Estimating Mosquito Population Size from Mark—Release—Recapture Data

Author(s): D. Cianci , J. Van Den Broek , B. Caputo , F. Marini , A. Della Torre , H. Heesterbeek , and N. Hartemink

Source: Journal of Medical Entomology, 50(3):533-542. 2013.
Published By: Entomological Society of America
URL: http://www.bioone.org/doi/full/10.1603/ME12126

Estimating Mosquito Population Size From Mark–Release–Recapture Data

D. CIANCI,[1,2] J. VAN DEN BROEK,[1] B. CAPUTO,[3] F. MARINI,[3] A. DELLA TORRE,[3] H. HEESTERBEEK,[1] AND N. HARTEMINK[4]

J. Med. Entomol. 50(3): 533–542 (2013); DOI: http://dx.doi.org/10.1603/ME12126

ABSTRACT Accurate estimation of population size is key to understanding the ecology of disease vectors, as well as the epidemiology of the pathogens they carry and to plan effective control activities. Population size can be estimated through mark–release–recapture (MRR) experiments that are based on the assumption that the ratio of recaptured individuals to the total captures approximates the ratio of marked individuals released to the total population. However, methods to obtain population size estimates usually consider pooled data and are often based on the total number of marked and unmarked captures. We here present a logistic regression model, based on the principle of the well-known Fisher–Ford method, specific for MRR experiments where the information available is the number of marked mosquitoes released, the number of marked and unmarked mosquitoes caught in each trap and on each day, and the geographic coordinates of the traps. The model estimates population size, taking into consideration the distance between release points and traps, the time between release and recapture, and the loss of marked mosquitoes to death or dispersal. The performance and accuracy of the logistic regression model has been assessed using simulated data from known population sizes. We then applied the model to data from MRR experiments with *Aedes albopictus* Skuse performed on the campus of "Sapienza" University in Rome (Italy).

KEY WORDS *Aedes albopictus*, vector population, logistic regression model, vector-borne disease control, mark–release–recapture

Knowledge of the population size of vectors in endemic or epidemic areas is valuable for understanding disease transmission dynamics and for determining the extent of control measures necessary to interrupt transmission. A measure developed in the field of infectious disease epidemiology, the value of which is indicative of control effort required, is the basic reproduction number, R_0. R_0 combines, in a weighted way, the factors that determine whether a pathogen can become established in an area where it is introduced. One of the factors included in the R_0 formula, to which the numerical value is very sensitive, is the ratio of vectors to hosts (Hartemink et al. 2009, 2011). The more precise the estimate of the population size, the better the estimate of R_0.

Population size and other parameters of animal populations, such as dispersal, survival, and flight range are estimated by mark–release–recapture (MRR) experiments (Pollock et al. 1990). MRR methods have a long history in ecology (Le Cren 1965) and a large literature on the modeling of capture–recapture methods exists (Begon 1979; Seber 1982, 1986; White et al. 1982).

Peterson was the first to estimate population size as the number of marked animals released divided by the proportion of marked in the group of captures (Lincoln 1930, Le Cren 1965, Seber 1982, Krebs 1999). The same idea was used afterwards in other models, such as Fisher–Ford's method, which can take into account a death rate from literature (Fisher 1947). Jackson (1937, 1939, 1940) obtained the estimate of population size by fitting a curve to a series of recapture rates, obtained from the change in recapture in time. The program MARK (White 1999) provides parameter estimates from marked animals, including population size. Capture and recapture probability also can be modeled as a function of time, but the software is based on the history of individual animals and therefore may not be useful for most mosquito experiments. Extensive overviews of MRR techniques to estimate animal population abundance are provided in the literature (Pollock 1980, Seber 1986, Schwarz and Seber 1999, Gratz 2004, Efford 2009). Not all the information

The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

[1] Faculty of Veterinary Medicine, Utrecht University, The Netherlands.

[2] Corresponding author: Daniela Cianci Faculty of Veterinary Medicine, Utrecht University Yalelaan seven 3584 CL Utrecht, The Netherlands (e-mail: d.cianci@uu.nl).

[3] Dipartimento di Scienze di Sanità Pubblica, Università di Roma "Sapienza", Piazzale Aldo Moro 5, 00198, Roma, Italy.

[4] Georges Lematre Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, ouvain-la-Neuve, Belgium.

collected with MRR experiments has been taken into account simultaneously in one model. For instance, it may be important to account for the animal loss rate estimated from the data, because of mortality or dispersal, that recapture occurred on more than one day and to consider whether there is a biological limit to their dispersal distance.

We present a specific method to estimate the population abundance of mosquitoes by using data based on MRR experiments, where recaptures take place on different dates (time effect) and at different distances (distance effect). The method is based on a logistic regression that models the fraction of marked and unmarked mosquitoes as a function of population size, loss rate, distance between the release site and the traps, and number of mosquitoes released. This model is specific for data that do not distinguish individuals. We illustrate the principle by applying the method to data collected during MRR experiments with *Aedes albopictus* Skuse mosquitoes on the campus of "Sapienza" University in Rome (Italy). We also apply the model to simulated data to show how the model performs in different scenarios, including different population size, different number of marked mosquitoes released, and different effect of including data on distance.

Materials and Methods

Study Area and Mark–Release–Recapture Experiments. Three mark–release–recapture experiments were carried out, using the experimental design described in detail in Marini et al. (2010), on the campus of "Sapienza" University in Rome. The campus is situated in an urbanized area in the center of Rome; the buildings—typically 10–20 m high—are separated by lanes and green spaces. For the experiments, 55 sticky traps (STs), specifically designed to collect *Ae. albopictus* females searching for oviposition and resting sites (Facchinelli et al. 2007), were located at ground level in sheltered positions, distributed over an area of ≈22 ha (Fig. 1a) measured with Google Earth, version 5.2.1 (Google Earth Beta 2010). The Euclidean distances between the STs and the release points were calculated from their georeferenced locations by using ArcGIS (version 9.2, ESRI 2009). The number of sticky traps increases with the distance from the release site to keep the density of sticky-traps constant over space (Marini et al. 2010). *Aedes albopictus* adults used for MRR experiments were obtained from eggs collected by ovitraps within the study area. Eggs were hatched and larvae reared to the adult stage in plastic basins. The pupae were collected and transferred to emergence cages, where adults were maintained with 10% sugar solution for 2–9 d. On the morning of the release, females blood-fed by membrane feeders were marked with orange fluorescent dust and released from the center of the campus. The number of marked mosquitoes released was 464 in MRR1 (August of 2008), 566 in MRR2 (September of 2008), and 552 in MMR3 (October of 2008). Marked and unmarked mosquitoes in each ST were counted at days 5, 9, 13, 17, and 21 after release in MRR1 and at days 2–5, 7, 9, 13, 17, and 21 after release in MRR2 and MRR3. The recapture percentage in the campus, i.e., the number of marked recaptures divided by the number of mosquitoes released, was 4.5% for MRR1 (21 recaptures), 5.1% for MRR2 (29 recaptures), and 3.3% for MRR3 (18 recaptures) (Marini et al. 2010).

Data Used in the Analysis. Fig. 1b indicates, by the size of the pie chart, the total number of mosquitoes captured in the three experiments, with the fraction of recaptured mosquitoes in black. As shown in the figure, few released mosquitoes have been recaptured in the eastern part of the campus. This is probably because of the fact that the western part of the campus (where the mosquitoes were released) is to a large extent separated from the eastern part by an almost continuous line of high buildings, interrupted by few relatively small openings. Because it is likely that these buildings could have represented a physical barrier for the mosquito dispersion eastwards, we decided to use only the trapping results from the area situated on the western side of the buildings crossing the study area for our analyses.

Fig. 1a shows the 28 STs selected for the analysis that were situated in an area of 10.6 ha, indicated by the red line. A breakdown of the number of captured and recaptured mosquitoes in the selected area during each experiment is presented in Table 1. Because no marked mosquitoes were captured more than 9 d after the release in each of the three experiments (Marini et al. 2010), the analysis was restricted to the first 9 d. No marked mosquitoes were found farther than 210 m in the selected area. Therefore, the analysis was restricted to the traps in the range of 210 m from the release site in this area.

To gauge the performance of the model and to compare this in different scenarios, we simulated data for the same experimental set-up. We did this for an existing population of 21,000 unmarked mosquitoes and with the number of mosquitoes released being 150, 510, and 1,005, respectively. The loss rate per day was fixed at 0.05. The distance was considered in the simulated data as three pooled areas; in one scenario marked and unmarked mosquitoes were equally distributed in the three areas, in another scenario only the unmarked mosquitoes were equally distributed and the marked decreased when the distance from the release point increased. The former scenario is unrealistic in nature, but we used it to investigate whether adding a measure of distance improves the performance of the model. In the latter scenario the marked mosquitoes were drawn from a binomial distribution of sample size 100, which is the number of mosquitoes caught in each trap, with a probability that changed according to time and distance. Details on how the data were simulated, including the R-code we used, are given in the supplementary material.

Logistic Regression Model. The formula derived by Petersen and Lincoln (Lincoln 1930, Seber 1982) to estimate population size is the simplest estimation obtained from a single marking occasion in a closed population. It was the starting point to develop more

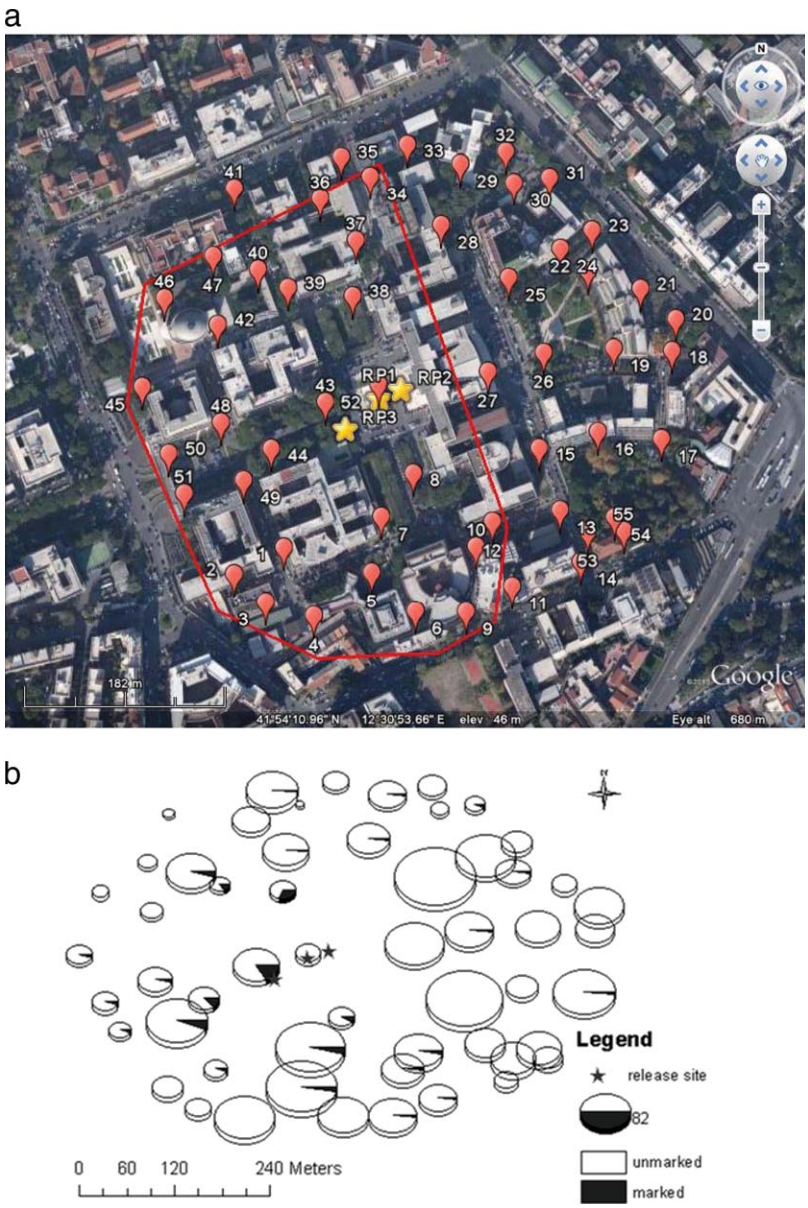

Fig. 1. (a) Distribution of the sticky traps and release points on the campus of the University of Rome, "Sapienza." The red markers indicate the 55 STs and the stars the release sites. The image is taken from Google Earth (source: "Sapienza," Rome; 41° 53′58.48″ N, 12° 31′13.59″ E. GOOGLE EARTH. July 2007; 25 April 2012). The red line indicates the area selected for the analysis. (b) Distribution of captured female *Ae. albopictus* in the 55 sticky traps on the campus of "Sapienza" University of Rome. The stars indicate the release sites, the black slices represent the number of marked captured mosquitoes, and the white slices represent the number of the unmarked. The size of the pies is proportional to the total number of mosquitoes trapped in the three MRR experiments. (Online figure in color.)

intricate and refined methods. We present how it was modified to include a mortality effect and a distance effect, to illustrate the similarities with the logistic regression model we developed, which is based on the same principle.

The principle behind the method is that a known number, *M*, of individuals is caught, marked and then returned to the environment. After a suitable interval to allow mixing with the population, a second population sample, of size *n*, is taken and the numbers of marked and unmarked individuals recorded. When the assumption of a closed population is met and the captured animals (marked and unmarked) constitute a simple random sample, the conditional distribution of the recaptured, *m*, given *M* and *n*, is hypergeometric (Chapman 1951, Seber 1982) (a list of the parameters

Table 1. Number of marked female *Ae. albopictus* recaptured (unmarked captured) per day in the three mark–release–recapture (MRR) experiments in the selected area (Fig. 2a) of the campus of "Sapienza" University in Rome (Italy)

	Days								
MRR	2	3	4	5	7	9	13	17	21
1	NS (NS)	NS (NS)	NS (NS)	15 (87)	NS (NS)	4 (87)	0 (88)	0 (109)	0 (127)
2	5 (32)	11 (37)	2 (14)	3 (17)	2 (38)	1 (17)	0 (36)	0 (36)	0 (45)
3	0 (41)	1 (56)	1 (42)	9 (33)	4 (35)	0 (30)	0 (56)	0 (75)	0 (33)

The number of marked mosquitoes released is 464 (MRR1), 566 (MRR2), 552 (MRR3).
NS = not sampled.

used is provided in the Table 2). Here, the hypergeometric distribution describes the random variable that counts, for n elements randomly drawn from the population of size N, the number of marked individuals recaptured from the subgroup of the marked individuals released. Applying the technique to the experiment on mosquitoes, the probability to get m marked mosquitoes is:

$$f(m|M,n) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \qquad [1]$$

Using a binomial approximation to the hypergeometric distribution, we have (Bailey 1951, 1952):

$$f(m|M,n) \approx \binom{n}{m}\left(\frac{M}{N}\right)^{m}\left(1-\frac{M}{N}\right)^{n-m} \qquad [2]$$

and the maximum likelihood estimate of $\hat{N}$ is the Petersen estimate (Seber 1982). The estimator is based on the assumption that the ratio of marked individuals to the total number in the second sample (both marked and unmarked) reflects the same ratio in the total population, so that:

$$\frac{m}{n} = \frac{M}{N} \qquad [3]$$

which provides the estimator:

$$\hat{N} = n\frac{M}{m} \qquad [4]$$

(Seber 1982, Pollock et al. 1990). A further extension of this formula includes a mortality factor:

$$\hat{N} = \phi^{t} n\frac{M}{m} \qquad [5]$$

where ϕ is the survival rate (actually more a "retention rate") and t is the period between the release and the recapture (Conway et al. 1974). This is known as Fisher–Ford's method (Dowdeswell et al. 1940, Fisher 1947). The mortality can be based on a laboratory population (MacLeod 1958) or obtained by independent assessment in the field. Because M marked adults, reared from eggs or pupa, are added to the natural population of which the size is being estimated, M needs to be subtracted from the estimate (Nelson 1978; Reisen et al. 1979, 1980):

$$\hat{N} = \left[\phi^{t} n\frac{M}{m}\right] - M \qquad [6]$$

Using the correction provided by Fisher and Ford for situations when the number recaptured is small (<20) (Bailey 1952), the formula becomes:

$$\hat{N} = \left[\frac{\phi^{t}(n+1)(M+1)}{m+1}\right] - 1 - M \qquad [7]$$

We applied these formulas (6 and 7) to the real and simulated data. The confidence intervals related to the estimates are calculated with the method of percentile bootstrap (Efron and Tibshirani 1993) based on 500 bootstrap replicates at 95% level. The survival rate was a value taken from Alto and Juliano (2001), who investigated the relationship between temperature and duration of life of *Ae. albopictus*, through laboratory experiments. They estimated a daily mortality rate at 24°C of 0.0114. This value provides an estimate made in conditions similar to the temperature experienced in Rome during the MRR experiments. However, because the other environmental conditions are different there may nevertheless be a difference between the mortality rate as estimated in the laboratory and that for the experiment is carried out in the field. Since, because of the low recaptures, it was not possible to estimate the mortality rate from the data, we also analyzed the case of a mortality rate of 0.15, which is considered more likely under field conditions.

We now describe our logistic regression approach, based on the same principle as the Fisher–Ford method. Although for the Fisher–Ford method we need to set the value of the mortality rate, the survival rate present in the logistic regression model does not need preset values from the literature, because it can be estimated simultaneously from the data. We expect

Table 2. Parameters involved in the formulas

Parameters	
N	Population size per day
M	Number of marked mosquitoes released
n	Number of mosquitoes captured
m	Number of marked mosquitoes recaptured
ϕ	Survival rate
t	Time between release and capture
S_m	Fraction of marked mosquitoes that can still be trapped at time t
λ	Parameter of the survival function
α	Intercept of the logistic regression, –log(N)
β_o	ln(λ)
p_i	Fraction that moves from an area to the one farther

to recapture more mosquitoes in the first days, because after that they may have spread to a bigger area and some of them will have died. To control for this, the model includes a time effect, which means the effect of mortality and emigration. Furthermore, the logistic regression allows the inclusion of the distance between release point and STs, which may be important, because of the limited flight range of *Ae. albopictus.* Usually in an MRR experiment with insects, traps are placed at different distances from the release point and therefore each trap is expected to have a different recapture probability. This can be controlled for by introducing the distance effect in the model.

The starting point for the logistic regression model is the same as that of Fisher–Ford's method: the fraction of recaptured mosquitoes. If we assume a zero death rate for the logistic regression model, we can see the similarity between the methods. The model aims to estimate the number of mosquitoes in the period of interest and first we assume this quantity is constant and there is no mortality in the entire period. Then, the population fraction of marked mosquitoes is:

$$\pi = \frac{M}{M + N} \qquad [8]$$

where the denominator considers that we introduced M mosquitoes in the total population (cf. equation 7. The ratio of the probability to capture a marked mosquito to the probability to capture an unmarked mosquito is:

$$\frac{\pi}{1 - \pi} = \frac{M}{N} \qquad [9]$$

π is estimated as the ratio of marked mosquitoes recaptured to the total number of mosquitoes captured ($\pi = m/n$). We can then express the above formula as:

$$\frac{m}{n - m} = \frac{M}{N} \qquad [10]$$

and it follows that:

$$N = \frac{M(n - m)}{m} \qquad [11]$$

which is similar to the Lincoln index.

In the following, we add first survival and then distance to the model.

We assume that N is the number of mosquitoes in the study area per day and that the number of marked mosquitoes released at time point $t = 0$ is M. The population of mosquitoes can increase through emergences or decrease through deaths, whereas the population of marked mosquitoes can only decrease. At time t, $MS_m(t)$ is the number of marked mosquitoes that can still be trapped at that time. For the wild population, we are assuming a closed population, which means that when a mosquito is lost, it is replaced by another one. That is, the number of emergences equals the number of deaths. Then the number of mosquitoes N will be constant over the time. We are also assuming a stable age distribution. Considering these assumptions, the population fraction of marked mosquitoes at time t is:

$$\pi(t) = \frac{MS_m(t)}{N + MS_m(t)} \qquad [12]$$

and the ratio of the probability to capture a marked mosquito to the probability to capture an unmarked mosquito is:

$$\frac{\pi(t)}{1 - \pi(t)} = \frac{MS_m(t)}{N} \qquad [13]$$

A sample of n mosquitoes at time t is taken and each is checked for marks. The number of marked mosquitoes in the sample has a binomial distribution with sample size n and population fraction π. From this, the logistic regression model is:

$$\log\left(\frac{\pi(t)}{1 - \pi(t)}\right) = -\log(N) + \log[S_m(t)] + \log(M) \qquad [14]$$

Assuming an exponential distribution for the survival function, i.e. $S_m = e^{-\lambda t}$ (Macdonald 1952), and a log-linear model for the rate, i.e. $\lambda = e^{\beta o}$ the logistic regression model with time effect is:

$$\log\left(\frac{\pi(t)}{1 - \pi(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M) \qquad [15]$$

The survival function in the model has a broader meaning than usual: it includes both a mortality factor and dispersal to outside the capture area. N can be estimated from the intercept of the model, i.e., $\alpha = -log(N)$:

$$\hat{N} = \exp(-\hat{\alpha}) \qquad [16]$$

and the approximate 95% confidence interval is given by:

$$\exp[-\hat{\alpha} - zSE(\hat{\alpha})] < N < \exp[-\hat{\alpha} + zSE(\hat{\alpha})] \qquad [17]$$

where z is a standard normal distribution.

Besides the time effect, we incorporate a distance effect, to check whether it plays an important role in the sense that a model with distance included provides markedly different estimates, or improved estimate-confidence. The distance can be included in the model as a continuous variable, to consider the individual trap-distances, or divided in zones. Given the spatial distribution of the recaptures of our data, we decided to consider three different pooled distance areas from the release site. We consider that a certain fraction p of marked mosquitoes moves from an area to the adjacent one, starting from the release point $p_iMS(t)$. If we consider three distances, we expect to find the mosquito population N uniformly distributed over the areas ($N/3$ in each area) and the fraction p_2 moving to $area_2$, the fraction p_3 moving to $area_3$, the fraction $p_1 = 1\text{-}p_2\text{-}p_3$ staying in $area_1$. Then $p_1MS(t)$ mosquitoes will be in $area_1$, $p_2MS(t)$ mosquitoes will be in $area_2$,

Table 3. Estimates of female *Ae. albopictus* population size in the three mark–release–recapture (MRR) experiments, day 1–9, selected part of the campus of "Sapienza" University in Rome (Italy), maximum distance 210 m

MRR	M	Method	Density per ha	N per day	95% CI for N	Width CI	λ	SE(β_0)	AIC
1	464	Model 1[a]	49	516	392–679	287	0.330	0.069	2,549.53
		Model 2[b]	23	244	37–1611	1574	0.307	0.490	61.07
		Fisher—Ford[c]	340	3603	2,245–7,219	4974	–	–	–
		Fisher—Ford[d]	55	580	204–1,505	1301	–	–	–
2	566	Model 1[a]	135	1,435	1,314–1,567	253	0.223	0.047	11,328.06
		Model 2[b]	63	673	232–1,953	1721	0.254	0.680	95.08
		Fisher—Ford[c]	294	3,115	1,932–5,451	3,519	–	–	–
		Fisher—Ford[d]	36	379	103–919	816	–	–	–
3	552	Model 1[a]	823	8,723	8,369–9,093	724	1.35E-07	68.792	13,224.78
		Model 2[b]	625	6,625	2,880–15,238	12,358	3.76E-06	174.820	89.53
		Fisher—Ford[c]	692	7,334	4,661–13,282	8,621	–	–	–
		Fisher-Ford[d]	139	1,472	827–2,988	2,161	–	–	–

Density is expressed as the number of female per hectare, N is the estimate of the population size per day, λ is the parameter of the survival function estimated with the logistic regression. The daily survival rate used for the method of Fisher—Ford is 0.9886 (Alto and Juliano 2001) and a more likely value for the field of 0.85; the confidence interval is based on 500 bootstrap replicates with the method of percentile bootstrap at 95% level.

M = marked mosquitoes released.

[a] Equation 15: $\log\left(\frac{\pi(t)}{1-\pi(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M)$.

[b] Equation 20: $\log\left(\frac{\pi_i(t)}{1-\pi_i(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M) + \log(3p_i)$.

[c] Fisher and Ford with daily survival rate 0.9886 (Alto and Juliano 2001) and correction for low recapture.

[d] Fisher and Ford with daily survival rate 0.85 and correction for small recaptures.

AIC = Akaike Information Criterion.

$p_3MS(t)$ mosquitoes will be in $area_3$. The population fraction of marked mosquitoes at time t in area i is then:

$$\pi_i(t) = \frac{p_iMS_m(t)}{\frac{1}{3}N + p_iMS_m(t)} = \frac{3p_iMS_m(t)}{N + 3p_iMS_m(t)} \qquad [18]$$

and the ratio of the probability to capture a marked mosquito to the probability to capture an unmarked mosquito becomes:

$$\frac{\pi_i}{1-\pi_i} = \frac{3p_iMS_m(t)}{N} \qquad [19]$$

The logistic regression model with time and distance effect is:

$$\log\left(\frac{\pi_i(t)}{1-\pi_i(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M) + \log(3p_i) \qquad [20]$$

In the application of the model to the real data, the distance was included in the model as a factor and the total experimental area around the release point was divided in three regions of equal surface area (distance smaller than 121 m, between 121 and 171.5 m, between 171.5 and 210 m), with an exception for MRR2 in which we recognized only two areas (distance smaller than 148.5 m, between 148.5 and 210 m), because no marked mosquitoes were found in the third area. Also for the simulated data, the distance was included in the model as a factor and the area was divided in three regions with the same surface area.

The parameters are estimated using the maximum likelihood approach with the package bbmle in R (Bolker and RDC Team 2012).

The parameters involved in the models are summarized in the Table 2.

Results

We applied to the data the logistic regression model with only the survival function (equation 15), the logistic regression model with survival function and distance effect (equation 20), and the Fisher–Ford method (equation 7). This is used as a possible comparison with one of the already existing methods; the choice is because of the fact that it is based on the same principle as our method. The same analysis also was applied to simulated data to show how the models perform in different scenarios, when the number of marked mosquitoes released, the population size and the distance effect change.

Analysis of *Ae. albopictus* data. Applying the model with the survival function, the estimated population size per day was 516 (95% CI 392–679) in MRR1, 1,435 (95% CI 1,314–1,567) in MRR2, and 8,723 (95% CI 8,369–9,093) in MMR3. The parameter of the survival function, which describes the amount of mosquitoes lost per time period, was estimated as 0.330 (SE 0.069) in MRR1, 0.223 (SE 0.047) in MRR2, 1.35E-07 (SE 68.792) in MRR3 (Table 3).

Applying the model with the survival function and the distance effect, the estimated population size per day was 244 (95% CI 37–1,611) in MRR1, 673 (95% CI 232–1,953) in MRR2, and 6,625 (95% CI 2,880–15,238) in MMR3. The parameter of the survival function was

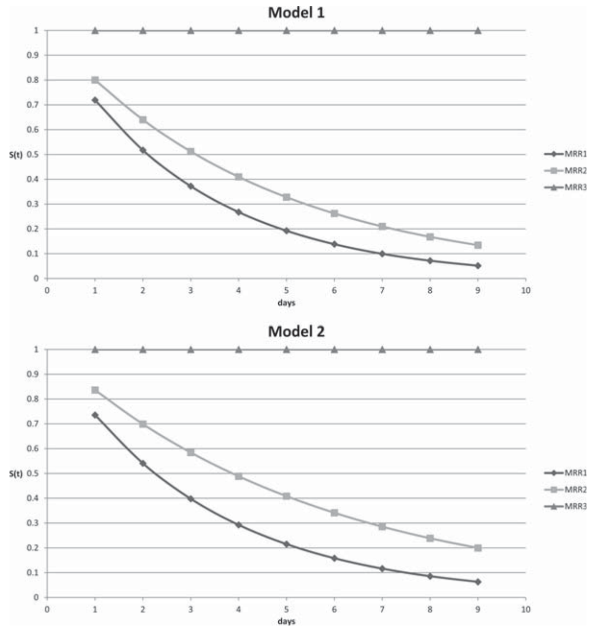

Fig. 2. Loss function in models 1 and 2 for each experiment.

estimated as 0.307 (SE 0.490) in MRR1, 0.254 (SE 0.680) in MRR2, 3.76E-06 (SE 174.820) in MRR3 (Table 3).

It is important to note that the estimates of λs for MRR3 are very small but are also based on very limited data with very low recapture at the beginning of the experiment. This limits the capability of the model to estimate the parameters, as we can see from the huge standard error (for both models).

The loss survival function for each experiment and model are shown in Fig. 2.

The population size estimated with Fisher–Ford method is very sensitive for the parameter that we choose for the mortality rate. In MRR1 is 3,603 with $\phi = 0.9886$ (95% CI 2,245–7,219) and 580 with $\phi = 0.85$ (95% CI 204–1,505); in MRR2 is 3115 with $\phi = 0.9886$ (95% CI 1,932–5,451) and 379 with $\phi = 0.85$ (95% CI 103–919); in MRR3 is 7,334 with $\phi = 0.9886$ (95% CI 4,661–13,282), 1,472 with $\phi = 0.85$ (95% CI 827–2,988).

Analysis on the Simulated Data. In all the simulated scenarios and with both logistic regression models (equation 15 and 20), we obtained an estimate of N very close to the predefined size of our simulated population. We present the median value of the estimate (Table 4), because the distribution of the estimates is skewed (Figure S1 in the supplemental material). The performance of both models is similar when the marked mosquitoes are equally distributed over the areas. In scenarios 1, 3, and 5 the AICs (we report the median value because the AIC has also a skewed distribution) differ <2 U. The model with the distance performs better in the scenarios 2, 4, and 6, when there is a relationship between the distance from the release site and the number of marked mosquitoes caught. In all the scenarios, the precision of the estimates, indicated by the interquartile range, does not substantially change.

In all the simulated scenarios, the true value of the population size is never present in the interquartile range of the estimates produced by Fisher–Ford method. The estimates are more precise, but they lack accuracy.

Discussion

We applied a logistic regression model to estimate the population size of *Ae. albopictus* on the campus of "Sapienza" University in Rome, using data from mark–release–recapture experiments (Marini et al. 2010). We also tested the performances of the model under different simulated scenarios.

The logistic regression model we present is built to accommodate most mosquito MRR protocols where population size is estimated from a single release, captures are carried out on several days after the release and traps are distributed in an area at different distances from the release site.

Besides providing an estimate of the population size, the model allowed us to estimate a combined survival and dispersal (or "retention") factor from the data and to consider the time needed to reach traps situated farther from the release point. This means that it is not necessary to provide a preset value for survival taken from the literature (laboratory experiments or other field experiments). Such values will usually not be a good match with the situation in one's own experiment. For example, the value we chose for the Fisher–Ford model, based on the only laboratory experimental evidence we could find, was markedly different from the values we estimated directly from the data of the MMR experiments. In our design, we do not have a preset value, but rather a preset *distribution* for the survival. In our analyses we chose an exponential

Table 4. Application of the model on simulated data: $N = 21{,}000$; survival rate = 0.95; number of simulation = 1,000; except for scenario 1 where the number of simulations is 516

Scenario	M in area 1	M in area 2	M in area 3	Model	Median N	First Q	Third Q	IQR	Median AIC
1	50	50	50	1	20,095	15,401	26,854	11,453	55.46
				2	20,080	15,370	26,820	11,450	57.16
				FF[a]	15,914	14,129	19,622	5,493	–
2	105	30	15	1	20,597	14,890	26,874	11,984	58.26
				2	20,500	14,780	26,820	12,040	50.61
				FF[a]	15,914	13,942	19,622	5,680	–
3	170	170	170	1	20,439	17,022	24,879	7,857	91.75
				2	20,430	16,980	24,830	7,850	93.67
				FF	16,507	14,987	17,955	2,968	–
4	357	102	51	1	20,675	17,075	24,411	7,336	113.1
				2	20,360	16,670	23,990	7,320	84.08
				FF	16,507	14,987	17,955	2,968	–
5	335	335	335	1	19,903	17,370	22,704	5,334	110.97
				2	19,900	17,340	22,690	5,350	113.26
				FF	15,927	14,830	16,997	2,167	–
6	704	201	100	1	19,990	17,605	22,920	5,315	162.4
				2	19,410	17,060	22,300	5,240	102.68
				FF	15,761	14,830	17,188	2,358	–

M = marked mosquitoes released.

Model 1: $\log\left(\frac{\pi(t)}{1-\pi(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M)$.

Model 2: $\log\left(\frac{\pi_i(t)}{1-\pi_i(t)}\right) = -\log(N) - e^{\beta_0}t + \log(M) + \log(3p_i)$.

First Q and third Q are the first and third quartile.

IQR is the interquartile range.

AIC = Akaike Information Criterion.

[a] FF: Fisher—Ford method with correction for low recapture (equation 7).

distribution, but this can be changed to another distribution, should more information about the survival function of mosquitoes be available for the field situation one is studying.

Taking into account the distance between release site and trap may be important in an experimental setup. We would expect to find more mosquitoes in the traps near the release point in the first days and then a spread to a wider area. For traps at the borders of the natural flight range, one would expect that too few marked individuals are caught to satisfy the (implicit) assumption in our approach that the naturally occurring (unmarked) population and the marked population are well-mixed. If we do not consider the distance effect, we could base our estimate on a too low recapture rate, because of the difficulty the mosquitoes had to reach the traps and because of the chance effects deviating from a well-mixed situation.

We evaluated whether the distance correction really improved the estimates, comparing a model with only time effect (model 1) and a model with time effect and distance effect (model 2) for each experiment. We observed that the introduction of the distance effect improved the model substantially, as shown by the AIC. Second, the width of the confidence intervals shows how the estimates became less precise, with a substantially larger variability. To draw more robust conclusions about the performance of the models, we examined the simulated data, because the real data did not have many observations. The estimates related to the simulated scenarios, appeared very similar for models 1 and 2. They were always close to the real value of 21,000 mosquitoes. The interquartile range assumed similar values in both models. The AIC shows a small improvement in the scenarios in which the number of mosquitoes decreases when the distance increases. The difference was negligible when the marked mosquitoes were distributed equally among the areas.

Examining the number of mosquitoes estimated by the models from the real data, one may be surprised by the relatively low estimates, considering what is known about the abundance of *Ae. albopictus* in Rome (Severini et al. 2008), and Sapienza campus area, in particular (B. C., A.D.T., F. M., personal communication). It is, however, important to remark that the estimates are based on the number of marked mosquitoes recaptured, which also depends on the characteristics of the method used to trap the mosquitoes in the experimental setup. Because the sticky trap attracts mainly gravid female mosquitoes, and all marked mosquitoes were released at the fed stage and presumably recaptured during oviposition (see Marini et al. 2010), it may be more appropriate to consider that our estimates refer mainly to this fraction of the total population. This could then explain why in the third experiment, carried out in October, we found a larger number of mosquitoes in the traps: in fact, at the end of the reproductive season it may be expected that no freshly emerged mosquito are produced and that, a higher fraction of the surviving female mosquitoes is looking for a place to lay eggs.

The logistic regression model proposed is able to include specific features of MRR experiments carried out with mosquitoes. It can be applied whenever there is interest in monitoring the population size of mosquitoes or other insects with mark–recapture experiments. To improve and further validate the model, additional experimental data from the field are needed. This would help to understand which variables are most important and the kind of relationship they have with the outcome. One could then also use the model in combination with simulated data to explore experimental set-ups, comparing different numbers of traps, at different distances and spatial distributions from a release point, and the influence of releasing fewer or more marked individuals. We have used simulated data to at least provide some validity of the approach, showing that the model is able to estimate preset population size with sufficient accuracy to serve as a basis to estimate vector to host ratio's accurately enough for (local) R_0 estimates.

Acknowledgments

This study was funded by EU grants FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext040 (http://www.edenext.eu).

References Cited

Alto, B. W., and S. A. Juliano. 2001. Temperature effects on the dynamics of *Aedes albopictus* (Diptera: Culicidae) populations in the laboratory. J. Med. Entomol. 38: 548–556.

Bailey, N.T.J. 1951. On estimating the size of mobile populations from recapture data. Biometrika 38: 293–306.

Bailey, N.T.J. 1952. Improvements in the interpretation of recapture data. J. Anim. Ecol. 21: 120–127.

Begon, M. 1979. Investigating animal abundance: capture-recapture for biologists. Arnold, Edward, London, United Kingdom.

Bolker, B., and RDC Team. 2012. bbmle: tools for general maximum likelihood estimation, http://cran.r-project.org/web/packages/bbmle/index.html.

Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. University of California Publications in Statistics, Berkeley, CA.

Conway, G. R., M. Trpis, and G.A.H. McClelland. 1974. Population parameters of the mosquito *Aedes aegypti* (L.) estimated by mark-release-recapture in a suburban habitat in Tanzania. J. Anim. Ecol. 43: 289–304.

Dowdeswell, W. H., R. A. Fisher, and E. B. Ford. 1940. The quantitative study of populations in the lepidoptera I. *Polyommatus icarus* rott. Ann. Hum. Genet. 10: 123–136.

Efford, M. G. 2009. Density estimation by spatially explicit capture–recapture: likelihood-based methods. Springer, New York.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, London, United Kingdom.

[ESRI] Environmental Systems Resource Institute. 2009. ArcGIS 9.2. Redlands, CA.

Facchinelli, L., L. Valerio, M. Pombi, P. Reiter, C. Costantini, and A. Della Torre. 2007. Development of a novel sticky trap for container - Breeding mosquitoes and evaluation of its sampling properties to monitor urban populations of *Aedes albopictus.* Med. Vet. Entomol. 21: 183–195.

Fisher, R. A. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. Heredity 1: 143–174.

Google Earth Beta. 2010. Google Earth Beta, version 5.2.1.

Gratz, N. G. 2004. Critical review of the vector status of *Aedes albopictus.* Med. Vet. Entomol. 18: 215–227.

Hartemink, N. A., B. V. Purse, R. Meiswinkel, H. E. Brown, A. De Koeijer, A.R.W. Elbers, G. J. Boender, D. J. Rogers, and J.A.P. Heesterbeek. 2009. Mapping the basic reproduction number (R0) for vector-borne diseases: a case study on bluetongue virus. Epidemics 1: 153–161.

Hartemink, N. A., S. O. Vanwambeke, H. Heesterbeek, D. Rogers, D. Morley, B. Pesson, C. Davies, S. Mahamdallie, and P. Ready. 2011. Integrated mapping of establishment risk for emerging vector-borne infections: a case study of canine leishmaniasis in southwest France. PLoS ONE 6: e20817.

Jackson, C.H.N. 1937. Some new methods in the study of *Glosmina morsitans.* Proc. Zool. Soc. 106: 811–896.

Jackson, C.H.N. 1939. The analysis of an animal population. J. Anim. Ecol. 8: 238–246.

Jackson, C.H.N. 1940. The analysis of a tsetse fly population. Ann. Eugen. 10: 332–369.

Krebs, C. J. 1999. Ecological methodology. H and R Publishers, New York.

Le Cren, E. D. 1965. A note on the history of mark-recapture population estimates. J. Anim. Ecol. 34: 453–454.

Lincoln, F. C. 1930. Calculating waterfowl abundance on the basis of banding returns. Circ. U.S. Dep. Agric. 118.

Macdonald, G. 1952. The analysis of the sporozoite rate. Trop. Dis. Bull. 49: 569–586.

MacLeod, J. 1958. The estimation of numbers of mobile insects from low - Incidence recapture data. Ecol. Entomol. 110: 363–392.

Marini, F., B. Caputo, M. Pombi, G. Tarsitani, and A. Della Torre. 2010. Study of *Aedes albopictus* dispersal in Rome, Italy, using sticky traps in mark-release-recapture experiments. Med. Vet. Entomol. 24: 361–368.

Nelson, R. L. 1978. Estimates of survival, population size, and emergence of *Culex tarsalis* at an isolated site. Ann. Entomol. Soc. Am. 71: 801–808.

Pollock, K. H. 1980. Capture-recapture models: a review of current methods, assumptions and experimental design. North Carolina State University, Raleigh, NC.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. Wildl. Monogr. 107: 1–97.

Reisen W. K., F. Mahmood, and T. Parveen. 1979. *Anopheles subpictus* Grassi: observations on survivorship and population size using mark-release-recapture and dissection methods. Res. Popul. Ecol. 21: 12–29.

Reisen, W. K., F. Mahmood, and T. Parveen. 1980. *Anopheles culicifacies* Giles: a release-recapture experiment with cohorts of known age with implications for malaria epidemiology and genetical control in Pakistan. Trans. R. Soc. Trop. Med. Hyg. 74: 307–317.

Schwarz, C. J., and G.A.F. Seber. 1999. Estimating animal abundance: review III. Stat. Sci. 14: 427–456.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. Griffin, London, United Kingdom.

Seber, G.A.F. 1986. A review of estimating animal abundance. Biometrics 42: 267–292.

Severini, F., M. Di Luca, L. Toma, and R. Romi. 2008. *Aedes albopictus* in Rome: results and perspectives after 10 years of monitoring. Parassitologia 50: 121–123.

White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, Los Alamos, NM.

White, G. C. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46: 120–139.

Received 8 June 2012; accepted 25 January 2013.