag{3.3a}$$
$$\varphi(A)_f = exp[S4(T)(exp(-S5(T)A) - exp(S6(T)A))] \tag{3.3b}$$

where A indicates the age and T the temperature. We also have

$$\begin{aligned} S1(T) &= S1 \times T \\ S2(T) &= S2 \times T \\ S3(T) &= S3 \times T \\ S4(T) &= S4 \times T \\ S5(T) &= S5 \times T \\ S6(T) &= S6 \times T \end{aligned}$$

where $S1$ to $S6$ are constants.

In the model, each adult fly individual is assigned a random number between 0 and 1. At the start of each day the death probability is calculated, multiplied with the survival probability of surviving to the start of that day and compared with the random number to determine if the death event occurs. For example, on the start of day t the death probability is given by:

$$P(t < T < t+1 | T > t) = 1 - \varphi(A_t)/\varphi(A_{t+1}) \tag{3.4}$$

where A_t and A_{t+1} are the age of the adult fly individual at time t and $t+1$ respectively. The temperature used is the average temperature for day t. The survival probability to time $t+1$ is then calculated by:

$$s_{t+1} = P(T > t+1) = s_t \times (1 - P(t < T < t+1 | T > t)) \tag{3.5}$$

where s_t is the accumulated survival probability of the previous t days since the adult fly individual was created. When the survival probability to time $t+1$ is larger than the random number the adult fly individual dies during

day t. During a particular day the deaths are distributed uniformly during the day; hence the death event competes with other events scheduled for that day. Density-dependence is not included in the survival probability. This process is stochastic to allow random variation in the mark-release-recapture data gathered from the model.

Pupal mortality: In the model, pupal mortality is set as a constant loss per day for each day the pupa individual is present in the model. At the start of each day the model assigns a random number between 0 and 1 to each pupa individual. If the number is smaller than the loss per day the pupa dies that day. Deaths are distributed uniformly during the day of death. The death event competes with the emergence event on the day of emergence.

Capture

Hargrove (1990) presents evidence that capture probability is affected by both the age and the gender of *G. m. morsitans*. In the model, at the start of each day each adult fly individual is assigned a random number between 0 and 1. If the random number is less than the capture probability, specified by the age and the gender of the individual, the mark event is scheduled for that day. During a particular day t the capture events occur uniformly between $t + 0.25$ and $t + 0.354$ or between $t + 0.646$ and $t + 0.75$ to coincide with the two fly rounds in the actual experiment. Each individual can only be captured ten times.

3.3 Model Tests

A thorough and aggressive testing regime is important for the development of a successful, efficient and credible model. It improves efficiency, because no time is wasted analysing system behaviours when they are the result of problems at lower levels. It improves credibility, because it shows that individual-level traits were analysed independently before system behaviours were examined. Examining individual-level behaviour often involves using unrealistic scenarios (Grimm and Railsback, 2005).

The following hierachy of code testing methods is the minimum necessary to provide reasonable assurance that an IBM's software is ready to be used (Grimm and Railsback, 2005):

Code reviews: A reviewer compares the code to the written formulation of the model to look for mistakes.

Visual tests: Visual tests are conducted simply by running the model and observing its behaviour.

Spot checks: Spot checks verify a few selected model calculations by comparing results to those calculated by hand.

The code tests were performed on the individual-based model. The following sections describe the results of these tests in turn.

3.3.1 Code reviews

Code reviews are usually the first line of defence against major programming errors and poor software design. In the individual-based model developed here, however, code reviews took place on the final model. A more experienced programmer provided a final validation of the software design and ensured that good programming practices were followed.

3.3.2 Visual tests

Visual tests are easy and important and were conducted every time the software of the model was modified. BioBrowser, the ModGen Biography Browser, is a stand-alone software product which supplements the ModGen language used to build the individual-based model. BioBrowser graphs the lifetimes of individuals and their characteristics as they change over the course of the individual's lifetime. BioBrowser can present one or many characteristics for an individual.

An example of the graphs produced by BioBrowser is given in Figure 3.5. In this particular case the characteristics associated with the marking process are presented. The first sub-plot (top grey bar) shows the lifetime, from creation

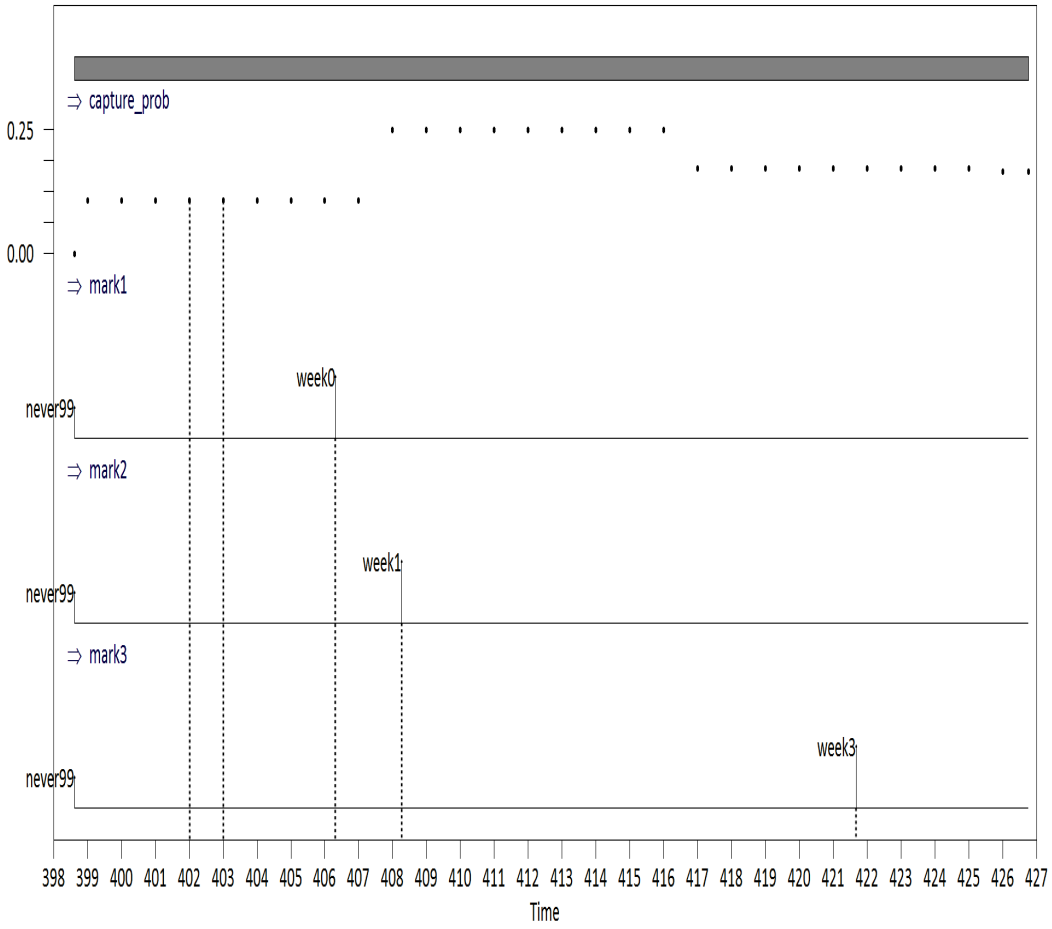


Figure 3.5: An example of a graph produced by BioBrowser showing the particular states associated with the marking process.

(emergence) of an adult fly individual at model time 398.62 until its removal (death) at model time 426.76, for a particular male adult fly individual. The second sub-plot (capture_prob) shows the changing capture probability at the start of each day. Even though the capture probability is defined for times before model time 400 days, the weekly marking experiment only started on day 400 and thus capture is only possible after model time 400. The next three sub-plots (mark1, mark2 and mark3) show the time the particular adult fly individual was captured during the simulated mark-release-recapture experiment. The first mark was at model time 406.31, the second at model time 408.27 and the third at model time 421.69. Similar representations are possible for other characteristics associated with the different processes in the model.

3.3.3 Spot checks

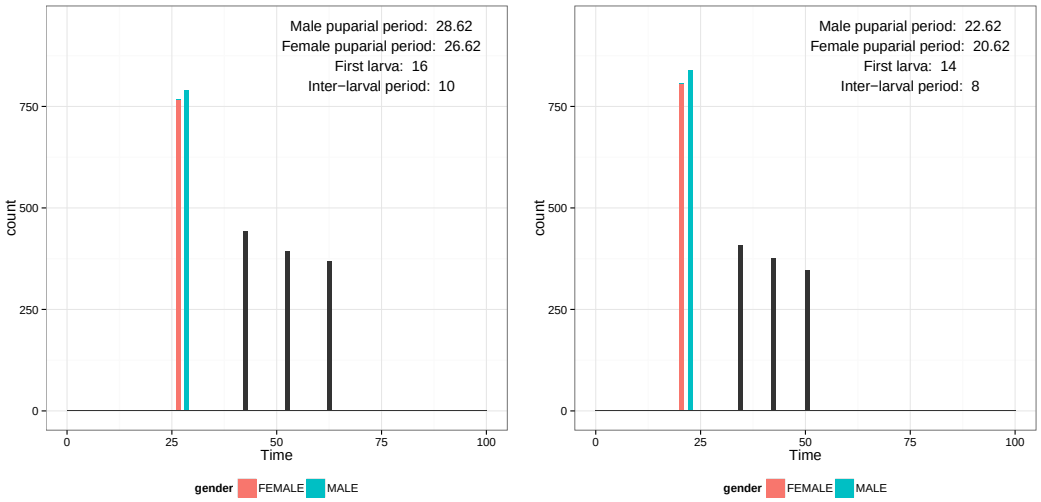
The sub-models for reproduction and mortality are tested in the developed individual-based model against hand calculated values. The hand calculated values are shown in Table 3.3. The graphical displays in Figure 3.6 are the output from the individual-based model. The scenarios were conducted using the base-line parameters and under conditions of constant temperatures of 25 and 30°C.

Reproduction

Figures 3.6a and 3.6b show the puparial period and larvipositions of a cohort of 2000 pupa individuals initialised at the same time (shown as time zero in the graphs). The puparial period in the individual-based model is 28.62 days for males and 26.62 days for females, when the temperature is constant at 25°C. These are essentially the same as the values calculated by hand (Table 3.3) allowing for model "rounding" as described in Section 3.2.3, *i.e.* if the exact time of emergence as calculated in Table 3.3 is on or before $t+0.625$ the event takes place at $t+0.625$. If it is after $t+0.625$, the event is postponed to the next day, thus to time $(t+1)+0.625$. The same is true for larviposition, but we need to consider that the female emerged at time $t+0.625$. For example, a female individual emerging after 26.62 days - temperature constant at 25°C - will deposit her first larva at $26.62+15.87=42.49$ days which will be "rounded" to 42.62 days giving the result of the deposit taking place 16 days after emergence. Allowing for the same calculations for the inter-larval period and a constant temperature of 30°C, we note that the sub-models for reproduction are behaving as expected in the individual-based simulation. The decrease in the heights of the bars is due to female mortality.

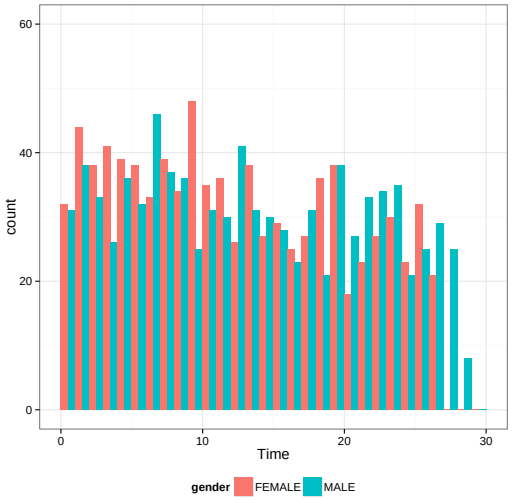
Figures 3.6c and 3.6d show the puparial periods of the founder individuals if the temperature was constant at 25 and 30°C. As described previously, the emergence times of the founder individuals are randomly distributed between model time zero and a maximum number of days. The maximum number of days is the puparial period had the founder pupa individual been initialised at model time zero. Emergence still occurs at $t+0.625$ for the founder individuals, where t is the randomly distributed integer. The results shown are as expected.

Figures 3.6e and 3.6f show the number of pupa individuals each female adult fly individual in the cohort deposited before she was removed from the simulation by the death event. We expect there to be little variation between the cohort simulated under a constant temperature of 25 and 30°C because, although the females are producing more pupa individuals per unit time at 30°C, they have shorter life spans at the higher temperature.

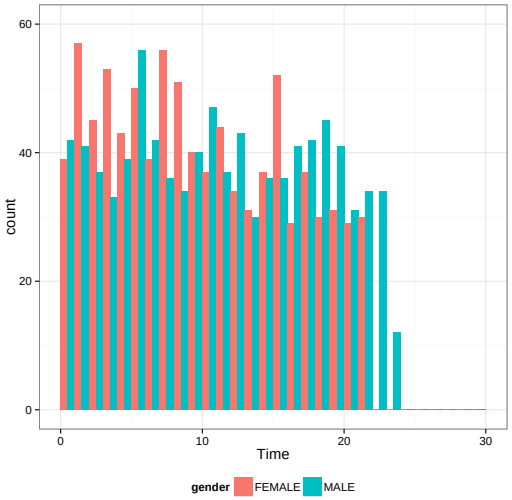


(a) 25°C

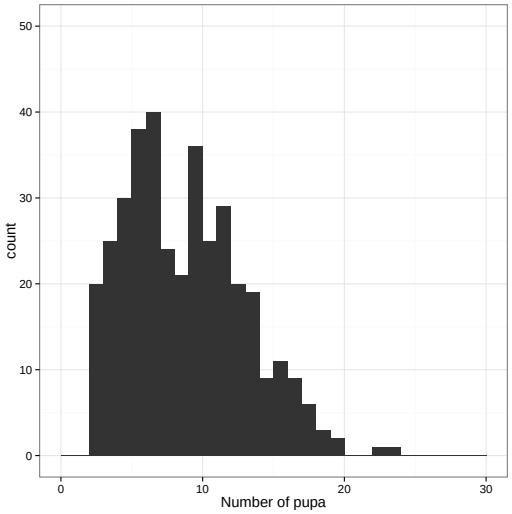
(b) 30°C



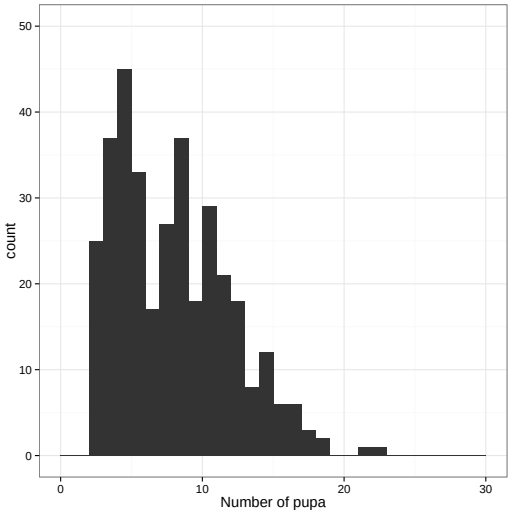
(c) 25°C



(d) 30°C



(e) 25°C



(f) 30°C

Figure 3.6: Spot checks performed on the reproduction process in the individual-based model for a cohort of 2000 individuals under constant temperatures of 25 and 30°C.

Table 3.3: The duration (in days) of pupal period, first and subsequent larvipositions calculated by hand for 25 and 30°C.

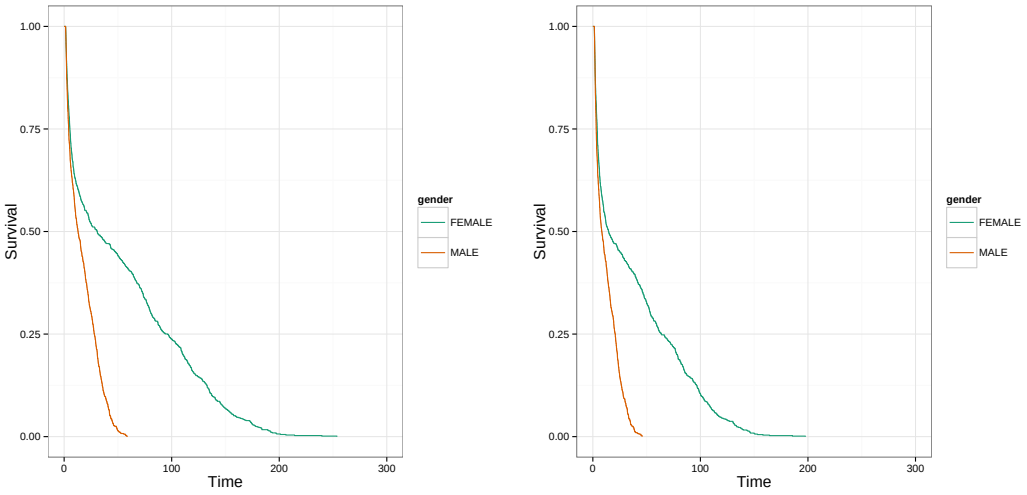
Temperature	Sub-model tested	Calculation
25°C	Male pupal period	$\frac{1+exp(5.3-0.24\times25)}{0.053} = 28.23746$
	Female pupal period	$\frac{1+exp(5.5-0.25\times25)}{0.057} = 25.83099$
	First larviposition	$\frac{1}{0.061+0.002(25-24)} = 15.87302$
	Subsequent larvipositions	$\frac{1}{0.1046+0.0052(25-24)} = 9.107468$
30°C	Male pupal period	$\frac{1+exp(5.3-0.24\times30)}{0.053} = 21.68997$
	Female pupal period	$\frac{1+exp(5.5-0.25\times30)}{0.057} = 19.91816$
	First larviposition	$\frac{1}{0.061+0.002(30-24)} = 13.69863$
	Subsequent larvipositions	$\frac{1}{0.1046+0.0052(30-24)} = 7.36377$

Mortality

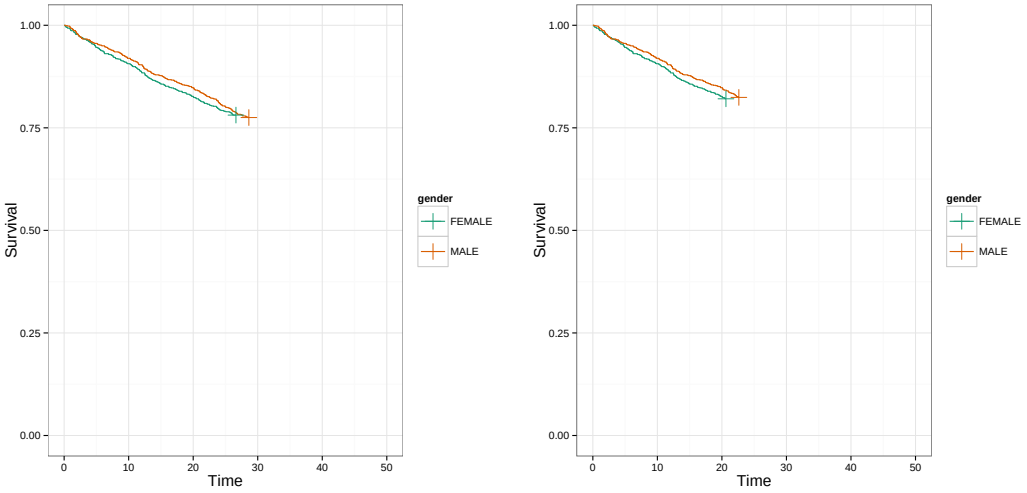
Figures 3.7a and 3.7b show the survival function for a cohort of adult fly individuals initialised at the same time (shown as time zero in the graphs). These graphs were constructed using time to event data obtained from the cohort as input data for R to estimate the survival functions. These functions are exactly as expected for the functions in Equation 3.3.

Figures 3.7c and 3.7d show the survival function for a cohort of pupa individuals initialised at the same time (shown as time zero in the graphs). These graphs were constructed using time to event data obtained from the cohort as input data for R to estimate the survival functions. Some of the event data include censored values indicated by the crosses in the graph. Censoring occurs when pupa individuals are converted to adult fly individuals. These functions are as expected given a constant rate of death.

Having established that the model is credible on the individual level we can examine some of the system level behaviours.



(a) Adult fly individual survival at 25°C (b) Adult fly individual survival at 30°C



(c) Survival in the pupa stage at 25°C (d) Survival in the pupa stage at 30°C

Figure 3.7: Spot checks performed on the mortality process in the individual-based model for a cohort of pupa individuals and adult fly individuals under constant temperatures of 25 and 30°C.

Chapter 4

Simulations

The individual-based model developed here includes some parameters that are not well known, namely the survival function parameters and the capture probabilities. Manual calibration may work since being structurally rich and having many parameters does not automatically mean that the model responds to the parameters in a complex way or even that the parameters must be calibrated (Grimm and Railsback, 2005). We explore below whether this is the case in the individual-based model developed here.

In the individual-based model developed here the survival function parameters are adjusted first by considering the simulated population over model time. A rough comparison is made between the simulated population and the number of tsetse flies captured during weekly samples taken on Antelope Island, Lake Kariba, Zimbabwe (Figure 4.1a), where it should be noted that model time 400 corresponds to the day when sampling began for the weekly marking mark-recaptyre exercise. When the model is executed using the base-line parameters, derived from the literature, the simulated population is decreasing and there are insufficient individuals to support the captures numbers observed in the field (Figure 4.1b). Accordingly, all mortality parameters are reduced by the same proportion. The simulated female population reaches viable levels when the model is executed using mortality parameters that are reduced to 77% of the original value (Figure 4.1c). With these reductions, however, the male populations were still not viable. Accordingly male mortality was decreased further. The simulated male population reaches viable levels when the model is executed using mortality parameters that are reduced to 59% of the original value (Figure 4.1d). We notice from Figure A.1 that temperatures will be increasing from model time 400 (October) onwards as this is the peak of the hot season. We would therefore expect the tsetse population to be under some strain at and after model time 400. Once the rains break, and temperatures decrease, in mid-December, the populations gradually increases again. This effect is indeed observed in the declining number of flies actually caught between October and December Figure A.1. The growth rate in the simulated population numbers of male flies drops to zero during this time, though

numbers do not actually decline, and the simulated population numbers of females continue to increases. This mismatch between expected and predicted numbers of tsetse populations points to the need for further refinement of the calibration of the model.

The distributions of colour markings using the adjusted mortality parameters are our next focus. The simulated distributions of colour markings are shown in Figure 4.2b and 4.2d. The simulated distribution of colour markings on males shows the same pattern and variation observed in the actual experiment. The simulated distribution of colour markings on females shows a marked difference from the pattern and variation observed in the actual experiment. Manually calibrating the simulated colour distribution proved unproductive. A more sophisticated approach needs to be adopted.

Identifying parameter values from observed data is still an open issue for individual-based models (Grimm and Railsback, 2005). Beaudouin *et al.* (2008) propose an approach that combines local and global sensitivity analysis with calibration. Conventional gradient-based optimization methods are difficult to conduct with individual-based models because the gradient must be laboriously estimated by repeated simulation (Beaudouin *et al.*, 2008).

The technical details that follow are based on Ginot *et al.* (2006) and Beaudouin *et al.* (2008). The notation used is that of Beaudouin *et al.* (2008).

Local sensitivity analysis

Let p_i denote a parameter i and M_j an output j. The local sensitivity coefficients are defined as

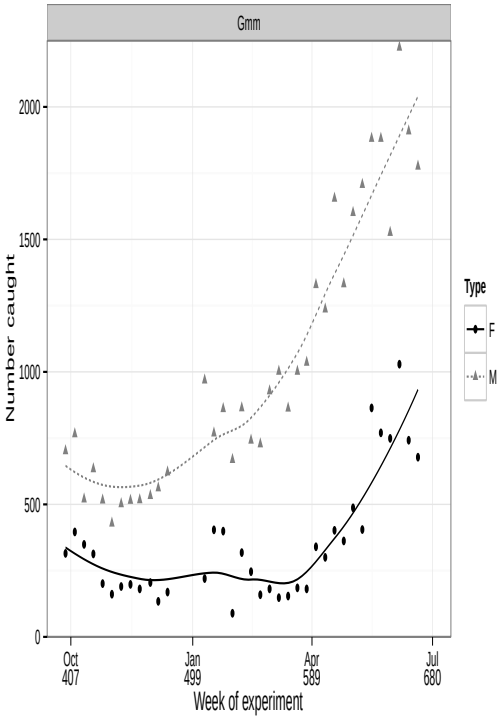
$$ls_{i,j} = \frac{\delta M_j}{\delta p_i} \tag{4.1}$$

Local sensitivity coefficients can be computed by varying all the parameters simultaneously, at random, around their prior values (e.g. 500 drawings with parameters varying between $\pm 5\%$) and performing linear regressions. The regression coefficients are then estimates of the local sensitivity coefficients for the parameters around the prior values. These coefficients are scaled to obtain the relative local sensitivity coefficients

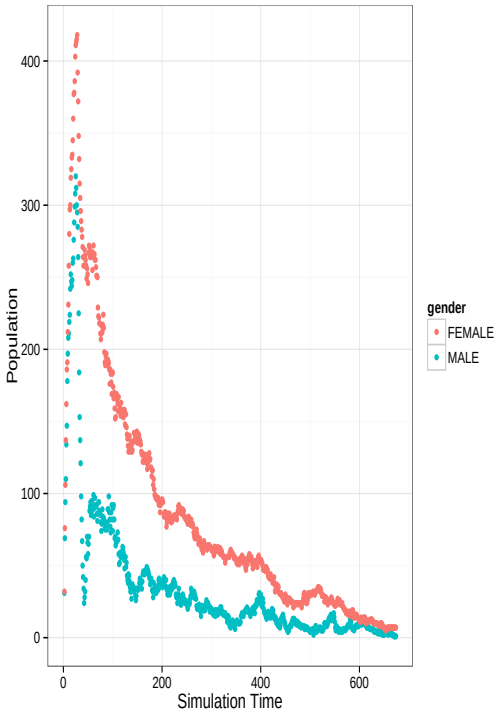
$$rls_{i,j} = \frac{\delta M_j}{\delta p_i} \frac{p_i}{M_j} \tag{4.2}$$

where p_i is the prior value for parameter i.

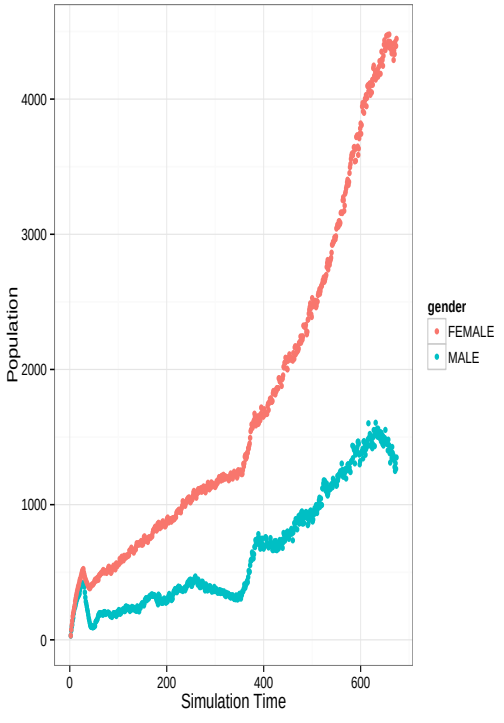
The values of these local sensitivity coefficients are subject to chance; their significance (or uncertainty) needed to be assessed. Beaudouin *et al.* (2008) does this by first including dummy parameters, with no influence on the model outputs, in the computation of the local sensitivity coefficients and then, calculates the total relative local sensitivity coefficients ($rls_{i,tot}$) for dummy and real parameters:



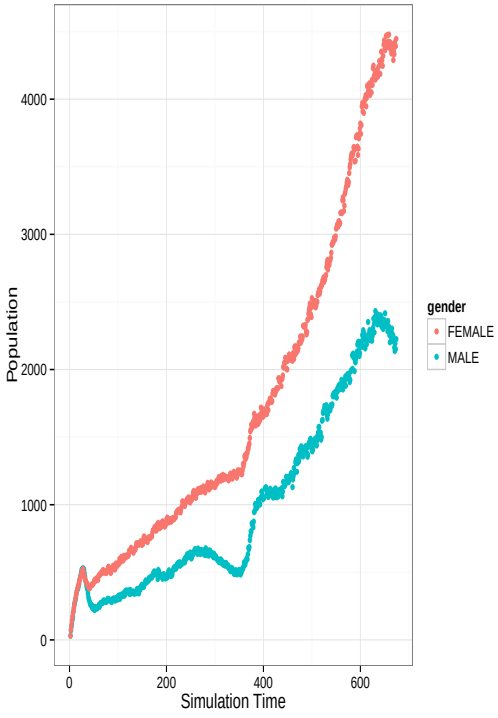
(a) Number of flies captured per week



(b) Base-line

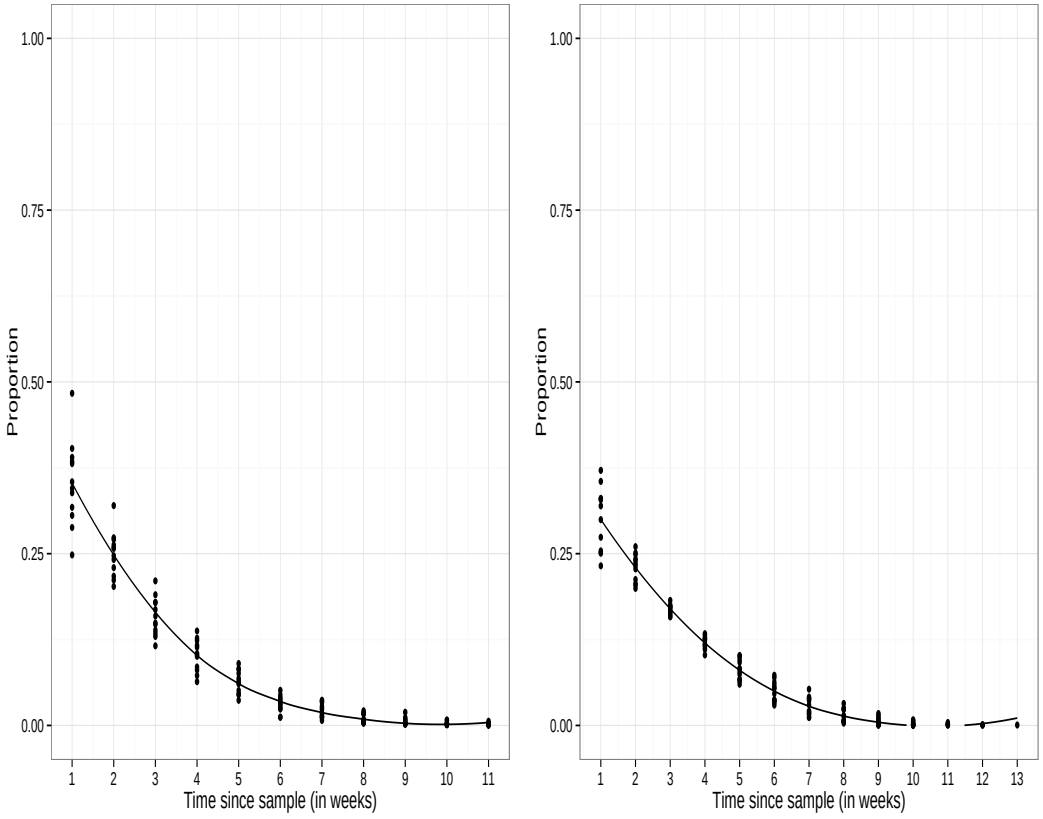


(c) Male and Female mortality reduced to 77%



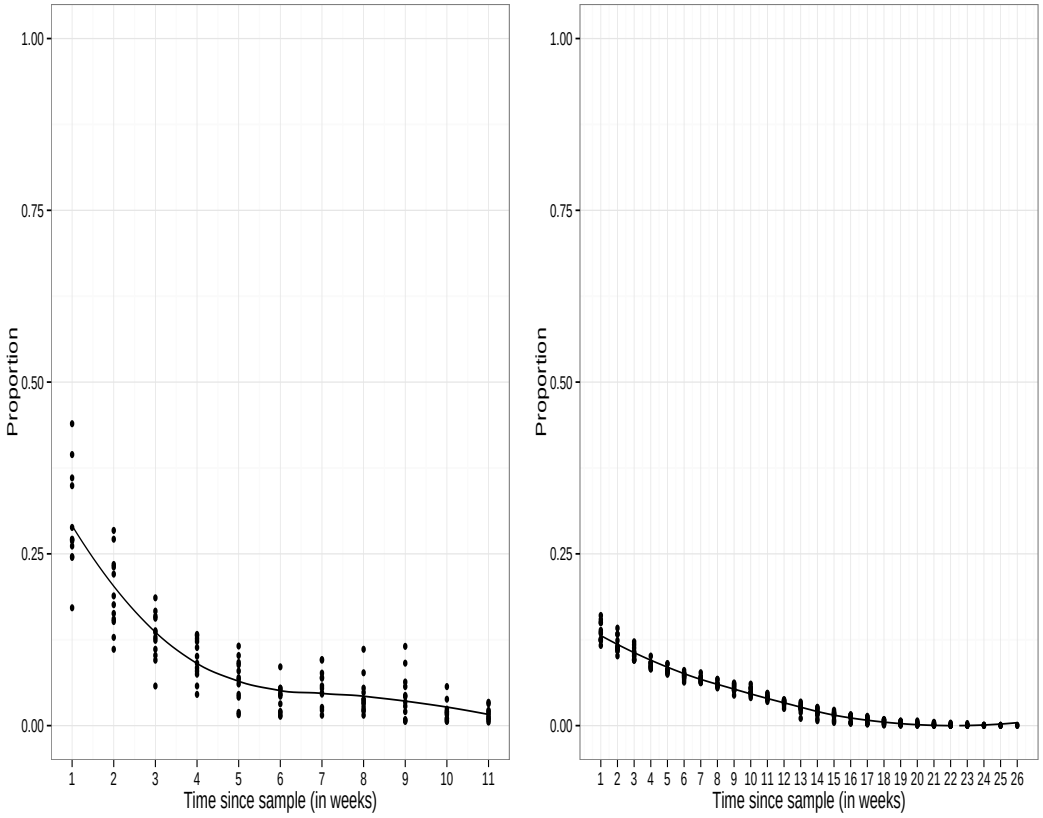
(d) Male mortality reduced to 59%

Figure 4.1: The effect of adjusting the mortality parameters on the simulated male and female populations. a) The catch numbers observed on Antelope Island, Lake Kariba, Zimbabwe. b) Simulated male and female populations using the base-line parameters. c) Simulated male and female populations using base-line mortality parameters reduced by 23%. d) Simulated male and female populations using base-line mortality parameters reduced by 23% for female and 41% for males.



(a) Observed distribution for males

(b) Simulated distribution for males



(c) Observed distribution for females

(d) Simulated distribution for females

Figure 4.2: Comparison between female/male observed and simulated distribution of colour markings.

$$rls_{i,tot} = \sqrt{\sum_j (rls_{i,j})^2} \tag{4.3}$$

The dummy parameters give the distribution of the total relative local sensitivity coefficients under the null hypothesis stating that the model output is unaffected by the parameter. Parameters with total relative local sensitivity coefficients lower than the 99% upper percentile of the distribution are not distinguishable from random effects. If too many parameters fall below the threshold the number of drawings should be increased.

Total relative local sensitivity coefficients can also be computed for a output with regards to all the parameters ($rls_{tot,j}$):

$$rls_{tot,j} = \sqrt{\sum_i (rls_{i,j})^2} \tag{4.4}$$

The relative local sensitivity coefficients can be placed into a matrix where the rows indicate the parameters and the columns the outputs. The correlation between the rows and the correlation between the columns are calculated. Hierarchical clustering is also performed on the rows and columns. Hierarchical clustering groups the parameters having similar actions on outputs and outputs having similar responses to parameters.

Knowing the uncertainties of the parameter we can calculate the relative local uncertainty coefficients which estimate which parameters mostly impact the outputs:

$$als_{i,j} = \frac{\delta M_j}{\delta p_i} \frac{\Delta p_i}{M_j} \tag{4.5}$$

where Δp_i is the uncertainty range (upper limit - lower limit).

Again the total relative local uncertainty coefficients are calculated:

$$rls_{i,tot} = \sqrt{\sum_j (als_{i,j})^2} \tag{4.6}$$

Global sensitivity analysis

Ginot *et al.* (2006) proposes using ANOVA for global sensitivity analysis to identify the interactions in sub-sets of parameters as well as to assess the global contribution of the parameters in their uncertainty range. It involves performing one at a time analysis *i.e.* plot the model output as a function of each parameter, for a wide range of values for that parameter keeping the rest constant. The shape and the evolution of the variance for each parameter is assessed. Since ANOVA is more reliable if the variance is constant and simpler to interpret if the effects are linear, transforming model output and the scale to enhance linearity is recommended. Keeping in mind that transformations alter the interpretation.

The one at a time analysis also helps assess the number of levels to use per parameter. Ginot *et al.* (2006) suggest two to be appropriate in most cases. Complete factoral design is then implemented and ANOVA is performed.

Calibration

Now having the results of the sensitivity analyse the steps for performing the calibration are outlined below.

Step 1 *Select one badly fitted output*

Identify all badly fitted outputs and choose the output having the highest total relative local sensitivity coefficient Equation 4.4.

Step 2 *Identify correlated outputs*

Examine the correlations between the columns of the relative local sensitivity coefficients matrix. Outputs that are correlated with the selected output could be affected by the calibration.

Step 3 *Select the parameters*

Use the results of the hierarchical clustering to identify the clusters of parameters which strongly influence the value of the selected output, while paying attention to their effects on the correlated outputs identified in Step 2. Select the parameters with both high total relative local sensitivity coefficients and total relative local uncertainty coefficients.

Step 4 *Refine the parameter selected*

Examine the correlations between the rows of the relative local sensitivity coefficients matrix. At the start of the calibration process select one of these for calibration to avoid an infinite number of parameterizations giving the same results. At the end of calibration process include all the correlated parameters so that the "biologically acceptable" parameter values can be chosen from the infinite number of equivalent parameterisations.

Step 5 *Included important interactions*

Order the parameters according to their local uncertainty coefficient and their global sensitivity coefficient. If the local uncertainty coefficients and global sensitivity coefficients hierarchies are similar, local analysis gives a good idea of the parameter effects on the model outputs. If local and global hierarchies are dissimilar, the impact of some parameters on outputs are dependent on the values of other parameters. If this is the case identify and include the interacting parameters in the calibration.

Step 6 *Build calibration design*

Choose the number of levels of each parameter selected in the first five steps to a build complete factorial design with acceptable computation cost.

Step 7 *Perform the calibration*

Select parameterisations that meet predefined criteria, for example appropriate simulated population levels. The best parameterisation can then be chosen by considering the weighted sum of least squares between simulated and observed values:

$$wls = \sum_{i=1}^{N} \frac{(sim_i - obs_i)^2}{\sigma^2 obs_i} \tag{4.7}$$

where N is the total number of outputs, sim_i is the simulated expectation (average from a few replications), obs_i is the observed replications and $\sigma^2 obs_i$ is the variance of the observed mean estimated by bootstrapping the replications.

The calibration process can be halted when all simulated values fall into the 95% bilateral confidence interval of the observed values. In the case of the distribution of colour marking the bilateral confidence intervals can be estimated by using bootstrap techniques.

Chapter 5

Conclusion and Outlook

The aim of this work was to develop an individual-based model to simulate, in all its detail, the results of a capture-recapture experiment carried out on Antelope Island, Lake Kariba, Zimbabwe. The model developed here was used to simulate the changes in fly numbers with time after the onset of experiment as well as the temporal acquisition of marking patterns on individual flies. Optimisation of the fit between observed and simulated data was beyond the scope of the current study, but the principles involved in such an optimisation procedure are outlined in the thesis.

The individual-based model developed here has potential applications well beyond the modelling of the Antelope Island mark-release-recapture data for which it was initially and specifically designed. Since the model essentially follows the entire life history of individual flies it will also, by implication, generate the age structure of the population, and keep track of the total numbers of pupae and adult flies, of each gender, present in the population at any moment in time. The model can, therefore, be used to model temporal and indeed spatiotemporal changes in total population, changes in sex, and species, ratio and changes in age structure.

Since it is possible to define the probability of capture of individual flies of different gender and age, the model can be used to study the difference between the age structure of female flies captured - using different sampling techniques - and the age structure of the actual population. Differences between the age structures of sample and actual population age structures have been identified as an important source of error in the estimation of adult female tsetse mortality from sample ovarian age structures derived from ovarian dissection studies (Hargrove, 1992*b*). Moreover, it is now apparent that, at least at sites like Rekomitjie Research Station, tsetse populations almost never have a stable age distribution (Hargrove and Ackley, 2014) and this greatly complicates the estimation of mortality from age distribution data.

Conversely, the benefit of the individual-based modelling approach is that there is no assumption about the stability of the age structure, nor indeed about the independence of probabilities of survival and/or of capture with

age. Instead these relationships can be allowed to vary and the effect of different assumed parameter values and formats on the resulting predictions of population structure can be explored.

A further exciting prospect for the use of the individual-based approach lies in the modelling of tsetse-borne trypanosomiasis. Since the model follows the life-steps of individual flies it can, *inter alia*, follow the feeding pattern of these flies and, thereby, follow the time course of the passage of trypanosomes from infected vertebrate host to uninfected fly and on to an uninfected vertebrate host. As such the individual-based approach could very easily be used to model, for example, the time course of the acquisition of infections by tsetse, which has been studied in detail in various settings (e.g., Woolhouse *et al.* (1993, 1996); Woolhouse and Hargrove (1998)).

It is then a simple step to use the model, further, to model the effects of various control interventions - aimed at controlling the disease, either directly by attacks on the trypanosome, or on the tsetse vectors of the disease.

A similar approach to the modelling, and control, of tsetse and trypanosome populations has been undertaken by Vale and Torr (2005). These authors have developed a suite of simulation programmes [see www.Tsetse.org] that can, similarly, be used to predict the growth of populations and of the age, gender and species structure. The programmes have, however, been written in Excel, rather than a high-level language such as is used here. Moreover, they deal deterministically with cohorts of flies rather than following the fates of individual flies in a fully stochastic way as is done here. The current work marks the first time that an individual-based modelling approach has been applied to the study of tsetse population dynamics.

Appendices

Appendix A

A.1 Local Linear Regression

Local linear regression (loess) is a non-parametric regression method that combines multiple regression models in a k-nearest-neighbour-based meta-model. Consider a target point in the data set. A line is fitted to a subset of the data where the subset contains points that are "close" to the target point. The line is fitted using weighted least squares where more weight is assigned to points near the target point than to points further away. Although we fit an entire linear model to the subset of the data, the linear model is only used to evaluate the fit at the target point. This process is performed for each point in the range of the data.

The subset of the data used for each point is flexible and was set to 75% in all loess graphs in the document, *i.e.* only the closest 75% of all the data is included in the subset.

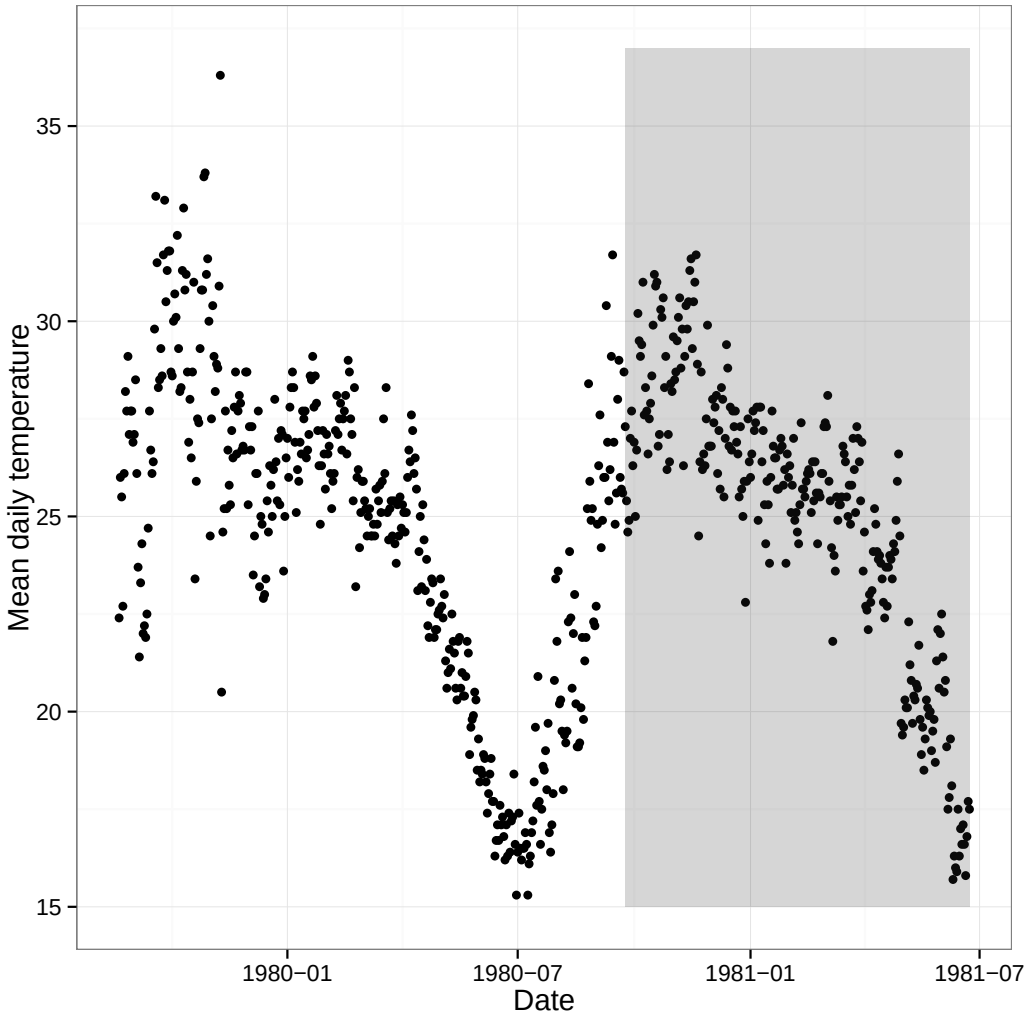


Figure A.1: The mean daily temperature on Antelope Island, Lake Kariba, Zimbabwe. Records were gathered by the weather station at Kariba Airport on the mainland about 5 km from Antelope Island. The mean daily temperature is taken as the average of the daily minimum and maximum temperature. The grey area indicates the period over which marking was taking place.

Table A.1: Colour and position of marks relating to the weeks of the experiment as well as exact dates of the week in question

Week	Volume Number	Mark	Start Date	End Date
1	3	v3R1	1980-09-24	1980-09-30
2	3	v3W1	1980-10-01	1980-10-07
3	3	v3B1	1980-10-08	1980-10-14
4	3	v3O1	1980-10-15	1980-10-21
5	3	v3Y1	1980-10-22	1980-10-28
6	3	v3G1	1980-10-29	1980-11-04
7	3	v3R2	1980-11-05	1980-11-11
8	3	v3W2	1980-11-12	1980-11-18
9	3	v3B2	1980-11-19	1980-11-25
10	3	v3O2	1980-11-27	1980-12-02
11	3	v3Y2	1980-12-03	1980-12-09
12	3	v3G2	1980-12-10	1980-12-16
13	4	v4R1	1981-01-07	1981-01-13
14	4	v4W1	1981-01-14	1981-01-20
15	4	v4B1	1981-01-21	1981-01-27
16	4	v4O1	1981-01-28	1981-02-03
17	4	v4Y1	1981-02-04	1981-02-10
18	4	v4G1	1981-02-11	1981-02-17
19	4	v4R2	1981-02-18	1981-02-24
20	4	v4W2	1981-02-25	1981-03-03
21	4	v4B2	1981-03-04	1981-03-10
22	4	v4O2	1981-03-11	1981-03-17
23	4	v4Y2	1981-03-18	1981-03-24
24	4	v4G2	1981-03-25	1981-03-31
25	5	v5R1	1981-04-01	1981-04-07
26	5	v5W1	1981-04-08	1981-04-14
27	5	v5B1	1981-04-15	1981-04-21
28	5	v5O1	1981-04-22	1981-04-28
29	5	v5Y1	1981-04-29	1981-05-05
30	5	v5G1	1981-05-06	1981-05-12
31	5	v5R2	1981-05-13	1981-05-19
32	5	v5W2	1981-05-20	1981-05-26
33	5	v5B2	1981-05-27	1981-06-02
34	5	v5O2	1981-06-03	1981-06-09
35	5	v5Y2	1981-06-10	1981-06-16
36	5	v5G2	1981-06-17	1981-06-23

List of References

Beaudouin, R., Monod, G. and Ginot, V. (2008). Selecting parameters for calibration via sensitivity analysis: An individual-based model of mosquitofish population dynamics. *Ecological Modelling*, vol. 218, no. 1, pp. 29–48.

Challier, A. (1965). Amélioration de la méthode de détermination de l'âge physiologique des glossines: études faites sur *Glossina palpalis* gambiensis Vanderplank, 1949. *Bulletin de la Société de Pathologie Exotique*, vol. 58, no. 2, pp. 250–259.

Ginot, V., Gaba, S., Beaudouin, R., Aries, F. and Monod, H. (2006). Combined use of local and ANOVA-based global sensitivity analyses for the investigation of a stochastic dynamic model: application to the case study of an individual-based model of a fish population. *Ecological Modelling*, vol. 193, no. 3, pp. 479–491.

Grimm, V. (1999). Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling*, vol. 115, no. 2, pp. 129–148.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G. *et al.* (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, vol. 198, no. 1, pp. 115–126.

Grimm, V. and Railsback, S.F. (2005). *Individual-based Modeling and Ecology*. Princeton Paperbacks. Princeton University Press. ISBN 9780691096667.

Hargrove, J.W. (1990). Age–dependent changes in the probabilities of survival and capture of the tsetse, *Glossina morsitans morsitans* Westwood. *International Journal of Tropical Insect Science*, vol. 11, no. 3, pp. 323–330.

Hargrove, J.W. (1991). Ovarian ages of tsetse flies (Diptera: *Glossinidae*) caught from mobile and stationary baits in the presence and absence of humans. *Bulletin of Entomological Research*, vol. 81, no. 1, pp. 43–50.

Hargrove, J.W. (1992*a*). Management of insect pest: nuclear and related molecular and genetic techniques; proceedings. In: *International Symposium on Management of Insect Pest: Nuclear and Related Molecular and Genetic Techniques. Vienna (Austria). 19-23 Oct 1992.*

Hargrove, J.W. (1992*b*). Problems associated with estimating mortality from sample age distributions. *Management of insect pests: nuclear and related molecular and genetic techniques*, p. 549.

Hargrove, J.W. (1994). Reproductive rates of tsetse flies in the field in Zimbabwe. *Physiological Entomology*, vol. 19, no. 4, pp. 307–318.

Hargrove, J.W. (1995). Towards a general rule for estimating the stage of pregnancy in field-caught tsetse flies. *Physiological Entomology*, vol. 20, no. 3, pp. 213–223.

Hargrove, J.W. (1999). Reproductive abnormalities in tsetse flies in Zimbabwe. *Entomologia Experimentalis et Applicata*, vol. 92, no. 1, pp. 89–99.

Hargrove, J.W. (2001*a*). Factors affecting density–independent survival of an island population of tsetse flies in Zimbabwe. *Entomologia Experimentalis et Applicata*, vol. 100, no. 2, pp. 151–164.

Hargrove, J.W. (2001*b*). Mark-recapture and Moran curve estimates of the survival probabilities of an island population of tsetse flies *Glossina morsitans morsitans* (Diptera: *Glossinidae*). *Bulletin of Entomological Research*, vol. 91, no. 1, pp. 25–35.

Hargrove, J.W. (2001*c*). The effect of temperature and saturation deficit on mortality in populations of male *Glossina m. morsitans* (Diptera: *Glossinidae*) in Zimbabwe and Tanzania. *Bulletin of Entomological Research*, vol. 91, no. 2, pp. 79–86.

Hargrove, J.W. (2012). Age-specific changes in sperm levels among female tsetse (*Glossina* spp.) with a model for the time course of insemination. *Physiological Entomology*, vol. 37, no. 3, pp. 278–290.

Hargrove, J.W. and Ackley, S.F. (2014). Mortality estimates from ovarian age distributions of the tsetse fly *Glossina pallidipes* Austen sampled in Zimbabwe suggest the need for new analytical approaches. *Journal of Animal Ecology [in preparation]*.

Hargrove, J.W., Ouifki, R. and Ameh, J.E. (2011). A general model for mortality in adult tsetse (Glossina spp.). *Medical and Veterinary Entomology*, vol. 25, no. 4, pp. 385–394.

Hargrove, J.W. and Williams, B.G. (1998). Optimized simulation as an aid to modelling, with an application to the study of a population of tsetse flies, *Glossina morsitans morsitans* (Diptera: *Glossinidae*). *Bulletin of Entomological Research*, vol. 88, no. 4, pp. 425–435.

Jolly, G.M. (1965). Explicit estimates from capture-recapture data with both death and immigration - stochastic model. *Biometrika*, vol. 52, pp. 225–247.

Maudlin, I., Welburn, S.C. and Milligan, P.J.M. (1998). Trypanosome infections and survival in tsetse. *Parasitology*, vol. 116, no. S1, pp. S23–S28.

Phelps, R.J. and Burrows, P.M. (1969*a*). Prediction of the pupal duration of *Glossina morsitans orientalis* Vanderplank under field conditions. *Journal of Applied Ecology*, vol. 6, no. 2, pp. 323–337.

Phelps, R.J. and Burrows, P.M. (1969*b*). Puparial duration in *Glossina morsitans* orientalis under conditions of constant temperature. *Entomologia Experimentalis et Applicata*, vol. 12, no. 1, pp. 33–43.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Available at: `http://www.R-project.org/`

Rogers, D. (1979). Tsetse population dynamics and distribution: a new analytical approach. *The Journal of Animal Ecology*, pp. 825–849.

Seber, G. (1965). A note on the multiple recapture census. *Biometrika*, vol. 52, pp. 249–259.

Statistics Canada (2014). *ModGen: A Generic Microsimulation Programming Language*. Ottawa, Ontario.
Available at: `http://www.statcan.gc.ca/microsimulation/modgen/modgen-eng.htm`

Vale, G.A., Hargrove, J.W., Cockbill, G.F. and Phelps, R.J. (1986). Field trials of baits to control populations of *Glossina morsitans morsitans* Westwood and *G. pallidipes* Austen (Diptera: *Glossinidae*). *Bulletin of Entomological Research*, vol. 76, no. 2, pp. 179–193.

Vale, G.A. and Torr, S.J. (2005). User-friendly models of the costs and efficacy of tsetse control: application to sterilizing and insecticidal techniques. *Medical and Veterinary Entomology*, vol. 19, pp. 293–305.

Woolhouse, M.E.J. and Hargrove, J.W. (1998). On the interpretation of age–prevalence curves for trypanosome infections of tsetse flies. *Parasitology*, vol. 116, no. 2, pp. 149–156.

Woolhouse, M.E.J., Hargrove, J.W. and McNamara, J.J. (1993). Epidemiology of trypanosome infections of the tsetse fly *Glossina pallidipes* in the Zambezi Valley. *Parasitology*, vol. 106, no. 5, pp. 479–485.

Woolhouse, M.E.J., McNamara, J.J., Hargrove, J.W. and Bealby, K.A. (1996). Distribution and abundance of trypanosome (subgenus *Nannomonas*) infections of the tsetse fly *Glossina pallidipes* in southern Africa. *Molecular Ecology*, vol. 5, no. 1, pp. 11–18.