Theoretical Population Biology 78 (2010) 71–76

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Fitting parameters of stochastic birth–death models to metapopulation data

Heinrich zu Dohna [a], Mario Pineda-Krch [b,*]

[a] Center for Animal Disease Modelling, Department of Veterinary Medicine, University of California Davis, One Shields Avenue, Davis, CA 95618, USA

[b] Centre for Mathematical Biology, Department of Mathematical & Statistical Sciences, 632 Central Academic Building, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

ARTICLE INFO

Article history:
Received 11 August 2009
Available online 17 June 2010

Keywords:
Metapopulation
Birth–death processes
SIS model
Logistic growth
Maximum likelihood

ABSTRACT

Populations that are structured into small local patches are a common feature of ecological and epidemiological systems. Models describing this structure are often referred to as metapopulation models in ecology or household models in epidemiology. Small local populations are subject to demographic stochasticity. Theoretical studies of household disease models without resistant stages (SIS models) have shown that local stochasticity can be ignored for between patch disease transmission if the number of connected patches is large. In that case the distribution of the number of infected individuals per household reaches a stationary distribution described by a birth–death process with a constant immigration term. Here we show how this result, in conjunction with the balancing condition for birth–death processes, provides a framework to estimate demographic parameters from a frequency distribution of local population sizes. The parameter estimation framework is applicable to estimate parameters of disease transmission models as well as metapopulation models.

1. Introduction

Metapopulation models in ecology and household models in epidemiology describe populations whose individuals are distributed over many loosely connected small local populations. An important approach in metapopulation biology is the incidence function model which estimates extinction and colonization probabilities of local populations from one or multiple snapshots of patch occupancy (Hanski, 1994; Moilanen, 1999). The incidence function model and similar approaches distinguish between empty and occupied patches but do not take data on local population sizes into account.

Household models in epidemiology have been used to estimate parameters of within and between household disease transmission from a frequency distribution of infected individuals per household for models that contain susceptible, infected and resistant individuals (SIR models) (Longini and Koopman, 1982; Longini et al., 1982). Parameter estimation approaches have been lacking so far for household models of diseases without a resistant stage (SIS models). Theoretical studies of SIS household models have shown that at equilibrium within household dynamics can be treated as stochastic birth–death process with constant immigration term when the number of households is large (Ball, 1999; Ghoshal et al., 2004).

Here we show how approximation results from household SIS models can be combined with analytical results for stochastic birth–death processes to estimate parameters for within and between household disease transmission from a frequency distribution of infected individuals per household. The same approach can be used to estimate immigration, birth and death rates from a frequency distribution of local population sizes in metapopulation data.

2. Methods

2.1. Problem description

Metapopulations with small local populations undergo demographic stochasticity at the local scale. If the immigration rate to any local population depends on a large number of other local populations, the local demographic stochasticity does not translate into stochastic fluctuations of the immigration rate to each patch (however, demographic stochasticity in the number of immigrants per time still occurs). In that case the distribution of local population sizes at equilibrium can be well approximated by a stationary distribution of a stochastic immigration–birth–death process whose immigration term depends on the mean local population size (Ghoshal et al., 2004). Such a stationary distribution whose mean equals the mean population size in its immigration term has been called a 'self-consistent field' and has been shown to approximate simulations of a household SIS model well for a surprisingly wide range of conditions (Ghoshal et al., 2004).

* Corresponding author.
E-mail addresses: hzudohna@ucdavis.edu (H. zu Dohna), mpineda@math.ualberta.ca (M. Pineda-Krch).

0040-5809/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2010.06.004

In this study we explore how the self-consistent field approach can be utilized to estimate parameters from counts of populations that have reached such a stationary distribution. We first review a key result for stationary distributions of stochastic birth–death processes and then show how this result can be used to estimate parameters from counts of individuals in local populations. Finally we apply the parameter estimation method to fit parameters of an SIS household model and a metapopulation model to simulated data and field data.

2.2. A general birth–death model

Let I denote the number of individuals in a population that undergoes a stochastic birth–death process. The birth–death process could describe population dynamics or disease infection dynamics, with I either denoting the number of individuals living in a population or the number of infected individuals, respectively. In the text below, the term birth will describe the addition of an individual and death the removal. These two events correspond to actual birth and death if the model describes population dynamics. If the model describes disease transmission, birth means the addition of a new infected individual through infection and death the removal of an infected individual through clearance of the disease, disease-induced mortality, quarantine or culling.

During the infinitesimal time interval $(t, t + \mathrm{d}t)$ a population with I individuals undergoes the two possible transitions to $I + 1$ or $I - 1$ with the following probabilities

$$P(I \to I + 1) = \lambda_I \cdot \mathrm{d}t \tag{1a}$$

$$P(I \to I - 1) = \mu_I \cdot \mathrm{d}t. \tag{1b}$$

The process described by Eqs. (1a) and (1b) reaches a stationary distribution characterized by the detailed balancing condition if the following conditions hold for birth rate λ_I and death rate μ_I (Gillespie, 1992) below a maximum population size m:

$$\lambda_I \text{ and } \mu_I \text{ are both } \begin{cases} > 0 & \text{if } I \in [1, m] \\ = 0 & \text{if } I \notin [1, m]. \end{cases} \tag{2}$$

In biological terms, the conditions above mean that the population has a positive birth rate as long as the population size ranges between zero and $m - 1$ and a positive death rate for the population ranging between 1 and m. Hence, for the detailed balancing condition to hold there must be a positive immigration rate (or infection rate for disease models) from the outside, leading to a positive birth rate when no individual is present, and a maximum population size m that the population can never exceed. In disease transmission models with fixed population size m is simply the number of individuals that can become infected.

The detailed balancing condition allows a recursive calculation of the equilibrium probabilities $P(I)$ for the stationary distribution of I (Gillespie, 1992; Nåsell, 2001):

$$P(I) = \begin{cases} K & \text{if } I = 0 \\ \dfrac{\lambda_{I-1}}{\mu_I} P(I-1) & \text{if } I \in [1, m] \end{cases} \tag{3}$$

where $K = (1 + \sum_{k=1}^{m} \prod_{j=1}^{k} \frac{\lambda_{j-1}}{\mu_j})^{-1}$.

2.3. Parameter estimation

The parameters of Eq. (3) can be estimated from multiple independent observations of separate populations. Let x_{im} denote the number of populations that have maximum population size m and observed number of individuals i. The log-likelihood of the observed population sizes x_{im} is given by

$$L = \sum_m \sum_{i=1}^{m} x_{im} \ln[P_m(i)]. \tag{4}$$

The subscript m indicates the dependence of the probabilities $P_m(i)$ on the maximum population size, which might be known or could be a parameter to be estimated. The parameters $\boldsymbol{\theta}$ that maximize the likelihood are given by the solutions to the equations

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_m \sum_{i=1}^{m} x_{im} \frac{\frac{\partial}{\partial \boldsymbol{\theta}}[P_m(i)]}{P_m(i)}.$$

The solutions can be found through Fisher scoring, an iterative procedure that replaces at each iteration step t the current vector of parameter estimates $\boldsymbol{\theta}_t$ with the improved estimates $\boldsymbol{\theta}_{t+1}$ (Rao, 1953)

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \mathbf{H}^{-1} \frac{\partial L}{\partial \boldsymbol{\theta}} \tag{5}$$

where $\mathbf{H}$ denotes the Fisher Information Matrix, whose elements (k, l) are given by $E[\frac{\partial L}{\partial \theta_k} \cdot \frac{\partial L}{\partial \theta_l}]$ (Rao, 1953). For the likelihood shown in Eq. (4) the elements of the Fisher Information Matrix become (Rao, 1953)

$$E\left[\frac{\partial L}{\partial \theta_k}\frac{\partial L}{\partial \theta_l}\right] = N \cdot \sum_m \sum_{i=1}^{m} \frac{\frac{\partial}{\partial \theta_k}[P_m(i)] \cdot \frac{\partial}{\partial \theta_l}[P_m(i)]}{P_m(i)}. \tag{6}$$

Hence solving Eq. (5) only involves calculating $P(i)$ and $\frac{\partial}{\partial \theta_k}[P(i)]$ for all parameters. The probabilities and their derivatives can be obtained efficiently by differentiating the recursive relationship of Eq. (3)

$$\begin{aligned} \frac{\partial}{\partial \theta_k}[P_m(i)] &= \frac{\partial}{\partial \theta_k}\left[\frac{\lambda_{i-1}}{\mu_i}\right] P_m(i-1) + \frac{\lambda_{i-1}}{\mu_i}\frac{\partial}{\partial \theta_k}[P_m(i-1)] \\ \frac{\partial}{\partial \theta_k}[P_m(0)] &= \frac{\partial}{\partial \theta_k}[K]. \end{aligned} \tag{7}$$

The general procedure to obtain maximum likelihood parameter estimates can therefore be summarized as follows: (i) start with an initial estimate for the parameter values; (ii) obtain $\frac{\partial}{\partial \theta}\left[\frac{\lambda_{i-1}}{\mu_i}\right]$ and $\left[\frac{\lambda_{i-1}}{\mu_i}\right]$ for all parameters θ and population sizes i; (iii) substitute results from step (ii) into recursions given by Eqs. (3) and (7) to calculate $P(i)$ and $\frac{\partial}{\partial \theta}[P(i)]$; (iv) substitute results from step (iii) into Eqs. (6) and (5) to obtain new parameter estimates and start with step (ii) again. These steps are repeated until two subsequent parameter estimates differ by less than a predetermined threshold. If n, the number of observations is large, the maximum likelihood estimates are approximately normally distributed with variance–covariance matrix of $1/n\mathbf{H}^{-1}$ (Rice, 1995). The above estimation procedure therefore yields not only the maximum likelihood parameter estimates but also confidence intervals around them.

To complete the parameter estimation procedure it is necessary to derive a method to obtain initial parameter estimates. This step is important since the iterative maximization of the likelihood function could converge to a local optimum if the initial parameter estimates are too far from the true values. Initial parameter estimates can be obtained using the recursive relationship of Eq. (3)

$$\frac{P_m(i)}{P_m(i-1)} = \left[\frac{\lambda_{i-1}}{\mu_i}\right]. \tag{8}$$

The ratio of count probabilities in the above equation can be replaced by the observed ratios x_{im}/x_{i-1m} to construct a set of equations that can be solved for the parameter values. Details of this procedure will be demonstrated in the examples below.

2.4. Example 1: the household SIS model and a metapopulation with logistic growth

Consider a collection of local populations of size m and disease transmission within and between populations. Each individual

becomes infectious immediately after infection, clears infection with rate γ (i.e. the average infectious period equals γ^{-1}) and becomes susceptible immediately after clearance of infection. This model is commonly referred to as SIS model. Depending on the parameter values, a disease described by this model either dies out or reaches a stationary distribution (Ball, 1999; Ghoshal et al., 2004). The probability intensities for birth (new infection), λ_I, and death (i.e. clearance of infection), μ_I, in a local population with I infected individuals are given by

$$\lambda_I = \left(\beta_b \cdot \bar{I} + \beta_w \cdot I\right) \cdot (m - I) \qquad (9)$$
$$\mu_I = \gamma \cdot I.$$

$\bar{I}$ denotes the mean number of infected individuals per sub-population. Treating $\bar{I}$ as a constant approximates the stochastic dynamics well if the number of populations is sufficiently large (Ball, 1999; Ghoshal et al., 2004). This model contains three parameters, β_b, the rate of between population disease transmission, β_w, the per pair rate of within-population disease transmission and γ, the rate of infection clearance. The birth–death rate ratios are given by

$$\frac{\lambda_{I-1}}{\mu_I} = \frac{\left(\beta_b \cdot \bar{I} + \beta_w \cdot (I-1)\right) \cdot (m - I)}{\gamma \cdot I}. \qquad (10)$$

The probabilities of the counts of infected number of individuals $P(I)$ are functions of birth–death rate ratios (see Eq. (3)). These ratios, and therefore the count probabilities, are not uniquely determined by the three parameters. For example, the same ratios resulting from three parameter values β_b, β_w, and γ, can be achieved by replacing these three values with β_b/γ, β_w/γ, and 1. Hence only values for the ratios β_b/γ and β_w/γ can be estimated but not for γ. That means in practice one can only estimate the relative values for between and within sub-population transmission from count data of infected individuals.

The derivatives of the birth–death ratios with respect to the parameters β_b and β_w are

$$\frac{\partial}{\partial \beta_b}\left[\frac{\lambda_{I-1}}{\mu_I}\right] = \frac{(m - I + 1) \cdot \bar{I}}{\gamma \cdot I} \qquad (11)$$

and

$$\frac{\partial}{\partial \beta_w}\left[\frac{\lambda_{I-1}}{\mu_I}\right] = \frac{(I-1) \cdot (m - I + 1)}{\gamma \cdot I}. \qquad (12)$$

Initial parameter estimates can be obtained from the observed ratios of consecutive count numbers. Given the number of infected individuals in a population, i, the following relationship follows from Eq. (3)

$$\frac{P(i+1)}{P(i)} \cdot \frac{i+1}{n-i} = \frac{\beta_b \cdot \bar{I}}{\gamma} + \frac{\beta_w}{\gamma} \cdot i. \qquad (13)$$

Denoting the ratio on the left hand side of the equation above by ζ_i, one can construct two equations for two different ratios ζ_i and ζ_j and solve them for the compound parameters β_b/γ and β_w/γ:

$$\frac{\beta_w}{\gamma} = \frac{\zeta_i - \zeta_j}{i - j} \qquad (14)$$

and

$$\frac{\beta_b}{\gamma} = \left(\zeta_i - i \cdot \frac{\zeta_i - \zeta_j}{i - j}\right) \Big/ \bar{I}. \qquad (15)$$

Replacing within each ratio ζ_i the probability ratio $P(i+1)/P(i)$ by the ratio of observed counts x_{im}/x_{i-1m} produces initial estimates of the parameters β_b/γ and β_w/γ. Since this procedure requires the denominator (x_{i-1m}) to be greater than zero it can only be used if there are at least three non-zero counts of consecutive numbers of infected individuals. In practice there might be more non-zero counts, such that several estimates for β_b/γ and β_w/γ can be obtained from multiple pairs of ratios ζ_i and ζ_j. In that case one can use the average of those values as starting parameter. Once initial parameter estimates have been obtained the maximum likelihood estimates can be calculated according to the iterative procedure outlined above.

The SIS model described above is mathematically equivalent to a metapopulation model with local stochastic logistic growth and a per capita birth and immigration rate that decreases linearly with local population density. In practice, however, fitting a logistic growth model is different from fitting an SIS model since the maximum population size m has to be estimated for a logistic growth model. Note that m is different from the carrying capacity. The carrying capacity is the population size I at which the birth rate equals the death rate and can be found by solving $\lambda_I = \mu_I$ for I. The derivative of the birth–death ratios with respect to the parameter m is

$$\frac{\partial}{\partial m}\left[\frac{\lambda_{I-1}}{\mu_I}\right] = \frac{\beta_b \cdot \bar{I} + \beta_w \cdot (I-1)}{\gamma \cdot I}. \qquad (16)$$

The derivative of the log-likelihood with respect to m is not defined since m can only change in integer increments and the number of terms in the log-likelihood sum (Eq. (4)) depends on m. Nevertheless, substituting the derivative of the birth–death ratios (Eq. (16)) into Eq. (7) provides some information on how the likelihood changes with changing m. We therefore investigated whether using this expression instead of a proper derivative would recover the true value of m.

2.5. Simulating the household SIS model

The parameter estimation procedure was written in *R* (R Development Core Team, 2009) and applied to simulated data. The simulations were performed in *R* using the Gillespie algorithm for continuous time stochastic processes (Gillespie, 1977; Pineda-Krch, 2008). The simulations were structured as follows: The parameters β_b and β_w were described as $\beta_b = \varepsilon \cdot \beta$ and $\beta_w = (1-\varepsilon) \cdot \beta$ for ε-values of 0.01, 0.11, 0.21, 0.31 and 0.41, $\beta = 0.15$ and $\gamma = 1$. The biological rationale was to compare different proportions of within and between population spread while keeping the overall transmission coefficient constant. The local population size m ranged from eight to twenty and the number of sub-populations equaled 10,000/m, rounded to the closest integer to keep the total population size roughly constant at 10,000.

In a first step the self-consistent field approximation (Ghoshal et al., 2004) was compared to the simulated data using the same parameter values as in the simulation to determine the quality of the approximation. In a second step the parameter estimation procedure was applied to the simulated data. The parameter estimation procedure used observations from 500 patches per run and either a single snapshot at time 200 or 100 time snapshots between time 100 and 200. In addition parameters were estimated from a single snapshot of 100 patches.

2.6. Example 2: local populations of Chagas disease vectors

In this example a model is fitted to counts of Chagas disease vectors *Triatoma infestans* (Klug), in Argentine villages. The data come from surveys that were conducted roughly biannually in three different villages from 1993 to 2002. Details of the data collection have been described elsewhere (Cecere et al., 2004; zu Dohna et al., 2007). Here we restrict our analysis to surveys after 1995 since vector populations are more likely to have reached a stationary distribution in that time period (Cecere et al., 2004; zu Dohna et al., 2009). The data consist of 8484 local population counts, ranging from 0 to 44 insects per count.

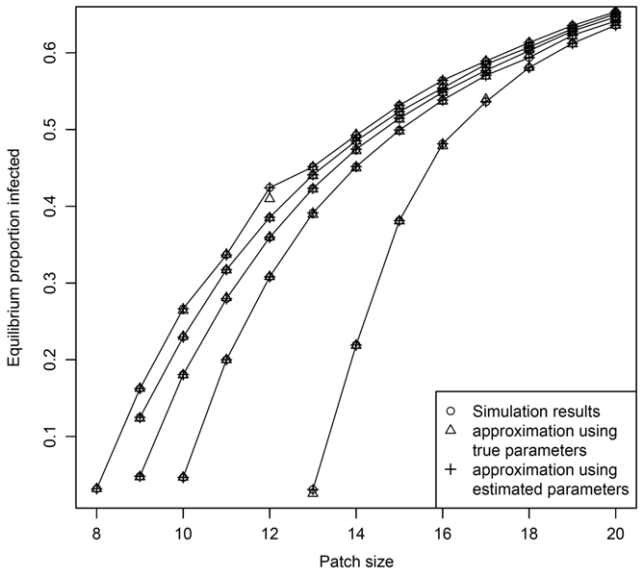

Fig. 1. Equilibrium proportion of infected individuals. Circles denote the proportion generated by the simulation model at time step 200, triangles denote proportions predicted by the self-consistent field approximation using the true parameter values and crosses denote proportions predicted by the self-consistent field approximation using the parameters estimated from the distribution of infected individuals per patch during the last 100 time steps. Each point corresponds to a simulation. Simulations were run for patch size varying from 8–20 individuals per patch (x-axis) and for $\varepsilon = 0.01, 0.11, 0.21, 0.31, 0.41$ (each line corresponds to a different ε).

In the fitted model local population growth is described by a Ricker function (Hassell, 1975). The birth–death rate ratios are hence given by

$$\frac{\lambda_{I-1}}{\mu_I} = \frac{\left(\beta_b \cdot \bar{I} + \beta_w \cdot (I-1)\right) \cdot \exp\left[-\alpha \cdot (I-1)\right]}{\gamma \cdot I}. \tag{17}$$

As in the previous example, only values for the ratios β_b/γ and β_w/γ can be estimated but not for γ. In order to fulfill the condition in Eq. (2), λ_I has to be set to 0 for large I. If this truncation threshold is chosen large enough (e.g. ten times the observed maximum count) it should not affect parameter estimates. Given the ratios

$$\zeta_I = \frac{P(I+1)}{P(I)} \cdot (I+1) \tag{18}$$

and

$$r_{In} = \frac{\zeta_{I+n}}{\zeta_I} \tag{19}$$

the initial parameter estimates can be obtained after substituting the ratios of probabilities by ratios of observed counts

$$\alpha = \frac{1}{2} \cdot \ln\left[\frac{r_{i1} - r_{j1}}{2 \cdot \left(r_{i2} - r_{j2}\right)}\right], \tag{20}$$

$$\frac{\beta_w}{\gamma} = \frac{e^{\alpha \cdot i}\zeta_i - e^{\alpha \cdot j}\zeta_j}{i-j} \tag{21}$$

and

$$\frac{\beta_b}{\gamma} = e^{\alpha \cdot i}\zeta_i - \beta_w \cdot i. \tag{22}$$

3. Results

The self-consistent field approximation fit the simulation results well when the true or the estimated parameters were used (Fig. 1). Three simulation runs ($\varepsilon = 0.11$ and $m = 18$, $\varepsilon = 0.11$ and $m = 19$, $\varepsilon = 0.41$ and $m = 16$) were re-run with a total population size of 1000 since these runs terminated prematurely with the original population size. The parameter estimation procedure is consistent, i.e. converges to the true parameter values, however, there is considerable variation when only one time step is used for parameter estimation (Fig. 2). Our method fails to recover the true parameter values when only 100 patches are sampled in one time step and maximum population size is estimated as well (Fig. 2F). The approximate procedure to estimate the maximum population size m works well for a large number of observations but tends to underestimate the true value for m when only one snapshot is used (Fig. 3).

The following estimates were obtained for the parameters of the Ricker model for Chagas disease vector data (with 95% confidence intervals): $\alpha = 0.001$ ($\pm 1.7 \times 10^{-5}$), $\beta_b \cdot \bar{I}/\gamma = 0.0099$ ($\pm 1.7 \times 10^{-5}$) and $\beta_w/\gamma = 0.97$ ($\pm 3.3 \times 10^{-4}$). The confidence intervals were calculated under the simplifying assumption that all observations are independent. A more detailed exploration of possible dependencies is beyond the scope of this paper. The effect of the truncation threshold on parameter estimates was below the numeric precision of R as long as the threshold exceeded 190. The fitted frequency distribution approximates the observed counts well (Fig. 4, P-value for g.o.f. test $= 0.87$).

4. Discussion

A method has been described to estimate parameters of a stochastic birth–death process from observations of local population sizes in a metapopulation setting. The method was used to fit a household SIS model to simulated data and a metapopulation model to field data of Chagas disease vectors. The estimation procedure can recover the true parameters once enough observations are available.

Our method requires positive birth and death rates within a finite range of possible population sizes and a zero birth rate above that range (Eq. (2)). In addition it requires a large neighborhood size, i.e. a large number of patches influencing the immigration rate on any focal patch. If the neighborhood size becomes too small, local demographic stochasticity leads to large-scale fluctuations (Ghoshal et al., 2004). The two requirements could be fulfilled in a wide variety of biological settings. Examples include SIS models with various transmission functions (McCallum et al., 2001), metapopulation dynamics models or models for the dynamics of macroparasites on hosts (Galvani, 2003).

Even though our method requires a maximum population size above which the birth rate is zero, the example of the insect count data showed that our method can approximate birth rates that reach zero asymptotically as population size reaches infinity by truncating the birth rate to zero above high population sizes. A truncation above four times the maximum observed population size eliminated any detectable effect of truncation threshold on parameter estimates. The model fitted to the insect count data showed evidence for within-population density dependence and provided estimates of the relative contribution of immigration and reproduction to local population growth.

The detailed balancing condition allows fast recursive calculations of the log-likelihood, of its derivatives and of the Fisher Information Matrix. Unfortunately, many model extensions that are important in the disease context such as exposed and immune stages or non-exponentially distributed infectious periods (Vergu et al., 2010) would violate the detailed balancing condition. Such extensions would require numerical optimization techniques that are computationally more expensive. The method presented here is most suitable for describing the dynamics of infectious agents that do not elicit an immune response but are epidemiologically

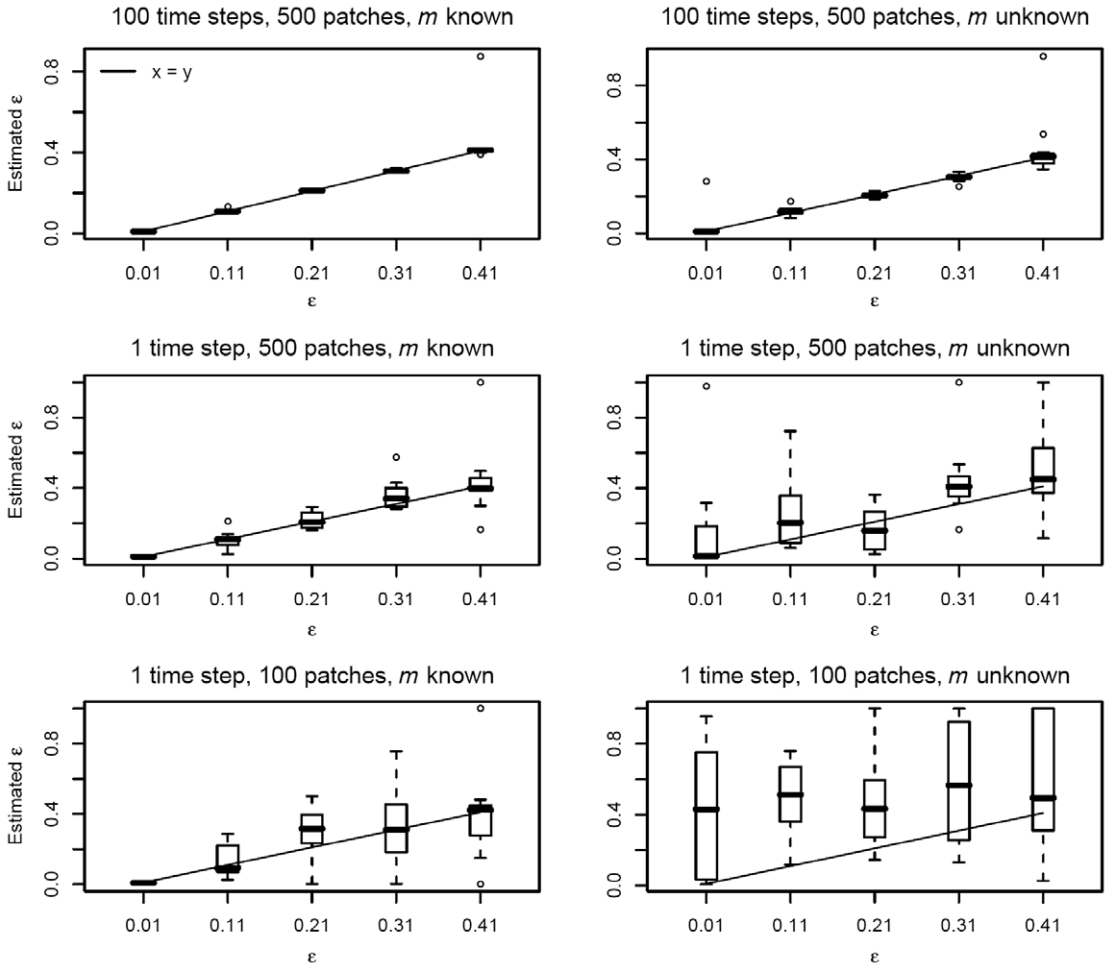

Fig. 2. True (x-axis) and estimated (y-axis) ε-values, based on observations from 500 patches in 100 time steps, 500 patches in one time step or 100 patches in one time step. The maximum population size was either known or estimated. The line indicates all points with $x = y$.

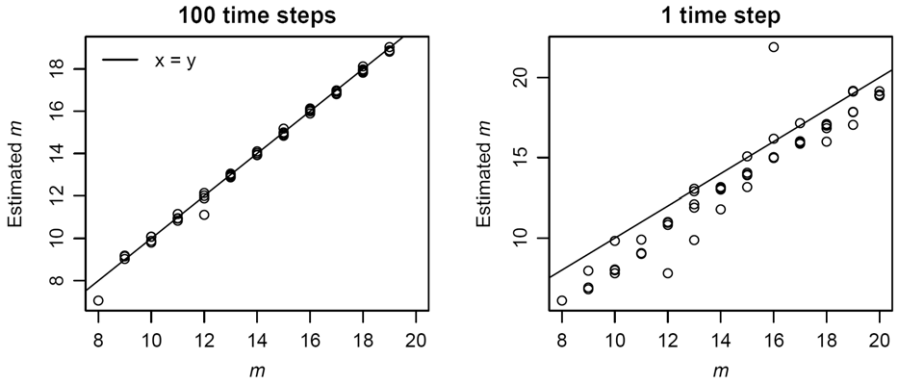

Fig. 3. True (x-axis) and estimated (y-axis) local maximum population sizes (m), based on observations from 500 patches in either 100 time steps or one time step. The line indicates all points with $x = y$. A vertical jitter was added to all points to make overlapping points visible.

important. Examples include commensals carrying antibiotic resistance (e.g. methicillin-resistant Staphylococcus aureus).

Our method could be extended to estimate parameters of a dispersal kernel by making the immigration term of each patch function of the sum of population sizes in all other patches, weighted by their distance. In that case the equilibrium population size of each patch depends on its neighborhood, and the parameter estimation procedure would require solving numerically the equilibrium distribution of patch specific population sizes for each parameter estimate. Such a numerical procedure could use the same recursive relationship presented here to calculate the derivatives of the mean population size of each patch with respect to the mean population size of all other patches. While computationally more expensive than the simple case presented here, the computational effort would not exceed previously proposed methods to estimate metapopulation parameters (Moilanen, 1999).

The parameter estimation method presented here fills two important gaps for fitting epidemiological and ecological models. In epidemiology our procedure extends model fitting methods for household data from models with an immune stage (SIR models)

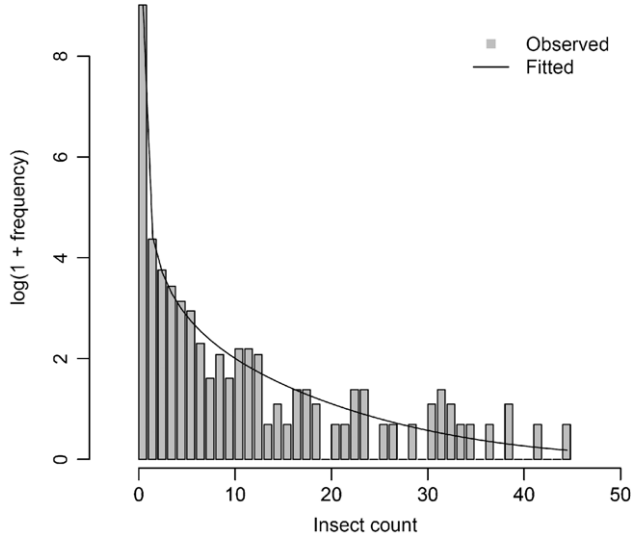

Fig. 4. Fitted (line) and observed (bars) frequency of insect counts. Frequencies are plotted on a logarithmic scale to show the high frequency zero counts and all other count frequencies on the same graph.

to models without an immune stage (SIS models). In ecology our procedure complements methods based on presence–absence data such as the incidence function model. Our procedure requires data on local population sizes rather than presence–absence data of the incidence function model but in return allows separating the effects of immigration from internal population dynamics. The incidence function model estimates parameters based on presence–absence data, i.e. the frequency of zero counts. While each combination of immigration, birth and death rate uniquely specifies the frequency distribution of population sizes, there are infinitely many parameter combinations that lead to different distributions with the same zero count probability. The incidence function model therefore requires other information such as patch size and distance between patches that correlate with the immigration and extinction probability to tease out immigration from patch internal dynamics.

Whether the incidence function model or our method is more appropriate depends on the biological context. The incidence function model is more suitable for cases in which data for patch sizes and interpatch distances are easy to measure and good indicators for extinction and immigration rates, respectively. Our method is more appropriate if local population counts are easy to obtain, and patch size and interpatch distances are either difficult to measure or poor indicators of extinction and immigration rates. An example of a biological setting that is more suitable for our method than the incidence function model could be macroparasite counts on hosts. In some cases count data are easy to obtain and distances between hosts might be hard to measure or uninformative if hosts are mobile.

Acknowledgments

We would like to thank Ricardo Gürtler for providing the data on Chagas disease vectors, Josh O'Brien for helpful discussions and two reviewers for comments that substantially improved the paper.

References

Ball, F., 1999. Stochastic and deterministic models for SIS epidemics among a population partitioned into households. Mathematical Biosciences 156, 41–67.
Cecere, M.C., et al., 2004. Spatio-temporal analysis of reinfestation by Triatoma infestans (Hemiptera: Reduviidae) following insecticide spraying in a rural community in northwestern Argentina. American Journal of Tropological Medicine and Hygiene 71, 803–810.
Galvani, A.P., 2003. Immunity, antigenic heterogeneity, and aggregation of helminth parasites. Journal of Parasitology 89, 232–241.
Ghoshal, G., et al., 2004. SIS epidemics with household structure: the self-consistent field method. Mathematical Biosciences 190, 71–85.
Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81, 2340–2361.
Gillespie, D., 1992. Markov Processes: An Introduction for Physical Scientists. Academic Press, Boston.
Hanski, I., 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63, 151–162.
Hassell, M.P., 1975. Density dependence in single species populations. Journal of Animal Ecology 44, 283–296.
Longini, I.M., et al., 1982. Estimating household and community transmission parameters for influenza. American Journal of Epidemiology 115, 736–751.
Longini, I.M., Koopman, J.S., 1982. Household and community transmission parameters from final distributions of infections in households. Biometrics 38, 115–126.
McCallum, H., et al., 2001. How should pathogen transmission be modelled? Trends in Ecology & Evolution 16, 295–300.
Moilanen, A., 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. Ecology 80, 1031–1043.
Nåsell, I., 2001. Extinction and quasi-stationarity in the Verhulst logistic model. Journal of Theoretical Biology 211, 11–27.
Pineda-Krch, M., 2008. GillespieSSA: implementing the stochastic simulation algorithm in R. Journal of Statistical Software 25, 1–18.
R Development Core Team, , 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
Rao, C., 1953. Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York.
Rice, J.A., 1995. Mathematical Statistics and Data Analysis. Duxbury Press, Belmont, CA.
Vergu, E., et al., 2010. Impact of the infection period distribution on the epidemic spread in a metapopulation model. PLoS One 5, e9371.
zu Dohna, H., et al., 2007. Re-establishment of local populations of vectors of Chagas disease after insecticide spraying. Journal of Applied Ecology 44, 220–227.
zu Dohna, H., et al., 2009. Spatial re-establishment dynamics of local populations of vectors of Chagas disease. PLoS Neglected Tropical Diseases 3, e490.