

# Optimized simulation as an aid to modelling, with an application to the study of a population of tsetse flies, *Glossina morsitans morsitans* (Diptera: Glossinidae)

J.W. Hargrove<sup>1\*</sup> and B.G. Williams<sup>2</sup>

<sup>1</sup>Tsetse Control Branch, Department of Veterinary Services, Box CY52, Causeway, Harare, Zimbabwe: <sup>2</sup>Epidemiological Research Unit, Box 30606, Braamfontein 2017, South Africa

## Abstract

A method is described for optimizing models by linking simulation procedures with a non-linear regression programme. It is then possible to work towards a parsimonious model which contains those, and only those, variables required for an optimum fit. Using the observed values, and the predicted values from each simulation, the optimizing routine calculates the value of an appropriate loss function. It then makes small changes in the parameters governing the simulation, recalculates the predicted values and the first and second derivative of the loss function with respect to each parameter. The algorithm uses this information to minimize the loss function for a given formulation of the model. The model is improved by adding variables which can be shown statistically to improve the fit, and by removing those which do not. The use of the technique is illustrated with reference to a series of weekly estimates of the total numbers, births and survival probabilities of a population of male and female tsetse flies *Glossina morsitans morsitans* Westwood. Simulation involved following the lives of cohorts of flies, and of all their progeny, from the time they were deposited as larvae. Development and reproduction were regarded as fixed functions of temperature, but mortality rates of pupae and of adult flies depended on meteorological and biological variables, plus the level of trapping imposed on the population. Potential factors were added singly and the model thereby improved in an objective, stepwise manner. The best fit was achieved when effects on adult survival due to maximum temperature, various modes of trapping, and an annual cycle were included in the model. The optimized simulation technique has been used here in improving a model which describes a biological population but it could equally be used to improve models in any situation where data are fitted using simulation procedures.

## Introduction

Many real-life modelling problems are sufficiently complex that they cannot satisfactorily be approximated by

a set of analytical equations. Workers sometimes then opt for the simpler approach of simulation which requires less mathematical sophistication, but can still give insights into the dynamics of the problem. As examples, Haile & Mount (1987) and Mount *et al.* (1991) modelled tick populations and Rogers (1990) and Brightwell *et al.* (1997) modelled tsetse (*Glossina* spp.) populations using simulation. The studies

\*Fax: 263 4 722684

E-mail: jhargrove@rttcp.gaia.co.zw

varied in their scope and detail, and the fit to the data was better in some cases than others. What the studies had in common were that: (i) no detail was given of the process by which it was decided that the 'best fit' had been achieved with the model chosen; (ii) variances and covariances were not estimated for the model parameters; and (iii) it was therefore impossible to judge whether an optimal fit had been achieved with the variables chosen, or whether all of these variables were necessary in the model.

The above is not meant as a criticism of the studies cited; they seem typical of many simulation studies in biology, probably because simple methodical techniques for producing optimal, parsimonious models are either not readily available, or are at least unknown to many workers in the field (D. Haile and R. Brightwell, personal communication). Whereas sensitivity analysis may be carried out to judge how the simulation is affected by changes in the model's parameters, this is often not carried out in such a way that parameter variances and covariances can be estimated; nor is it used to obtain an optimal fit to the data.

These problems are addressed here by linking a simulation procedure to a non-linear optimization routine. For a given model, this enables the iterative estimation of a set of parameter values associated with an optimum fit to the data. The model is improved by adding or removing variables following statistical comparisons between the best fits achieved when these variables are included or excluded from the model. The variances and covariances for each parameter are estimated as part of the process. There is nothing new in the use either of simulation or of non-linear minimization techniques. What is unusual is the linking of the two to allow iterative optimization of the simulation procedure.

## Methods

### *Linking simulation and iterative minimization*

In the standard use of iterative minimization routines, the parametric model under consideration is such that, given a set of values for the independent parameters, the predicted values can be computed directly. Simulation procedures are typically used in cases where the predicted values are not easily generated from explicitly defined equations. Nonetheless, the simulation process produces a set of predicted data that is dependent on the specified values of the parameters in the model. The aim is then to select a set of parameter values that minimizes the difference between the observed and predicted data.

It is shown below that this aim can be achieved by linking the simulation process to an iterative minimization routine. In many standard statistical packages this option is not available, but it is possible with SEARCH (© Copyright Kenneth Lange, 1985–1991) which is a FORTRAN 77 subprogram for function minimization. In order to carry out the minimization the following are required: (i) a set of observed data; (ii) a parametric model to describe the data; (iii) a loss function which measures the difference between the observed and predicted data; and (iv) starting values for each parameter in the model.

The minimization starts by producing a set of predicted data with the initial parameter values as inputs. In the present case, these predictions are produced using simulation. The difference between the observed and predicted

data is used to calculate a value for an appropriate loss function. Each parameter in turn is then perturbed slightly (positively then negatively, with all other parameters kept at their initial values) and the predicted data and loss functions recalculated via simulation. This allows the programme to identify the change in the loss function, and its first and second derivatives, with respect to each parameter. The information is used to select a new set of parameter values which, when used again to recalculate the predicted data, results in a reduction in the loss function. This procedure is iterated automatically until no perturbation can be found which results in a decrease in the loss function.

SEARCH follows the prescriptions of Powell (1978) on recursive quadratic programming. It differs in the application of a sweep operator for solving quadratic programming problems (Jennrich & Sampson, 1978) and in the use of the matrix of second derivatives of the function to be minimized.

### *Application to a particular problem*

The data to be modelled consist of a set of Jolly-Seber (J-S) estimates (Jolly, 1965; Seber, 1965) of survival probabilities, births and total population numbers, resulting from a multiple mark-recapture exercise carried out on a population of tsetse flies *Glossina morsitans morsitans* Westwood (Diptera: Glossinidae) in 1980–1983 (Vale *et al.*, 1986). The population grew from a stock of pupae introduced onto Antelope Island, Lake Kariba, Zimbabwe with a view to testing odour-baited traps and 'targets' against a tsetse population of known size. Vale *et al.* (1986) provide the background to the experiment and a preliminary analysis of the data.

The delays involved in reproduction, and the fact that the rate of development changes with temperature from day to day during larval and pupal life, make it particularly difficult to develop a realistic analytical model of tsetse populations. These factors do not pose a problem for a simulation routine; at each point in the simulation the development rates are simply modified according to the appropriate temperature (see below).

As a preliminary to the description of this example of optimized simulation, a brief account of the tsetse fly life cycle is provided for those unfamiliar with its biology.

### *The tsetse life cycle*

Female tsetse mature only one egg at a time. They ovulate for the first time at age *c.* 8 days; the egg hatches in the uterus and the resulting larva is nourished through three instars via a specialized 'milk gland'. Pregnancy lasts *c.* 9 days at 25°C. The mature third instar larva does not feed after parturition; it burrows under the ground to the depth of an inch or so, and pupates inside a puparial case. It then develops over the following *c.* 30 days at 25°C into the adult fly.

### *Simulation procedure*

The fates of cohorts of individual males and female, and the progeny of the latter, were followed. This simulation procedure involved the following steps:

1. Define the data set to be fitted. The J-S estimates of the total population of male and of female *G. m. morsitans*, and the numbers of births, were viewed as a series of estimates of a single dependent variable. The sum of squared differences between observed and predicted estimates was used as a loss function; for the purposes of estimation, each value was weighted by the inverse of the J-S estimate of its variance.

2. Define an initial distribution of (cohorts of) individuals. In the actual experiment, tsetse pupae were collected in August–October 1979, in the field at Rekomitjie Research Station, Zambezi Valley, Zimbabwe. The pupae were placed in trays of sand and allowed to emerge naturally. The results only record the number of pupae collected on each day and there is no way of knowing the stage of development of each pupa, nor its sex. At a first approximation, it was assumed that there was a 1:1 sex ratio and that the numbers of each were uniformly distributed on  $[0, P]$  where  $P$  is the pupal period pertaining to the mean temperature ( $T$ , in °C) at the time when the pupae were collected. The expected puparial durations in days for males ( $P_m$ ) and females ( $P_f$ ) were estimated by:

$$P_m = 1 / (0.05415 / (1 + \exp(4.8184 - T \cdot 0.2149)))$$

$$P_f = 1 / (0.05884 / (1 + \exp(4.8829 - T \cdot 0.2159)))$$

(Phelps & Burrows, 1969). A starting distribution of males and females, each on their day of deposition as fully developed third instar larvae, was thereby estimated (fig. 1). Preliminary investigations showed that long-term population changes were not seriously affected by the distribution of the pioneering stock.

3. Define the time course of reproduction in females. For poikilotherms such as tsetse the rates depend on the mean temperature ( $T$ ). The expected time in days ( $I_f$ ) taken for a female to produce her first larva is given by:

$$I_f = 33.38 - T \cdot 0.690$$

(Hargrove, 1994) and the interval ( $I_n$ ) in days between subsequent pupae by:

$$I_n = 18.39 - T \cdot 0.352$$

(Hargrove, 1994, 1995). The preceding four equations were used, with temperature data, to produce tables which gave the emergence day for any male or female fly deposited as a larva on a given day, and the days on which a female, emerging on a given day, is expected to produce a larva. There was a problem in the selection of the mean temperature ( $T$ ) in these calculations. Hargrove (1994) found that tsetse development rates on an island in Lake Kariba were significantly higher than at the same screen temperature at Rekomitjie Research Station, Zambezi Valley, Zimbabwe. A possible reason for this was that the flies on the island were living at a temperature 3°C higher than indicated by the Stevenson screen temperatures. We assume that this is the case for the purposes of the present paper.

4. Define a model for calculating mortality during the experiment. The factors which might affect mortality can be climatic, e.g. saturation deficit (Nash, 1937) or temperature (Rogers, 1979), or biological, in the form of density dependent effects (Rogers & Randolph, 1985; Rogers, 1990) or related to the levels of mortality imposed by the experimenters using traps and odour-baited targets (Vale *et al.*, 1986). The idea is to find the best fit with a very simple model, then add one candidate variable factor at a time in an

attempt to improve the fit. Initially, the mortality was taken as zero in the immature phase, constant in males and a linear function of maximum temperature in females. Formally:

$$\mu(I) = 0.0$$

$$\mu(M) = m_0$$

$$\mu(F) = f_0 + f_1 T_{max} \quad (1)$$

where  $\mu(I)$ ,  $\mu(M)$  and  $\mu(F)$  are the mortality rates in immature, adult male and adult female tsetse respectively;  $f_0$  and  $m_0$  are constants;  $f_1$  is the coefficient for the effect of maximum temperature  $T_{max}$ . The most complete model used, with a list of all of the variables and their associated parameters, is shown in table 1.

5. Create tables defining each mortality factor at all times during the experiment. These include tables of  $T_{max}$  and the number of traps and targets deployed on each day. Odour-baited traps were used during the last 3 h of the afternoon in weeks 129–204; and for the first 3 h of the morning in weeks 168–204. From week 205 until the end of the experiment the traps were replaced with odour-baited 'targets' which operated all day (Vale *et al.*, 1986). The targets consisted of black cotton cloth mounted on wire frames and sprayed with insecticide; tsetse reacting to the visual and olfactory stimuli of the targets alighted on the cloth and acquired a lethal dose of the insecticide.

6. Pick starting values for the coefficients defined in step 4.

7. Using the information provided by steps 2–6, produce a particular trajectory of the population. This procedure involves the following steps:

A. Consider the fate of the cohort of female pupae deposited on day 1 of the experiment.

i. From the appropriate table, predict the emergence day of females born on day 1.

ii. For each day of female pupal life, calculate the survival probability using Equation 1.

iii. The product of these survival values gives the proportion of the cohort emerging as adults.

iv. Using the daily adult survival, estimate the number of females from this cohort alive at the beginning of each week of the experiment. Accumulate these values.

v. When the cohort falls below an arbitrary value (we chose 1.0) it is considered 'extinct'.

B. Calculate the pupal production by each cohort.

vi. Check, using the information in the table referred to in step 3 in the text, whether a surviving female would have deposited a larva during each week. If so, store a number of pupae (on the appropriate day) equal to the number of females (rounded to the nearest integer) surviving to that time. Assume equality of sexes among the pupae unless the number is odd, in which case the extra offspring is assumed female. This reflects the slight sex bias seen in the laboratory.

C. Repeat for all days of the experiment

vii. Repeat steps i–vi for each day of the experiment. Notice that each cohort of females produces larvae whose fate is in turn considered as the simulation proceeds.

D. Consider the fate of all male pupae deposited on each day of the experiment.

viii. Repeat steps i–v for the male pupae.

The resulting trajectories, of births and total populations for each sex, form the input for the minimization routine, described in steps 8–17. These steps are independent of the

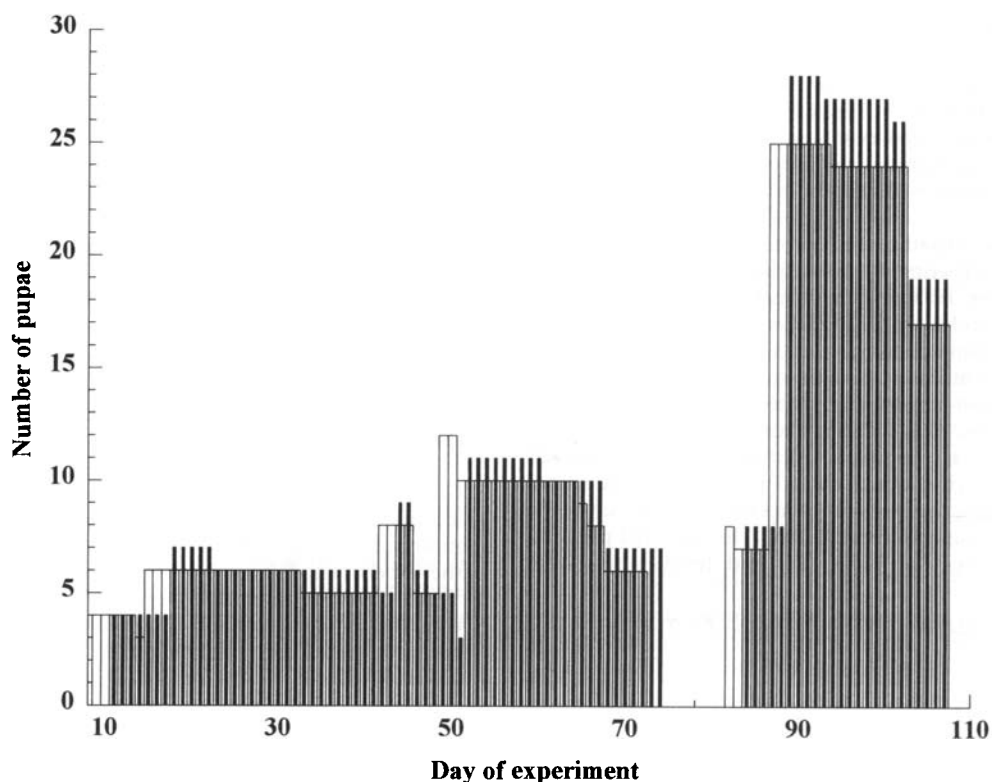


Fig. 1. The estimated distribution of 0-day-old pupae of *Glossina morsitans morsitans* introduced to Antelope Island, Lake Kariba, Zimbabwe. Key: Open histograms, males; solid histograms, females.

problem being tackled. For a different modelling problem it would therefore be necessary only to provide input, using appropriate simulation procedures, analogous to that provided by steps 1–7.

8. Calculate the residual sums of squared differences

(RSS) between the observed and predicted functions. In this particular case we sum, over all weeks and for both sexes, the squares of differences between the observed and expected populations and births.

9. Make a small increase in one coefficient in Equation 1,

Table 1. The equations determining mortality in immature and adult stages of *Glossina morsitans morsitans* in a model of the dynamics of an island population of that species.

Symbol	Description
$T_{max}(i)$	Maximum temperature on day $i$ ( $^{\circ}\text{C}$ ).
$C(i)$	Point on the annual cycle of day $i$ .
$A(i)$	Number of traps operating on the morning of day $i$ .
$P(i)$	Number of traps operating on the afternoon of day $i$ .
$S(i)$	Number of targets operating on day $i$ .
$\phi_p(i)$	Probability of survival of a pupa on day $i$ .
$\phi_f(i)$	Probability of survival of an adult female on day $i$ .
$\phi_m(i)$	Probability of survival of an adult male on day $i$ .

The full model defining mortality is:

$$\phi_p(i) = \exp(-(p_0 + p_1 T_{max}(i)))$$

$$\phi_f(i) = \exp(-(f_0 + f_1 T_{max}(i) + f_2 C(i) + f_3 A(i) + f_4 P(i) + f_5 S(i)))$$

$$\phi_m(i) = \exp(-(m_0 + m_1 T_{max}(i) + m_2 C(i) + m_3 A(i) + m_4 P(i) + m_5 S(i)))$$

where the  $p^*$ ,  $f^*$  and  $m^*$  are the parameters which are varied to achieve the least squares fit. Subsets of the full model are produced by setting particular parameters to zero. Notice that, in all forms of the model considered here, the survival probability is of the form  $\phi = \exp(-\mu)$  where  $\mu$ , the mortality rate, is a linear function of the model variables.

keeping all other coefficients at their original levels. Repeat steps 7 and 8.

10. Repeat step 9 making a small decrease in this coefficient. (The size of the changes in steps 9 and 10 decreases with the sensitivity of the RSS to changes in the parameter).

11. Repeat steps 9 and 10 for each coefficient with all other coefficients at their original levels.

12. Lange's algorithm now has a measure of the change, and rate of change, in the RSS resulting from changes in each of the coefficients in Equation 1. It uses this information to select a new set of values of the coefficients which moves the RSS towards its minimum value.

13. Repeat steps 7–12 until the difference between the RSS in consecutive iterations is less than some arbitrarily stipulated value.

14. The algorithm calculates the standard errors of all coefficients and their covariance matrix.

15. Calculate the  $F$  statistic (Snedecor & Cochran, 1980):

$$F_{a,b} \cong ((RSS_R - RSS_F)/(P_F - P_R))/(RSS_F/(N - P_F))$$

to test for a significant reduction in the RSS as a consequence of adding the variable in question.  $RSS_R$  is the RSS in the restricted model;  $RSS_F$  is the RSS in the full model;  $P_F$  and  $P_R$  are the corresponding numbers of parameters and  $N$  is the number of observations. The degrees of freedom are given by  $a = N - P_F$  and  $b = P_F - P_R$ . In this study,  $N = 820$  and  $P_F - P_R$  was always 1, since only one variable was ever added or removed at a time. The critical levels of  $F$  at the 0.05, 0.01, 0.001 and 0.0005 levels are c. 3.8, 6.6, 10.8 and 12.1 respectively.

16. Retain the variable if it had a significant effect, reject it otherwise.

17. Add another variable to Equation 1 (step 4) and repeat steps 1–16.

It is emphasized that optimized simulation can only be used to improve models by comparing the best fits achieved with different formulations of that model; the onus is on the modeller to select candidate variables which might be of use in this regard.

## Results

### Adult female mortality

The initial model assumed constant adult mortality in male *G. m. morsitans* and a linear increase with maximum temperature in females. Not surprisingly, the resulting best fit to estimates of total population and births was poor (figs 2 and 3; trajectory 1) with an RSS value  $>10^5$  (table 2, row 1). For male survival values (assumed constant), the fit was obviously poor (fig. 4, males). Surprisingly, however, the best fit to the female survival data was quite good, even though the model only included the effect of  $T_{max}$  (fig. 4, females) and the fits to the population and birth data were poor (cf. trajectory 1 in figs 2–4, females). Clearly, quite small changes in female survival probability have major effects on population levels. This is expected from the known dynamics of tsetse populations (Hargrove, 1988) and forms the basis of 'bait' methods of tsetse control (Vale *et al.*, 1986, 1988).

Since females form the productive part of any population, changes in their mortality strongly affect the numbers of both sexes. Accordingly, model development concentrated first on the factors affecting females. Introducing terms for the

effects, on female mortality, of the numbers of targets deployed, and of the numbers of traps deployed in the afternoon and morning reduced the RSS by 70% (table 2; rows 1–4).

### Adult male mortality

Inclusion of an effect of maximum temperature on male mortality significantly reduced the RSS (table 2; row 5). Interestingly, the parameter estimate was twice as high in males as in females, suggesting that the latter are more sensitive to high temperatures. Similarly, the addition of each factor relating to the degree of killing of males by traps and targets caused a significant decline in the RSS and in each case the parameter estimates were higher for males than for females (table 2; rows 6–8).

Attempts were now made to improve the model by including other meteorological factors such as saturation deficit and humidity but none of these was successful. Nonetheless, there did appear to be other seasonal effects. Inspection of the mortality data suggested that after variance due to temperature and trapping had been removed there was a residual effect, which showed a strong annual cycle, not obviously related to any meteorological measurement. This phenomenon will be investigated more thoroughly elsewhere; for present purposes a factor was entered consisting of a sinusoidal wave with period one year, range  $[-1.0, 1.0]$ , with maximum in the second week of May and minimum in the second week of November. Inclusion of this factor for males and females gave rise to a further substantial reduction in the RSS (table 2; rows 9 and 10).

### Mortality in the immature stages

The factors entered up to this point refer only to mortality in the adult stages. It has been assumed so far that there was no mortality during the immature stages. The results in rows 11 and 12 (table 2) are for simulations where this mortality was introduced, either as a fixed level of mortality, or at a level increasing linearly with maximum temperature. There were further significant decreases in the RSS, although the coefficients of variation for the two factors were large. A visual estimate of the progressive improvement in the fit is shown in figs 2–4. If a model with  $p$  parameters accounted for all of the variability (apart from that due to noise), the RSS would come from a  $\chi^2$  distribution with  $N - p$  degrees of freedom. The critical level (when  $N = 820$  and  $p = 12$ ) is 877. The RSS for the final model is 4026, indicating that there is still some unexplained variation in the data.

The coefficients of variation (CVs) of the parameter estimates in the final run (table 2, row 12) show that, for all coefficients relating to adult mortality, the CVs are  $<1\%$ . For the parameters  $p_0$  and  $p_1$ , associated with losses in the pupal phase, the CVs are much higher (table 2) and are more highly correlated with each other, and with several other parameters, than the adult parameters are with each other (table 3). The correlations are still within reasonable bounds, however, and there are good biological reasons for retaining the two parameters.

Thus, larval/pupal mortality is certainly greater than zero, and there are good reasons for thinking that it depends on temperature. The higher CVs may relate to the discrepancy between the data sets for population numbers and for births. The mean values for the latter set are smaller

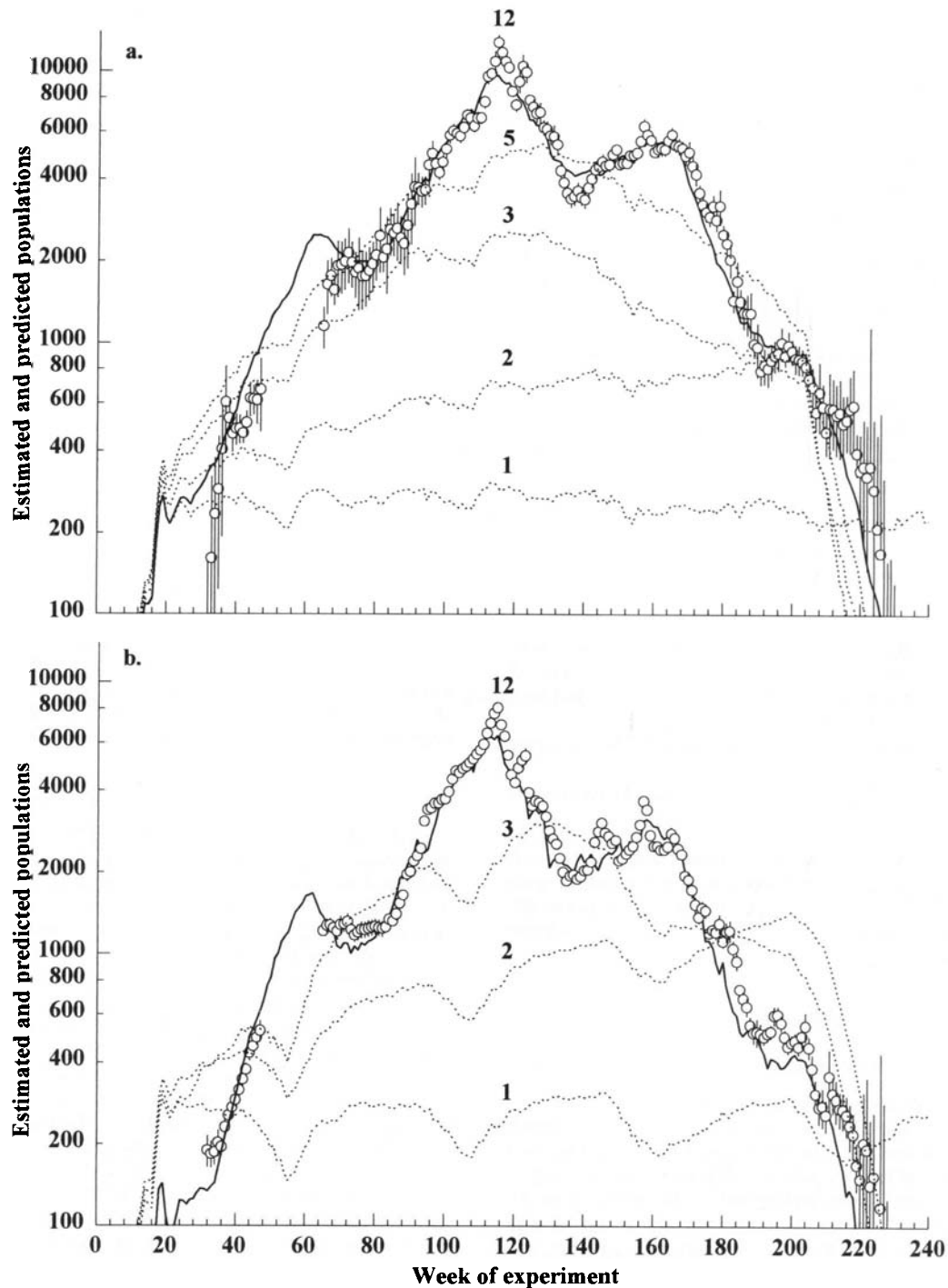


Fig. 2. Weekly Jolly-Seber estimates (circles; with standard error bars) of the total populations of female (a) and male (b) *Glossina morsitans morsitans* on Antelope Island, Lake Kariba, Zimbabwe. The numbered lines are the best fits achieved with the models summarized by the appropriate row numbers in table 2.

and their standard errors relatively much bigger (cf. figs 2 and 3). The birth estimates therefore have very much less weight in the fitting process than the estimates of total population; but it is specifically the information on births which inform us about losses in the immature stages.

#### *Density dependent mortality*

The introduction into the model of density-dependent effects on adult or pupal mortality failed to result in any significant reduction in the RSS. Given that the measured

changes in population were greater than three orders of magnitude, this was surprising. Equally surprising was the fact that it was possible to obtain a good fit to the data without recourse to density dependent

terms. Further investigation of this complex matter is beyond the scope of the present study which aims primarily to establish the technique of optimized simulation.

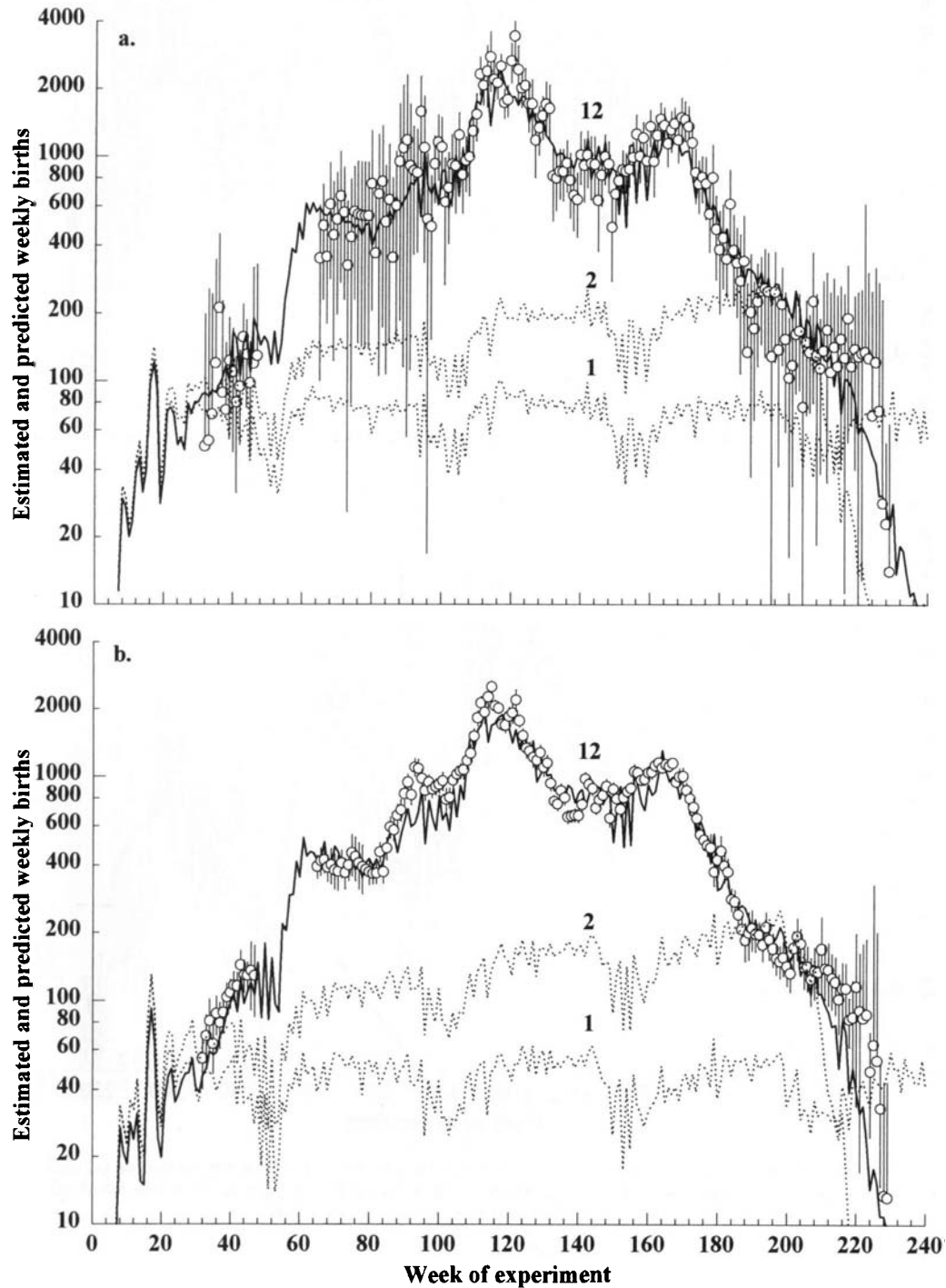


Fig. 3. Weekly Jolly-Seber estimates (circles; with standard error bars) of the numbers of births of female (a) and male (b) *Glossina morsitans morsitans* on Antelope Island, Lake Kariba, Zimbabwe. The numbered lines are the best fits achieved with the models summarized by the appropriate row numbers in table 2.

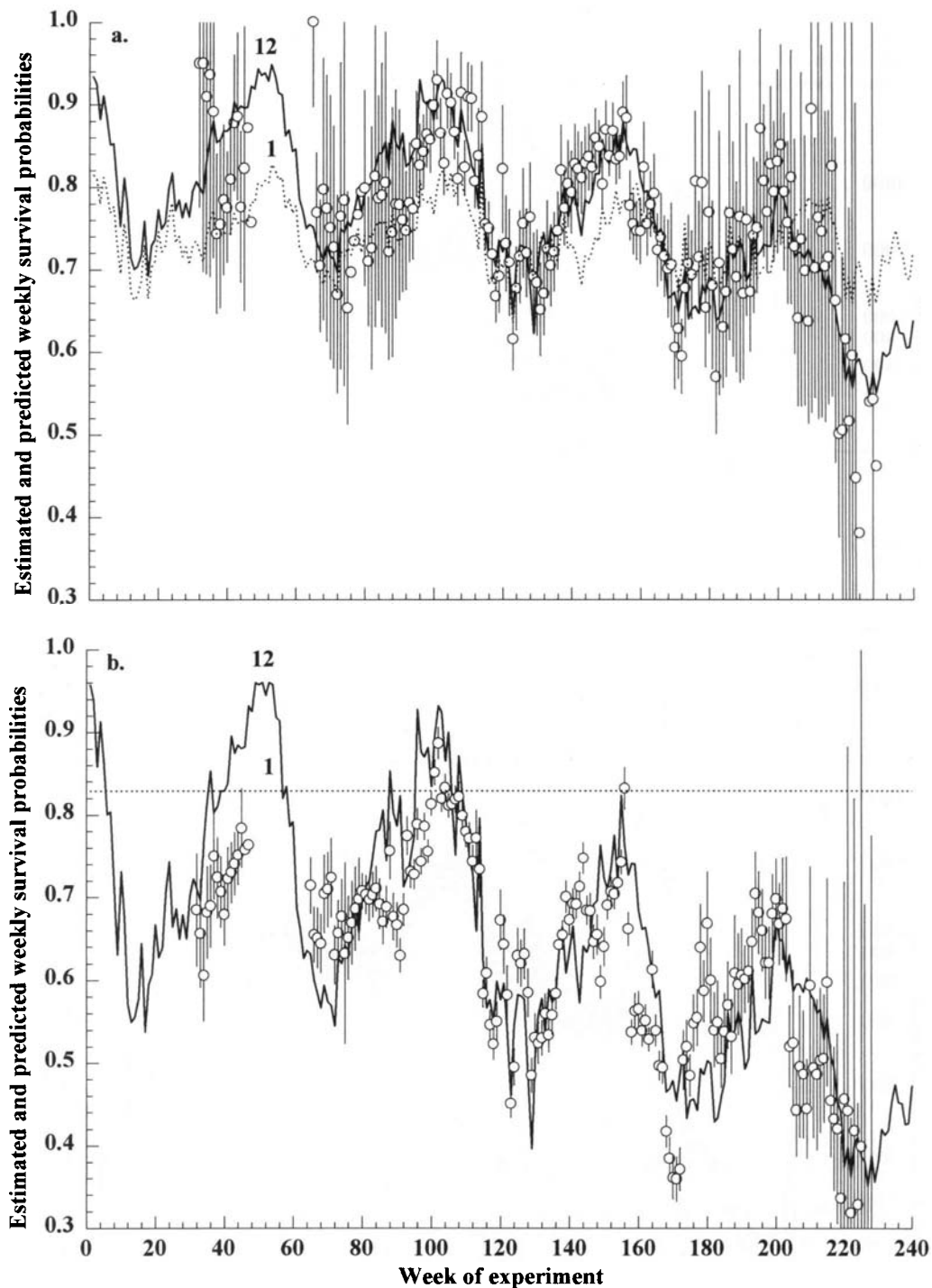


Fig. 4. Weekly Jolly-Seber estimates (circles; with standard error bars) of the survival probability of female (a) and male (b) *Glossina morsitans morsitans* on Antelope Island, Lake Kariba, Zimbabwe. The numbered lines are the best fits achieved with the models summarized by the appropriate row numbers in table 2.

#### *Fitting subsets of the data*

The fitting technique was applied with success to various subsets of the *G. m. morsitans* data set. Thus it was possible to obtain good fits to the adult data only, the data for each

sex taken singly, and for subsets of the experimental period. Indeed it was helpful, in getting a first approximation to the effect of meteorological factors, to focus initially on the early part of the experiment (weeks 29–43) when there was no mortality due to trapping devices. The subset was then



Table 2. Parameter estimates in various simulations of the Antelope Island tsetse population data. The values below the coefficients in row 12 are their standard errors (SE) and coefficients of variation (CV = standard error  $\times$  100/mean).

n	Parameters															RSS	F
	Pupae		Mature females						Mature males								
	$p_0$ Const.	$p_t$ $T_{max}$	$f_0$ Const.	$f_i$ $T_{max}$	$f_c$ Cycle	$f_a$ am trap	$f_p$ pm trap	$f_{is}$ Target	$m_0$ Const.	$m_t$ $T_{max}$	$m_c$ Cycle	$m_a$ am trap	$m_p$ pm trap	$m_{is}$ Target			
1			-0.219	0.0169					0.188							0.163.10 <sup>6</sup>	–
2			-0.202	0.0156				0.337	0.165							0.113.10 <sup>6</sup>	388.6
3			-0.225	0.0152			0.0655	0.285	0.234							0.535.10 <sup>5</sup>	911.0
4			-0.225	0.0151		0.0356	0.0652	0.285	0.232							0.463.10 <sup>5</sup>	121.2
5			-0.226	0.0148		0.0352	0.0641	0.285	-0.311	0.0247						0.326.10 <sup>5</sup>	340.9
6			-0.226	0.0148		0.0355	0.0639	0.286	-0.313	0.0242		0.0311				0.317.10 <sup>5</sup>	24.9
7			-0.228	0.0148		0.0362	0.0634	0.300	-0.318	0.0224		0.0336	0.187			0.311.10 <sup>5</sup>	13.7
8			-0.245	0.0153		0.0309	0.0753	0.296	-0.568	0.0300		0.0513	0.143	0.502		0.261.10 <sup>5</sup>	157.3
9			-0.261	0.0157	-0.0635	0.0346	0.0803	0.261	-0.582	0.0288		0.0710	0.173	0.493		9743	1355.3
10			-0.265	0.0159	-0.0709	0.0406	0.0768	0.280	-0.565	0.0283	-0.127	0.0850	0.215	0.485		4532	929.4
11	0.006		-0.308	0.0163	-0.0670	0.0463	0.0755	0.228	-0.693	0.0320	-0.124	0.0949	0.212	0.410		4283	46.8
12	0.00463	0.00102	-0.308	0.0165	-0.0663	0.0448	0.0731	0.221	-0.693	0.0321	-0.124	0.0919	0.206	0.397		4022	52.3
SE	0.00006	0.00010	0.0001	0.00001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001			
CV	(1.37)	(9.54)	(0.03)	(0.08)	(0.14)	(0.21)	(0.14)	(0.04)	(0.01)	(0.27)	(0.08)	(0.11)	(0.05)	(0.02)			

Key for survival parameters: The variables corresponding to each parameter are shown below the parameter in question; Const., constant term; RSS, residual sums of squares; F, F statistic for the effect of the added variable.

Table 3. Asymptotic correlation matrix of the parameters used in run 12 (see table 2).

	Parameters													
	Pupae		Mature females						Mature males					
	$p_0$ Const.	$p_t$ $T_{max}$	$f_0$ Const.	$f_t$ $T_{max}$	$f_c$ Cycle	$f_a$ $am$ trap	$f_p$ $pm$ trap	$f_{ig}$ Target	$m_0$ Const.	$m_t$ $T_{max}$	$m_c$ Cycle	$m_a$ $am$ trap	$m_p$ $pm$ trap	$m_{ig}$ Target
$p_0$	1.000													
$p_t$	0.261	1.000												
$f_0$	-0.424	-0.374	1.000											
$f_t$	0.594	-0.472	-0.118	1.000										
$f_c$	0.005	0.008	-0.004	-0.0001	1.000									
$f_a$	-0.387	-0.433	0.028	-0.055	-0.004	1.000								
$f_p$	-0.476	-0.422	0.037	-0.131	-0.005	0.033	1.000							
$f_{ig}$	0.0002	0.0004	-0.0002	-0.0001	0.00001	-0.0002	-0.0002	1.000						
$m_0$	0.005	0.008	-0.005	-0.001	0.0001	-0.004	-0.005	0.00001	1.000					
$m_t$	0.094	0.155	-0.086	-0.010	0.002	-0.084	-0.097	0.0001	0.002	1.000				
$m_c$	0.006	0.012	-0.007	-0.002	0.0002	-0.006	-0.007	0.00001	0.0002	0.003	1.000			
$m_a$	-0.001	-0.002	0.001	0.0005	-0.00003	0.001	0.001	-0.00001	-0.00003	-0.0006	-0.00005	1.000		
$m_p$	0.003	0.006	-0.003	-0.001	0.0001	-0.003	-0.004	0.00001	0.0001	0.002	0.0001	-0.00002	1.000	
$m_{ig}$	-0.00003	-0.00006	0.00003	0.00001	-0.00001	0.00002	0.00004	-0.00001	-0.00001	-0.00001	-0.00001	0.00001	-0.00001	1.000

Key for survival parameters: The variables corresponding to each parameter are shown below the parameter in question; Const., constant term.

expanded to include all data up to week 110, when the populations peaked, then to week 168 to include the period when traps were used in the afternoon only. Finally, all data were included. Good fits were obtained with every subset and there was always an effect of temperature on adult mortality.

### Discussion

The optimized simulation technique has the advantage of automating the process of estimating the parameter values which give the best fit to a given data set with a particular model formulation. In so doing, the procedure estimates the variance and covariance of the model parameters and hence the sensitivity of the fit to changes in those parameters. None of these objectives is generally achieved using existing techniques and simulation studies often therefore lack these details (see Introduction for examples). The present method also facilitates comparisons between the best fits achieved with various models. In the present case it was thus possible to show, on the one hand, that there were separate effects of the different trapping regimes and, on the other, that factors such as saturation deficit and humidity were not required in the model once an effect of temperature had been included.

In modelling the tsetse data, we are aware that we have ignored difficult problems concerning numerical stability, multi-modal surfaces and parameter and observational correlation. These arise in the modelling of any complex data set and will be dealt with in the production of more definitive models of the Antelope Island tsetse populations. They are, however, irrelevant to the principle of optimized simulation whose introduction was our aim.

The technique has been applied, in this instance, to an extraordinarily complete data set where birth, survival and total population estimates were available over a long period. However, it could equally be used on less complete data, such as estimates of total population alone. Indeed, it could be used in any situation where simulation is applicable. Its great advantage is its simplicity. In that respect the technique is a useful adjunct to more sophisticated modelling, where it is not clear how to formulate the problem mathematically. Instead of solving a series of difficult mathematical problems, optimized simulation might allow a sufficiently good approximation to be achieved such that only one final mathematical formulation is necessary.

In the simulation approach developed above, the fate of cohorts of flies was followed using a deterministic process. Other approaches to population simulation have been investigated. One method, which preserved better the spirit of the biology, was to follow the fates of individual flies and to make the simulation entirely stochastic. The survival level on each day depended, as before, on the given model and the conditions prevailing at the time. The fly's survival on that day was then decided by the value of a random number selected from a uniform distribution on [0,1]. In principle the minimization could be carried out as above, but two problems arose. First, the simulations took much longer. Second, because of the discrete, stochastic nature of the problem, small changes in parameter values led to relatively large changes in the RSS. There were thus large numbers of local minima in the loss function that made it impossible to calculate derivatives.

The problem was overcome by using a simpler minimization routine. In this approach, as before, the value

of the loss function was calculated for given parameter starting values and again when small changes were made in each parameter separately. The changes which resulted in the largest decrease in the loss function were then used to define the parameter values for the next iteration. Derivatives of the loss function were not calculated so that the problems referred to above did not arise. On the other hand the rate of convergence was now linear and the minimization took 10–100 times as long as the deterministic process.

The model developed for the Antelope Island data set can undoubtedly be improved. The difficulty experienced in entering variables relating to immature survival rates has been mentioned, and no density dependent effects could be detected. Moreover, mortality was assumed independent of age among adults, whereas it is higher than average in very young and in old flies (Hargrove, 1990). Trap and target-related mortalities also increase with age (Hargrove, 1991). Further work is in progress on the model and on attempts to extend it to tsetse populations in other areas. The current exercise merely serves to illustrate the principle of optimized simulation and to show how it can be used to effect statistically significant improvements to a model in a simple, methodical, step-wise manner.

### Acknowledgements

We are grateful for the use of Zimbabwe Government facilities. We thank Professor Lange (University of Michigan) for supplying SEARCH, and the following for reading of the manuscript: Drs J. Holt, K. Lange, S. Torr, G. Vale, M. Warnes and M. Woolhouse. The work was supported by the Overseas Development Administration of the United Kingdom Government (Research Scheme 4259) and the International Atomic Energy Agency (Technical Co-operation Project Zimbabwe/5/004).

### References

- Brightwell, R., Dransfield, R.D., Stevenson, P. & Williams, B. (1997) Changes over twelve years in the populations of *G. pallidipes* and *Glossina longipennis* (Diptera: Glossinidae) subject to varying trapping pressure at Nguruman, south west Kenya. *Bulletin of Entomological Research* **87**, 349–370.
- Haile, D.G. & Mount, G.A. (1987) Computer simulation of population dynamics of the Lone Star tick, *Amblyomma americanum* (Acari: Ixodidae). *Journal of Medical Entomology* **24**, 356–369.
- Hargrove, J.W. (1988) Tsetse: the limits to population growth. *Medical and Veterinary Entomology* **2**, 203–217.
- Hargrove, J.W. (1990) Age-dependent changes in the probabilities of survival and capture of the tsetse fly *Glossina morsitans morsitans* Westwood. *Insect Science and its Application* **11**, 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* **81**, 43–50.
- Hargrove, J.W. (1994) Reproductive rates of tsetse flies in the field in Zimbabwe. *Physiological Entomology* **19**, 307–318.
- Hargrove, J.W. (1995) Towards a general rule for estimating the day of pregnancy of field-caught tsetse flies. *Physiological Entomology* **20**, 213–223.

- Jennrich, R.I. & Sampson, P.F. (1978) Some problems faced in making a variance component algorithm into a general mixed model program. In Gallant, A.R. & Gerig, T.R. (Eds) *Proceedings of the Eleventh Annual Symposium on the Interface*. Institute of Statistics, North Carolina State University.
- Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration – stochastic model. *Biometrika* **52**, 225–247.
- Mount, G.A., Haile, D.G., Davey, R.B. & Cooksey, L.M. (1991) Computer simulation of *Boophilus* cattle ticks (Acari: Ixodidae) population dynamics. *Journal of Medical Entomology* **28**, 223–240.
- Nash, T.A.M. (1937) Climate the vital factor in the ecology of *Glossina*. *Bulletin of Entomological Research* **28**, 75–127.
- Phelps, R.J. & Burrows, P.M. (1969) Prediction of the puparial duration in *Glossina morsitans orientalis* Vanderplank under field conditions. *Journal of Applied Ecology* **6**, 323–337.
- Powell, M.J.D. (1978) A fast algorithm for linearly constrained optimisation calculations. In Watson, G.A. (Ed.) *Proceedings of the 1977 Dundee Conference on Numerical Algorithms*. Berlin, Springer-Verlag.
- Rogers, D.J. (1979) Tsetse population dynamics and distribution: a new analytical approach. *Journal of Animal Ecology* **48**, 825–849.
- Rogers, D.J. (1990) A general model for tsetse populations. *Insect Science and its Application* **11**, 331–346.
- Rogers, D.J. & Randolph, S.E. (1985) Population ecology of tsetse. *Annual Review of Entomology* **30**, 197–216.
- Seber, G.A.F. (1965) A note on the multiple recapture census. *Biometrika* **52**, 249–259.
- Snedecor, G.W. & Cochran, W.G. (1980) *Statistical methods*. Ames, Iowa, Iowa State University Press.
- Vale, G.A., Hargrove, J.W., Cockbill, G.F. & 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* **76**, 179–193.
- Vale, G.A., Lovemore, D.F., Flint, S. & Cockbill, G.F. (1988) Odour-baited targets to control tsetse flies, *Glossina* spp. (Diptera: Glossinidae), in Zimbabwe. *Bulletin of Entomological Research* **78**, 31–49.

(Accepted 6 April 1998)

© CAB INTERNATIONAL, 1998