Implications of empirical data quality to metapopulation model parameter estimation and application

Atte Moilanen

Parameter estimation is a critical step in the use of any metapopulation model for predictive purposes. Typically metapopulation studies assume that empirical data are of good quality and any errors are so insignificant that they can be ignored. However, three types of errors occur commonly in metapopulation data sets. First, patch areas can be mis-estimated. Second, unknown habitat patches may be located within or around the study area. Third, there may be false zeros in the data set, that is, some patches were observed to be empty while there truly was a population in the patch. This study investigates biases induced into metapopulation model parameter estimates by these three types of errors. It was found that mis-estimated areas influence the scaling of extinction risk with patch area; extinction probabilities for large patches become overestimated. Missing patches cause overestimation of migration distances and colonization ability of the species. False zeros can affect very strongly all model components, the extinction risk in large patches, intrinsic extinction rates in general, migration distances and colonization ability may become all overestimated. Biases in parameter estimates translate into corresponding biases in model predictions, which are serious particularly if metapopulation persistence becomes overestimated. This happens for example when the migration capability of the species is overestimated. Awareness of these biases helps in understanding seemingly anomalous parameter estimation results. There are also implications for field work: it may be reasonable to allocate effort so that serious errors in data are minimized. It is particularly important to avoid observing false zeros for large and/or isolated patches.

A. Moilanen, Metapopulation Research Group, Dept of Ecology and Systematics, P.O. Box 17, FIN-00014 Univ. of Helsinki, Finland (atte.moilanen@helsinki.fi).

Metapopulation models, or spatially explicit population models in general, have been suggested to be helpful tools for conservation (e.g. Burgman et al. 1993, Turner et al. 1995, Hanski 1998a, b). Despite the type of the model, parameter estimation is a necessary and important step in the model application (Conroy et al. 1995). Stochastic patch occupancy models (see Moilanen 1999 and references therein) represent a class of metapopulation models that make the presence/absence assumption, that is, a habitat patch has only two possible states, occupied by the focal species or empty. Ignoring local dynamics makes the model relatively parameter sparse, and, importantly, data collection is facilitated tremendously because only the presence of the species

in the habitat patches has to be observed. This simplification makes stochastic patch occupancy models prime candidates for use in real-world applications. Steinberg and Kareiva (1997) argue that the presence/absence assumption is almost compulsory in any large-scale spatial ecology study, as gathering large-scale time series data about population densities requires enormous resources.

Various parameter estimation methods applicable to stochastic patch occupancy models have been presented in the literature (Hanski 1994, Sjögren Gulve and Ray 1996, ter Braak et al. 1998, Moilanen 1999, 2000). A common feature of these studies is that empirical data are assumed to be error free. The same is generally true

of studies that estimate parameters from empirical metapopulation data: errors in empirical data are either entirely ignored or their effects on parameter estimates are assumed to be negligible. Nevertheless, many empirical data sets will necessarily contain errors that can cause systematic and large biases in parameter estimates.

Three types of errors are likely to occur commonly in metapopulation data: (1) Patch areas are estimated/measured inaccurately, or analogously, there is undetected variation in patch quality. (2) There are unknown patches within or around the study area. These are called 'missing patches' below. (3) Patch occupancy is incorrectly observed. This is the 'false zero' problem, with a patch observed to be empty actually containing a population of the focal species.

The aim of this study is to investigate the kinds of systematic biases that are caused in metapopulation model parameter estimates by the three error types described above. The effects of these biases on the basic components of metapopulation models, namely the scaling of local extinction risk with patch area, the distribution of migration distances, and the species colonization ability, are evaluated. To do this, a large number of simulated data sets are produced for metapopulations with different properties. Next, errors are added to the data sets. Parameters are then estimated from both the original and the modified data sets, and parameter estimates are compared in order to find out what kinds of biases were produced. It was found that all error types can cause significant biases in parameter estimates. Understanding the effects of these biases on model predictions and possible conservation decisions is important. Therefore, the biases detected are discussed in terms of their qualitative effects on estimates of metapopulation persistence. Also, consequences for metapopulation data collection are discussed.

Material and methods

Effects of errors in empirical data were tested using simulated data sets, which were generated using the incidence function metapopulation model (IFM; Hanski 1994). In this section I first review the IFM, then the data generation scheme is described, and finally I describe parameter estimation and the statistical analysis of biases.

The incidence function model

The IFM is summarized only briefly here, as the model and its parameter estimation have been extensively discussed elsewhere (Hanski 1994, 1997, 1998a, b, 1999, ter Braak et al. 1998, Moilanen et al. 1998,

Moilanen 1999, 2000). The IFM is used in this study as a convenient example of a stochastic patch occupancy model (SPOM). SPOMs only model the presence/absence of species in habitat patches. A SPOM can be understood as a Markov chain, in which the occupancy of patches at time $t + 1$ is determined by the pattern of occupancy at time t. Different SPOMs can be constructed by making different assumptions about which factors have what kinds of effects on local extinction and colonization of empty habitat patches. It is important to note that the results presented in this study are likely to be qualitatively representative of a large range of metapopulation models. This is because the assumptions of the IFM, (1) colonization probability of an empty patch increases with increasing connectivity, (2) connectivity decreases with increasing distance, and (3) extinction probability of a local population decreases with increasing patch area, are all biologically reasonable and are commonly made in metapopulation models. The specific functional forms assumed by the incidence function model are of secondary importance for the current purpose. See Moilanen (1999) for details about SPOMs and parameter estimation for SPOMs.

In the IFM the probability of local extinction, E_i, depends on patch area by the power function $E^i = e/A_i^x$, where A_i is the area of patch i. Parameters e and x determine the intrinsic extinction rates (extinction rates in the absence of immigration) of local populations and the scaling of extinction risk by patch area, respectively. The colonization probability of an empty habitat patch, $C_i(t)$, depends on patch connectivity by a sigmoid function $C_i(t) = S_i(t)^2/(S_i(t)^2 + y^2)$, where $S_i(t)$ is the connectivity of patch i at time t, and y is a parameter. The connectivity of patch i is computed from the observed patch occupancy pattern at time t:

$$S_i(t) = \sum p_j(t) \exp(-\alpha d_{ij}) A_j^b, \quad j \neq i \qquad (1)$$

where $p_j(t)$ is the state of occupancy of patch j at time t, and d_{ij} is the distance between patches i and j. Parameter α determines the distribution of dispersal distances for the species, with $1/\alpha$ being the average migration distance. Exponent b is a parameter defining the scaling of the emigration rate by patch area. When the rescue effect (Brown and Kodric-Brown 1977) is included in the IFM, E_i is replaced by $E_i(t) = (1 - C_i(t))E_i$ in model parameterization and simulation (using the above functions for E_i and $C_i(t)$). One interpretation for the rescue effect is that the local population is saved from extinction due to increased population size caused by past migration. This means that an isolated population is more likely to go extinct than a well-connected population. For discussion on the rescue effect and the IFM see Hanski (1997).

Generation of simulated data

It is possible that the effects of errors depend on the properties of the species and characteristics of the fragmented landscape. Therefore, simulated data were generated for a range of different systems. The factors that were considered potentially important here were (1) level of patch aggregation in the patch network, (2) distribution of migration distances, (3) strength of the dependence of local extinction risk on patch area and (4) the turnover rate in the metapopulation (the rate at which local populations go extinct and empty patches become colonized). To avoid complications, two levels were set for each of these factors, so that they can be entered as categorical factors into a mixed factorial model. The combination of four variables at two levels produces 16 different combinations. Five different replicate simulated data sets were produced for each combination, which gives a total of 80 simulated data sets. Next, data with errors were generated by modifying the 80 original data sets. For each error class, five modified data sets with additional error were generated from each original data set, which resulted in 400 data sets per error class. The amount of error added into each data set was randomly drawn from a specified range so that the amount of error could be entered as a continuous factor into a mixed factorial model. This is explained in detail below.

Fig. 1 shows the two original patch networks used to produce simulated data. In panel A, 120 patches were randomly distributed into a unit square whereas in panel B the distribution of patches is aggregated. The number of patches was chosen to allow a large proportion ($>50\%$) of patches to be removed without much affecting the persistence of the metapopulation. This may have relevance for the stability of parameter estimation as systems close to extinction may exhibit transient non-equilibrium dynamics that add undesirable variation into parameter estimates.

Two levels of migration capability were produced by using parameters $\alpha = 2$ and $\alpha = 8$ in the negative exponential dispersal kernel ($\exp(-\alpha d_{ij})$) of the incidence function model. The fourfold difference in average migration distance ($1/\alpha$) causes interactions among patches to be either quite localized ($\alpha = 8$) or nearly global ($\alpha = 2$).

The strength of the scaling of extinction risk by patch area is determined in the incidence function model by parameter x, which occurs in the power function $E_i = e/A_i^x$ (see Hanski 1998a, b). Levels of low dependence and high dependence were set to $x = 0.6$ and $x = 1.4$, respectively. Values around 1 for x have been observed in several empirical studies (Hanski 1999).

The final categorical factor used in this study is the turnover rate of local populations. The turnover rate predicted by the incidence function model can be tuned by changing the values of parameters e and y. This has

to be done separately for each combination of aggregation, α and x. For the low and high turnover rates, the numbers of turnover events were tuned to be 6.2 ± 0.2 and 24.5 ± 0.5, respectively. The proportion of occupied patches in the metapopulation was simultaneously kept constant at 0.75 ± 0.02.

The 80 original data sets were produced with simulation using the specified levels of aggregation, x and α, and the parameter values giving the desired level of turnover. Each data set was produced by starting a simulation with the incidence function model with each patch having a 50% probability of being occupied. Then the IFM was simulated for 1000 yr so that the metapopulation had plenty of time to reach stochastic equilibrium from the initial state (recall that the number of turnover events was 6–25 per year). Starting from year 1000, the patch occupancy in the metapopulation was sampled for 20 consecutive years. Each original data set thus consists of patch locations, areas and species presence/absence for 120 patches and 20 yr. This would be a lot of data in an empirical study.

Simulated data were generated for each error class separately by modifying the original data sets. Data sets with mis-estimated areas were generated simply by adding variation to patch areas. Up to 60% of error in patch areas was allowed for, which may be a conserva-
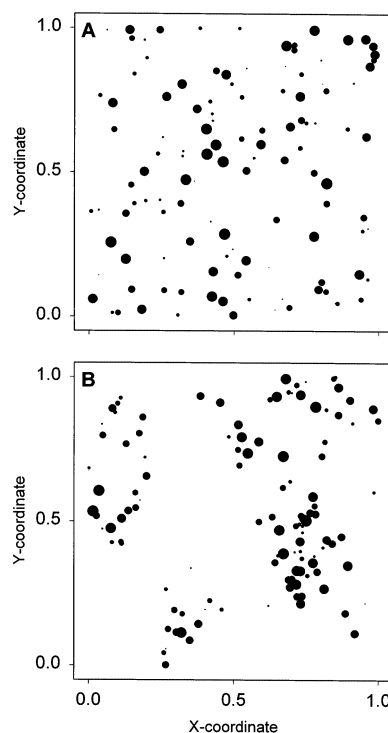

Fig. 1. The two patch networks used to generate simulated data. The distribution of patches is random in panel (A) and aggregated in panel (B). Each patch network has 120 patches with areas drawn from a uniform [0.2, 2] distribution. Circle diameter is scaled by patch area.

tive figure for some systems. The following procedure was used to produce one modified data set. First draw a random number r_1 from a uniform [0, 0.6] distribution. For each patch, multiply patch area by a number randomly drawn from a uniform $[1 - r_1, 1 + r_1]$ distribution. The number r_1 is the range of error associated with the data set. The average error $r_1/2$ was entered as a continuous independent factor into the mixed factorial design.

Data sets with missing patches were generated analogously. It was decided that at most 50% of the patches could be missing. First, a random number r_1 was drawn from a uniform [0, 0.5] distribution, then $r_1 \times 120$ patches were deleted from the data set. Two different patch deletion schemes were used with patches deleted either randomly or aggregatedly. In the latter case the patch network was cut with a line that had a randomly picked direction. The line was positioned so that $r_1 \times 120$ patches were left on the other side of the line, and these patches were removed.

The fourth error type studied here is the false observation of species absence (false zeros). Up to 30% of false zeros were allowed for. The original data set was modified by replacing each occurrence of a population (1) with an empty patch (0) with probability r_1 drawn from a uniform [0, 0.3] distribution.

Parameter estimation and statistical analysis

Parameter estimation was done using the TMC method of Moilanen (1999). This method computes a maximum likelihood estimate for a time sequence of observed patch occupancy patterns without making the equilibrium assumption (Moilanen 2000). The equilibrium assumption here means that the metapopulation that is used in parameter estimation is assumed to be in long-term stochastic quasi-equilibrium. Making the equilibrium assumption is not critical here because the length of a data set (20 time units) is so long that it is unlikely that the data set contains a strong trend (regardless of this the accuracy of the original parameter estimates was checked). Also, the TMC method computes exact likelihoods without resorting to time-consuming Monte Carlo techniques, which are necessary when making the equilibrium assumption (Moilanen 1999). The parameter estimation returns an optimal parameter set $\theta = (x, e, y, \alpha)$. The per-patch log-likelihood model fit error is denoted by $l(\theta)$.

Parameters were first estimated from the 80 original data sets and then from the 400 modified data sets for each error class. The average turnover rates and incidences were determined for all data sets by simulating the IFM with the estimated parameter values. Three replicate simulation runs were used with the first 500 yr ignored and the statistics computed for the next 1000

yr. Average turnover and incidence were calculated for the three replicate runs.

To investigate biases in parameter estimates, the ratios of new (estimated from data with additional error) to original (estimated from original data) parameter values were analyzed using analysis of variance. An alternative to using ratios would have been to use differences between original and new parameter values. The ratio was preferred because now coefficients, that is, the strengths of biases, are directly comparable between different variables, which would not have been the case with differences. The level of patch aggregation, x, α and the turnover rate were entered into a mixed factorial design as independent categorical factors. For aggregation, the factor levels were 0 and 1 for the random and aggregated networks, respectively. For x, α and the turnover rate corresponding factor levels were 0 and 1 for the low (0.6, 2 and 6.2) and high (1.4, 8 and 24.5) values, respectively. Note that for α the low value indicates long migration distances. The amount of error generated into the respective data set was used as an independent continuous factor. Various interactions between the amount of error and one or two categorical factors were also entered into the statistical model as explanatory factors. The dependent variables are the ratios between the new (with error) and original parameter values (x, e, y and α), model fit error ($l(\alpha)$), predicted turnover rate (TO), and average occupancy (p); the ratio is denoted by a prime over the respective parameter. Statistical analysis was done using the general linear models module of Systat for Windows version 5.0.

Including the probability of false zeros into parameter estimation

Moilanen (1999) describes a method of maximum likelihood parameter estimation for SPOMs. This method is based on maximizing the likelihood of the empirically observed sequence of patch occupancy patterns, which can be separated into the probability of arriving at the first observed pattern of patch occupancy multiplied by the probabilities of the subsequent transitions:

$$\max P[O(t_0)] \times P[O(t_1)|O(t_0)] \times \cdots \times P[O(t_M)|O(t_{M-1})]$$

(2)

In eq. (2), $O(t)$ stands for the patch occupancy pattern observed at time t and $M + 1$ is the number of observations. Moilanen (1999) presents equations and Monte Carlo approximation techniques needed to compute the components of eq. (2). An important special case is the transition probability between two snapshots with one time unit (often year) between observations. Denoting by $C_i(t)$ and $E_i(t)$ the extinction and colonization functions that define the SPOM, this probability is:

$$P[O(t+1)|O(t)]$$

$$= \prod_{i=1}^{N} \begin{bmatrix} 1 - C_i(t), & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 0 \\ C_i(t), & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 1 \\ E_i(t), & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 0 \\ 1 - E_i(t), & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 1 \end{bmatrix} \quad (3)$$

where N is the number of patches in the network, $p_i(t)$ is the occupancy of patch i at time t and $E_i(t) = [1 - C_i(t)]E_i$ is the extinction probability of a population in patch i corrected for the rescue-effect (E_i is the intrinsic extinction probability usually defined as a function of patch area). Eq. (3) can be extended to include the probability of observing a false zero (z) by taking into account the fact that observed zeros are not necessarily true zeros. (For example the probability of an observed extinction is the probability of a true extinction $E_i(t)$ plus the probability of a false extinction that remains unobserved, $[1 - E_i(t)]z$.) Now the corresponding equation is:

$$P[O(t+1)|O(t)] = \prod_{i=1}^{N} \begin{bmatrix} (1-z^n)[(1-C_i(t)) + zC_i(t)] \\ \quad + z^n[z(1-E_i(t)) + E_i(t)], & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 0 \\ (1-z^n)C_i(t) + z^n[1-E_i(t)], & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 1 \\ E_i(t) + [1-E_i(t)]z, & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 0 \\ 1 - E_i(t), & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 1 \end{bmatrix} \quad (4)$$

where n is the number of time units patch i has been observed to be empty before and including time t and z^n is the probability of observing n false zeros in a row (the more zeros there are in a row, the less likely it is

$$P[O(t+1)|X(t)] = \prod_{i=1}^{N} \begin{bmatrix} [1 - C_i(t)] + zC_i(t), & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 0 \\ C_i(t), & \text{if } p_i(t) = 0 \wedge p_i(t+1) = 1 \\ [E_i(t)] + [1 - E_i(t)]z, & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 0 \\ 1 - E_i(t), & \text{if } p_i(t) = 1 \wedge p_i(t+1) = 1 \end{bmatrix} \quad (5)$$

that the complete observed extinction-recolonization is false). Eq. (4) is exact for $n = 1$ only; otherwise it is an approximation because it assumes that a row of n zeros is either true or that all zeros are false. This ignores the possibility of intermediate recolonizations and extinctions, which remain unobserved (and which can be observed only with data sets of four or more observations). Such events have the probability of $zC_i(t)[1 - C_i(t+1)]E_i$, which is much smaller than the probability of an truly empty patch staying empty, $1 - C_i(t)$, and thus unobserved recoloniztions and extinctions can be ignored in eq. (4). Essentially, eq. (4) captures the important question whether a sequence of zeros contains at least one true extinction-recolonization event or whether the entire observed row of zeros is a false observation.

At first it might appear that z could be estimated along with the metapopulation model parameters by substituting eq. (4) into eq. (2), but this does not work. What happens is that z is estimated to be 1 (all zeros are thus assumed to be false ones) and the likelihood of the data becomes 1. Instead, what can be done is to estimate z in an independent study and use this value of z in parameter estimation (using eqs [2] and [4]), which is the approach adopted in this study.

The replacement of eq. (3) by eq. (4) is enough to modify the TMC method of Moilanen (1999) to include the effect of false zeros. Another important quantity used in Moilanen (1999) is the probability of moving from a simulated patch occupancy pattern $X(t)$ to an empirically observed pattern $O(t+1)$, which is used in certain Monte Carlo computations (e.g. in the MC method, where the use of the equilibrium assumption requires the estimation of the probability of arriving to the first snapshot). A modified version of this equation is given for reference even though the MC method is

not used in the present study. Given that the simulated occupancy pattern $X(t)$ is known exactly, but the observed pattern may contain false zeros, the probability in question is

Note that a simplistic way of correcting for false zeros would be simply to exchange '101' observations (patch is empty between two occupied years) for large/isolated patches into '111' observations, which essentially removes statistical outliers from the data. Sometimes this might work, but deciding which of the zeros actually are false zeros will be difficult. Hence the method presented above is preferable.

Results

The statistical analyses amount to seven mixed factorial analyses (x', e', y', α', TO$'$, p' and $l(\theta)'$) for each of the four error types. For brevity, statistical tables or figures

Table 1. Key for symbols used in statistical tests.

Symbol	Explanation
A%	average error in patch area
M%	proportion of missing patches
FZ%	probability of false zero
AG	categorical level of aggregation in the data set
X	categorical level of x
α	categorical level of α
TO	categorical level of turnover

are shown only when several moderately strong effects were found (see Table 1 for explanations for symbols). A table with full statistical results is presented only when many independent factors and/or interactions have strong effects. Table 2 compiles important sum-

mary information for the 28 statistical analyses. Among other things, Table 2 reports the proportion of variance explained, up to four of the strongest statistically significant effects, and ranges for summed coefficients (to give an idea of the minimal and maximal effects caused by the error type). Usually two or fewer dominant factors were found, so for almost all cases Table 2 summarizes the relevant features of the statistical analysis.

To check for the correctness of the parameter estimation procedure, the originally estimated and true parameter values were first compared. Averages (and standard errors) in data sets with low (true 0.6) and high (true 1.4) values for parameter x were 0.61 (0.12) and 1.37 (0.15), respectively. Similarly, for α, estimates

Table 2. Summary of results of the 28 mixed factorial analyses used to explore effects of errors in metapopulation data. Columns are: (1) bias that is investigated (ratio of new to original value), (2) R^2; the squared multiple correlation coefficient giving the proportion of variance explained, (3) the number of significant factors ($P < 0.05$) found in the analysis, (4) and (5) up to four factors/interactions having the highest magnitude coefficients; the F-ratio (DF = 1,400), coefficient and name are given for each factor, (7) and (8) the minimum and maximum coefficients summed for the significant factors/interactions. See Table 1 for explanations of symbols. [F-ratio values for $P = 0.05$ and $P = 0.01$ levels of statistical significance equal $F_{0.05}(1,400) \approx 3.9$ and $F_{0.01}(1,400) \approx 6.8$, respectively.]

Effect	R^2	#sign.	Effects 1/3 F	Effects 1/3 COEFF	Effects 1/3 Name	Effects 2/4 F	Effects 2/4 COEFF	Effects 2/4 Name	Coeff. range Min	Coeff. range Max
					Mis-estimated patch areas					
α'	0.031	0		–			–			
y'	0.020	0		–			–			
x'	0.550	6	305.4	−0.363	A%	34.6	0.122	A%×X	−0.363	−0.171
e'	0.208	4	25.7	0.041	A%×AG	13.9	−0.030	A%	−0.030	0.038
TO′	0.018	0		–			–			
p'	0.049	2	4.02	0.003	TO	6.43	−0.002	A%×TO	0	0.001
$l(\theta)'$	0.494	4	181.6	0.034	A%	80.2	−0.023	A%×X	0.020	0.040
					Randomly missing patches					
α'	0.257	7	12.4	−0.233	M%×X	9.67	−0.204	M%×AG	−0.562	−0.419
y'	0.407	8	116	−0.674	M%	12.1	−0.218	M%×α	−0.892	−0.339
x'	0.065	2	10.6	−0.066	M%×X×AG	4.69	0.027	X	−0.039	0
e'	0.081	1	15.8	0.059	M%×X×α		–		0	0.059
TO′[1]	0.205	8	10.8	−0.074	M%	6.36	−0.057	M%×α	−0.254	−0.074
p'[1]	0.060	2	12.3	−0.025	M%×X×α	4.91	0.016	M%×AG×α	−0.025	0.016
$l(\theta)'$	0.092	3	18.7	0.058	M%	8.21	−0.039	M%×X	0.021	0.058
					Aggregatedly missing patches					
α'	0.084	2	6.38	−0.209	M%	8.12	−0.121	M%×AG×α	−0.330	−0.209
y'	0.193	3	12.7	−0.359	M%×A	8.66	0.295	M%×TO	−0.359	0.095
			9.46	0.159	M%×AG×α		–			
x'	0.111	2	29.1	0.262	M%	4.40	−0.103	M%×AG	0.159	0.262
e'	0.067	1	4.34	−0.057	M%×TO		–		−0.057	0
TO’	0.107	2	6.11	−0.082	M%×X	15.4	0.067	M%×AG×X	−0.082	−0.015
p'	0.224	4	71.6	−0.131	M%×X×α	8.10	−0.086	M%×α	−0.217	−0.197
$l(\theta)'$	0.440	3	45.6	−0.150	M%×X	122	0.125	M%×X×α	−0.150	0.067
					False zeros					
α'	0.431	4	157	−1.51	FZ%	6.13	0.297	FZ%×α	−1.212	−1.564
			12.0	−0.204	FZ%×AG×α	4.82	−0.130	FZ%×X×α		
y'	0.757	6	81.0	−1.99	FZ%×α	78.6	−1.97	FZ%×TO	−1.807	1.840
			68.1	1.84	FZ%	100	1.09	FZ%×AG×α		
x'	0.811	3	1235	−2.10	FZ%	4.34	0.125	FZ%×TO	−2.105	−2.087
e'	0.933	6	2596	29.7	FZ%	1554	22.8	FZ%×TO	24.7	62.7
			143	6.92	FZ%×AG	37.3	−5.01	FZ%×X		
TO′	0.407	6	91.1	−5.07	FZ%	13.9	−1.97	FZ%×α	−8.53	−4.649
p'	0.775	5	1092	−3.80	FZ%	7.23	−0.308	FZ%×α	−4.11	−3.71
$l(\theta)'$	0.827	5	1170	1.36	FZ%	41.2	−0.253	FZ%×α	0.91	1.36

[1] Three data sets that produced parameters predicting metapopulation extinction were omitted.
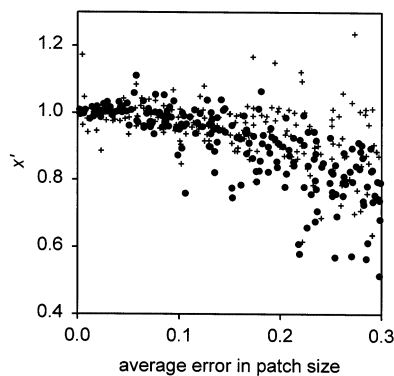
Fig. 2. The effect of mis-estimated patch areas on the scaling of extinction risk. Crosses and circles are for data sets generated using high and low levels for parameter x, respectively.
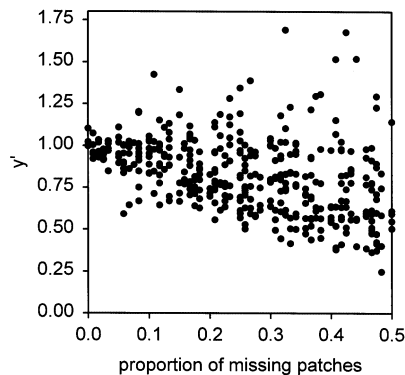

Fig. 3. The effect of randomly missing patches on the colonization parameter of the IFM (y). Decreased values for y indicate that colonization is possible with lower connectivity values. Note that simultaneously increased migration distances (decrease in α) actually indicate increased connectivity.

were 2.0 (0.65) and 7.9 (1.2) for data sets with true values of 2 and 8, respectively. Thus original estimates for x and α were unbiased and reasonably accurate. Predicted turnover rates were also compared to true rates. Rates 5.1 (0.59) and 20.5 (0.86) were obtained for data sets with originally 6.2 and 24.5 turnover events

per time unit. Thus the parameter estimation slightly underestimated the turnover rates, but the difference to true values was small enough to be of little concern. In summary, the TMC method (no equilibrium assumption) produced parameter estimates that were sufficiently accurate and unbiased for the present purpose.

Mis-estimated patch areas

This is the error type that produces clearly the smallest and simplest biases into IFM parameter estimates; only the scaling of extinction risk by patch area is strongly affected. Fig. 2 shows how the value of x' decreases with increasing error in patch areas (see Table 2 for the dominant effects in the associated statistical analysis). The other extinction parameter e was only weakly affected; four mutually conflicting (effects in different directions) statistically significant effects were found, but the largest coefficient was only 0.041 for average error in area (A%) × aggregation level of the patch network (AG), which means that the effect is small enough to be negligible for practical purposes. Note that the number of simulated data sets (400) was so large that even very weak effects could be statistically highly significant. No significant effects at (5% level) were found for the colonization parameters (α and y) or turnover rate, and the effect on incidence of occupancy was very weak (Table 2). The model fit error was slightly increased by A% (COEFF = 0.034, Table 2).

Randomly missing patches

Randomly missing patches cause large biases in the colonization parameters α and y. For y these effects are very much expected (Fig. 3, Table 3). When patches are missing, the observed connectivity values (Eq. [1]) will be smaller than when all patches are known. This is simply because the contribution of the unknown patches to connectivity is missing. It follows that the observed colonization events have to be explained with

Table 3. Mixed factorial analysis for the effects of randomly missing patches to colonization (y'). $N = 400$, multiple $R^2 = 0.407$. Coefficients are given for statistically significant effects ($P < 0.05$) only. See Table 1 for explanations of symbols.

Source	SS	DF	MS	F-ratio	P	Coeff.
α	0.032	1	0.032	1.063	0.303	
AG	0.031	1	0.031	1.020	0.313	
X	0.292	1	0.292	9.713	0.002	−0.057
TO	0.185	1	0.185	6.164	0.013	−0.045
M%	3.489	1	3.489	116.094	0.000	−0.674
M% × α	0.364	1	0.364	12.126	0.001	−0.218
M% × AG	0.385	1	0.385	12.824	0.000	0.233
M% × X	0.410	1	0.410	13.648	0.000	0.232
M% × TO	0.118	1	0.118	3.933	0.048	0.124
M% × AG × α	0.109	1	0.109	3.642	0.057	
M% × X × α	0.150	1	0.150	4.988	0.026	0.066
M% × X × AG	0.000	1	0.000	0.002	0.964	

522

Table 4. Mixed factorial analysis for the effects of randomly missing patches to migration distances (α'). $N = 400$, multiple $R^2 = 0.257$. Coefficients are given for statistically significant effects ($P < 0.05$) only. See Table 1 for explanations of symbols.

Source	SS	DF	MS	F-ratio	P	Coeff.
α	0.006	1	0.006	0.185	0.668	
AGGR	0.020	1	0.020	0.586	0.444	
X	0.165	1	0.165	4.947	0.027	0.043
TO	0.166	1	0.166	4.969	0.026	0.043
M%	0.216	1	0.216	6.466	0.011	-0.168
M% $\times \alpha$	0.306	1	0.306	9.171	0.003	0.199
M% $\times$ AGGR	0.323	1	0.323	9.673	0.002	-0.204
M% $\times$ X	0.415	1	0.415	12.442	0.000	-0.233
M% $\times$ TO	0.003	1	0.003	0.076	0.783	
M% $\times$ AG $\times \alpha$	0.009	1	0.009	0.283	0.595	
M% $\times$ X $\times \alpha$	0.332	1	0.332	9.970	0.002	-0.099
M% $\times$ AG $\times$ X	0.005	1	0.005	0.145	0.704	

lower connectivity values, which in the IFM translates into a reduced value of y (proportion of missing patches, COEFF $= -0.674$, Table 2).

Importantly, randomly missing patches also affect the estimate of migration distances (α). At first it may seem that this effect should be expected, because average interpatch distances increase when patches are missing. However, the dispersal kernel scales the relative contributions of patches at different distances to connectivity and it is not obvious that α should change when the relative contributions of patches at different distances do not actually change. Even so, α is moderately strongly affected by randomly missing patches (Table 4). Apparently some of the effects of missing patches are mixed between y and α, α becomes underestimated, and the migration distances are thus overestimated. Note that changes in y and α are related to each other. A decrease in α will increase connectivity values, which requires an increase in y to have constant colonization probabilities. In this case both α and y decrease, which means that the effect of decreased y is stronger than it appears.

Apart from the effects on migration and colonization, randomly missing patches have only minor effects. The scaling of extinction risk is weakly affected, the effects on x and e have coefficients that are small enough to be negligible (Table 2). At first it appears that the predicted turnover rate (TO') and proportion of occupied patches (p') are moderately affected by randomly missing patches. For TO' there are eight weak effects that sum to a maximum coefficient of -0.413 for an aggregated system and for p' there are two effects that sum to -0.084 for an aggregated system. However, these effects are mostly caused by three of the 400 estimations, which produced parameter values that do not predict metapopulation persistence. When these cases are excluded from the data, the effects on TO' and p' mostly disappear (Table 2). For p' there still are two significant effects, but the minimum summed coefficient is only -0.025. For TO' there are eight significant effects, and the minimum summed coefficient is now -0.253. Thus the effect on incidence

is insignificant and the effect on turnover is not large enough to be of great importance. The model fit error is slightly increased by missing patches (COEFF = 0.058, Table 2).

Aggregatedly missing patches

The pattern of effects for aggregatedly missing patches is similar to that for randomly missing patches, but overall the effects are much weaker. For α' the significant effects have a coefficient range from -0.330 to -0.209 and for y' from -0.359 to -0.095 (Table 2, Fig. 4). Comparing the proportion of variance explained (multiple R^2), the effects are less than half as strong as with randomly missing patches – now only patches at the edge of the metapopulation near the missing patch aggregate are strongly affected.

The effect of aggregatedly missing patches on the scaling of extinction risk (x and e) is small (Table 2). Interestingly, if the original patch network and patch occupancy pattern are aggregated (aggregation, high α, high x), the proportion of occupied patches in the metapopulation is predicted to be slightly too small (coefficient range from -0.217 to -0.197, Table 2). It
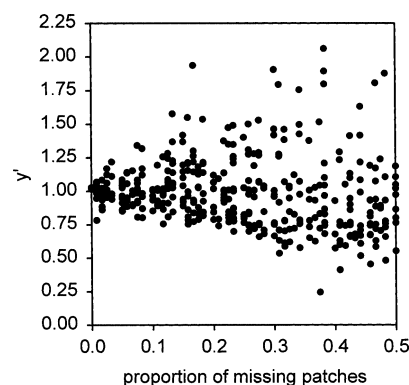

Fig. 4. The effect of aggregatedly missing patches on the colonization parameter of the IFM (y).

was discovered by examining simulations that this can occur when a patch cluster is partially cut off by the line of missing patches. The remaining part of the cluster may then be predicted to be unviable, and the proportion of occupied patches in the metapopulation is then predicted to be too small. Turnover is very weakly affected by aggregatedly missing patches (maximum coefficient -0.082, Table 2). Model fit error is weakly affected in mutually conflicting ways by interactions of aggregation parameters (Table 2).

False patch occupancy information

All effects seen so far pale in comparison with biases caused by false zeros in the data set. All model parameters, the predicted turnover rate, the average occupancy level and model fit error are very strongly affected by false zeros. The following results are essentially caused by interactions of three primary effects: (1) Populations on large patches have intrinsically low extinction rates. Therefore false zeros observed for large patches may cause a strong overestimation of extinction rates for large patches. (2) False zeros observed for isolated patches, followed by equally false colonization events, cause a potentially enormous overestimation of species migration/colonization ability. (3) Overestimated colonization probabilities lead to an increase in the strength of the rescue effect. Therefore false zeros observed for well-connected and possibly large patches lead to hugely overestimated intrinsic extinction rates, because otherwise extinction events would not be observed due to the strong rescue effect.

The effects on migration distances caused by false zeros are clear (Fig. 5A, Table 2). The value of α' is decreased by approximately $1.5 \times FZ\%$ (the probability of observing a false zero), which is a large enough effect to cause a serious overestimation of long-distance migration. Parameter y is affected in conflicting ways. The effect is smallest for a system with no aggregation, low α and low turnover rate. In this case y is increased by $1.84 \times FZ\%$, which however still means an increase in colonization probabilities, because α is simultaneously strongly decreased. Most data sets, however, showed a decrease in y (Fig. 5B, Table 2).

The scaling of extinction risk by patch area is very strongly affected by false zeros (Fig. 5C, Table 2). In these data, 10% of false zeros cause a 50% decrease in x. The strongest effect of false zeros is on model parameter e, the value of which is hugely increased (Fig. 5D, Table 2), for reasons given above.

The observed proportion of occupied patches in the metapopulation is decreased by false zeros, which is reflected in parameter estimates and predictions (Fig. 5E, Table 2). While the proportion of occupied patches decreases, the turnover rate of the metapopulation increases (Fig. 5F), which follows from the numerous extra turnover events created by false zeros. With many false zeros, the metapopulation size may drop to zero, that is, the estimated parameters no longer predict long-term persistence. The effects of false zeros observed for different patches may be mutually conflicting, which is reflected in wide variation in subsequent model predictions (Fig. 5E, F). In this particular case, even 10% of false zeros can cause so much confusion in model parameterization that the metapopulation can erroneously be predicted to lose its long-term persistence. For comparison, only a couple of the patch system variants with up to 60% missing patches were predicted to become unviable.

The model fit error is enormously increased by false zeros (COEFF $= 1.358$, Table 2), which means that the goodness-of-fit of the model is reduced, because randomly applied false zeros do not obey the rules of metapopulation dynamics.

Including false zeros into parameter estimation

Fig. 6 shows similar results as Fig. 5, but with the difference that in the estimations used to produce Fig. 6 the correction for the probability of false zeros (Eq. [4]) has been employed. Now almost all of the biases evident in Fig. 5 have disappeared. However, as the probability of false zeros increases, much variability enters into the parameter estimates, which implies that frequent false zeros are still quite detrimental for parameter estimation. Interestingly, when the probability of false zeros is high, x and α are sometimes badly overestimated. Apparently in some data sets all absences in large (and especially large and isolated) patches can be explained as false zeros, and x and α can correspondingly be increased, because stochastic extinction then never has to happen in these patches.

Discussion

This study investigates the effects of errors in empirical data to metapopulation model parameter estimates and subsequent predictions. In summary, mis-estimation of patch areas causes decreased scaling of effect (here extinction risk) with patch area. Other phenomena that may scale with patch area include per-capita emigration and immigration rates. Missing patches cause the colonization and migration abilities of the species to be potentially badly overestimated, which may cause an overestimation of the ability of the species to survive in the presence of habitat loss. By far the largest problems in parameter estimation are caused by error in the measurement of patch occupancy (false zeros). False zeros affect strongly all model components, and they significantly decrease the reliability of predictions made using metapopulation models.
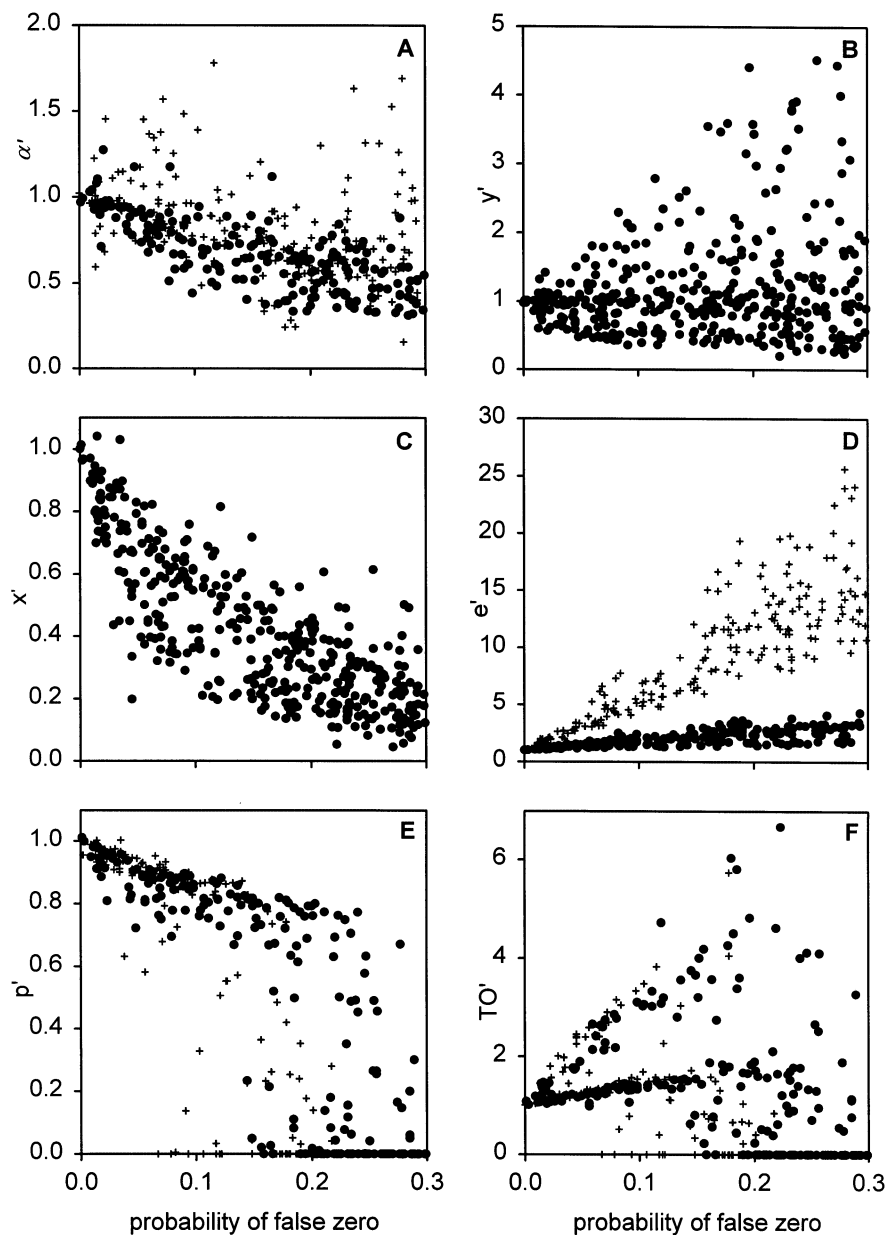
Fig. 5. Effects of false zeros on IFM parameters and model predictions. (A) and (B) show how migration distances and the species colonization abilities are overestimated. In (A) data sets with high original α are marked with crosses. (C) shows how the dependence of extinction risk on patch area decreases very strongly. (D) shows how overall extinction rates increase enormously, which is related to an increased turnover rate (data sets with a high original level of turnover are marked with crosses). (E) and (F) show how a lot of variation is introduced into model predictions due to false zeros (data sets with high original α are marked with crosses).

A moderately strong decrease in the strength of the scaling of extinction risk by patch area is the only notable effect caused by mis-estimated patch areas. Analogous behavior can be expected for any area-related parameter in any metapopulation model, as the explanatory effect of patch area should vanish when the correlation between observed and true patch areas decreases to zero. Fortunately, gaining accurate estimates

of patch areas should be relatively easy using remote sensing and GIS-based methods, and therefore the mis-estimation of geographical patch area should not be a large source of error. However, unaccounted variation in patch quality may have a similar effect, and habitat quality is much more difficult to measure than patch area. Note that there may be a general trend that areas of large patches will be overestimated, because large
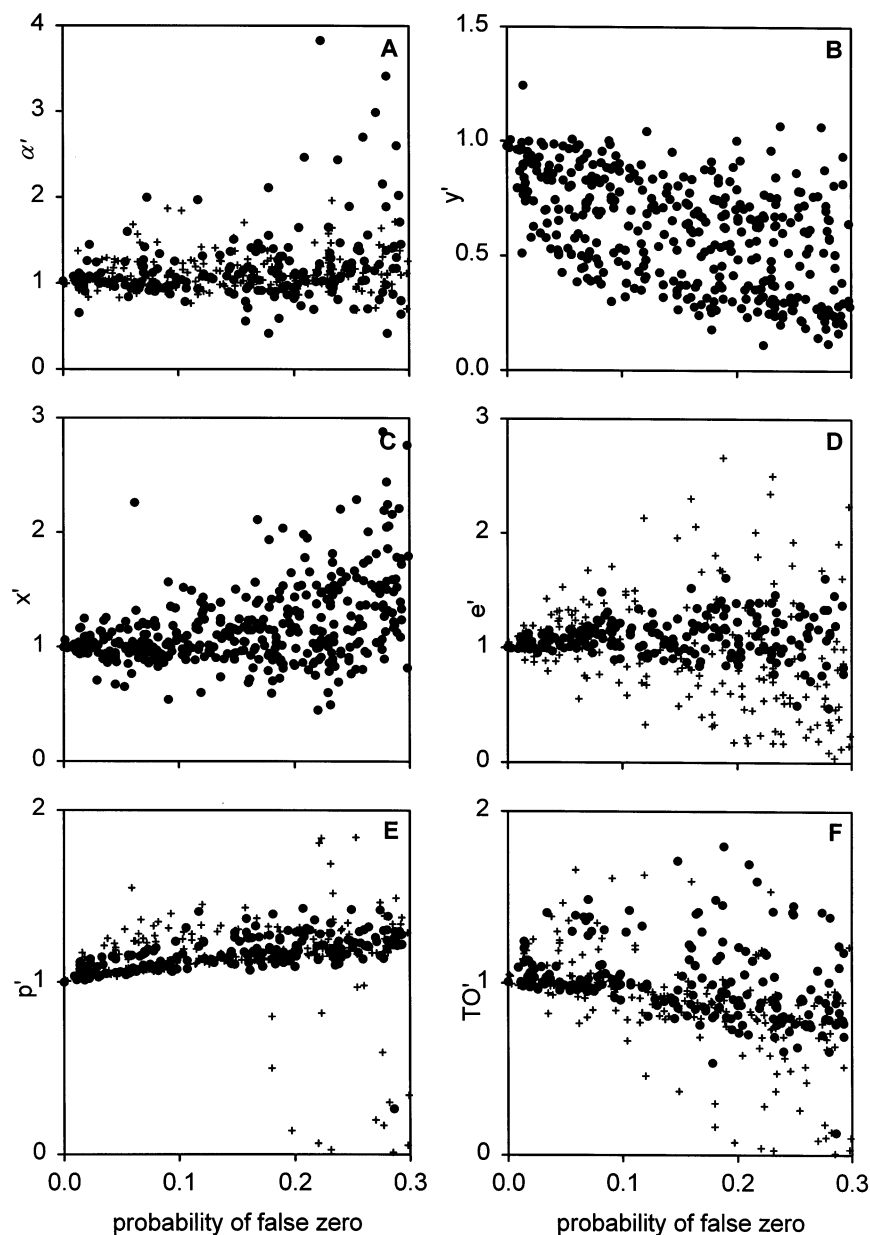
Fig. 6. As Fig. 5, but after correction for the probability of false zeros. Note different vertical scales in the panels of Figs 5 and 6.

patches are likely to also include low-quality habitat, but this will depend on the particulars of the patch network and the method used to estimate patch sizes. Differences in patch shapes, and consequent differences in emigration and immigration, could also affect the persistence of local populations by confusing the assumed relationship between patch area, population size and extinction probability. The present results agree with those of Ruckelhaus et al. (1997), who investigate the effects of mis-classification of habitat quality to dispersal success in a spatially explicit population model (parameters were not estimated). They find that only 1–17% of error is caused in dispersal rates, which

result is dependent on the specific assumptions made in the study. Ruckelhaus et al. (1997) did not consider consequences other than errors in predicted dispersal rates.

The major effects caused by missing patches are an overestimation of colonization ability and an increase in estimated migration distances. Missing patches cause a reduction in apparent connectivity, and therefore the observed colonization events have to be explained with lower connectivity levels, which leads to an overestimation of colonization ability. This is not necessarily a problem if predictions are produced for the same incompletely known patch network. However, if the in-

correctly parameterized model is applied to another completely known patch network, too high levels of patch occupancy and too strong metapopulation persistence will be predicted.

Colonization ability is overestimated more strongly when patches are missing randomly than when patches are missing in an aggregated manner at one edge of the metapopulation. This can be understood by thinking about the contribution of the missing patches to the dynamics of the known patches. The effect of a missing patch declines with distance. If patches are missing from inside the study area, then the missing patches are well connected to the rest of the metapopulation and thus their effect is large. If patches are missing aggregatedly from outside the study area, then the missing patches are relatively far away and possibly only one edge of the metapopulation is strongly affected. This result also implies that a couple of unknown isolated patches (a common situation in empirical studies) should not have any serious effects on a metapopulation study. However, the smallish effects seen for aggregatedly missing patches apply to this study, but not for all possible situations. Here patches were deleted from one edge of the patch network only. If the study area is small compared to the migration distances of the species and if patches are missing from all around it, effects will naturally be larger.

Some guidelines for field studies may be given based on these results. It is better to have a smaller area completely surveyed than a larger area with incomplete knowledge of the habitat patches. If it is known that resources will not suffice for a survey of all habitat patches, then, if possible, it is desirable to select a part of the patch network that is as little connected to missing patches as possible. Incidentally, it may be possible to detect missing patches based on local lack of fit between observed and predicted patch occupancy – if some small and isolated patches are occupied, it may be that an unknown cluster of patches is located near them. Fortunately, missing patches are in general more likely to be small than large, which means that their contribution to connectivity is smaller than the number of missing patches might give reason to believe. Also, patches are more likely to be missing from around an intensively studied core area than from within it. Consequently, one might draw the tentative conclusion that missing patches are unlikely to be a large problem for most metapopulation studies.

Long-distance migration and colonization can be particularly important for the persistence of metapopulations that consist of relatively isolated groups of patches, as opposed to the situation in which the patch network is one dense cluster. If long-distance migration and colonization have to be parameterized reliably, missing patches can become a problem. Statistically reliable parameterization of long-distance colonization requires some observations of these rare events. However, when the study area is expanded to include a number of isolated patches, the geographical area covered by the metapopulation may increase enormously. It may then require a big effort to survey the large piece of landscape in order to ensure that the supposedly isolated patches are truly isolated. If missing patches happen to be located around the supposedly isolated patches, long-distance colonization probabilities will naturally become overestimated.

False zeros caused the largest biases found in this study – extinction probabilities of large patches, intrinsic extinction rates, migration distances and colonization ability are all potentially very strongly overestimated. Recalling that large patches can be critical for metapopulation persistence and that long-distance migration is critical for large-scale metapopulation dynamics, it is obvious that false zeros in data that are used in parameter estimation can substantially decrease the predictive ability of the metapopulation model. Essentially, false zeros introduce statistical outliers into the occupancy data. When a false zero is observed, there is a false extinction event followed by a false colonization event. The strength of the biases caused by false zeros depends on the relative likelihoods of observing false zeros, likelihoods of extinction events of local populations on large patches and likelihoods of colonization for empty isolated patches.

Overall, it is highly desirable to avoid false zeros in data sets that are used in parameter estimation. Worst consequences are caused by false zeros observed for isolated patches as false long-distance colonization events may cause a strong overestimation of migration distances and colonization ability, which translates into an overestimate of metapopulation persistence. However, note that the effects of false zeros may here depend on the number of years with empirical data. With many years of data, colonization rates and dispersal distances are bound to be overestimated, but with only one year of data the situation may be the opposite. This is because there usually are only a few isolated and occupied patches in metapopulation data, and with one observation only, the occupancy of these isolated patches may be missed and consequently dispersal distances underestimated.

The possibility of having false zeros in empirical data casts doubt on our ability to estimate metapopulation model parameters, unless the probability of false zeros (z) can be reliably estimated from an independent study, in which case a correction such as described in this study can be used in parameter estimation. The control study for determining z can be performed at the same time as the metapopulation survey: typically patches are initially surveyed using (for example) a constant search time per area unit. Following the initial survey, z can be estimated by resurveying a number of patches that were initially observed to be empty using a

substantially increased effort. Now z is the proportion of patches that were empty in the initial survey, but occupied in the control study. The caveat here is that if the number of initially empty patches is small, then the estimate of z is likely to be inaccurate. Also, the control study for z naturally requires more work, but considering the very strong adverse effects caused by false zeros observed in this study, this effort is likely to be well spent.

The seriousness of the consequences of false zeros may affect the allocation of effort in an empirical study. It obviously is more important to avoid false zeros than to observe patch areas very accurately. Also, effects of false zeros are much stronger than the effects of an incompletely surveyed patch network (missing patches). Consequently, it may be preferable to go for a thorough survey of a part of a known (very large) patch network rather than do a hasty survey of the complete set of patches. However, it should be remembered that for a metapopulation study to be successful, the patch network must contain patches with a large enough range of area and connectivity (isolation) values. Otherwise statistically significant effects of area or isolation will be difficult to detect, and confidence intervals for parameters will be large. Although it is hard to be specific, in practical terms this requires at least some tens of patches with varying sizes, and a region which is a couple of times the average dispersal distance in diameter. All these different aspects of data amount and quality must be considered when designing an empirical study. It is possible to analyze the effects of study system size and structure, data quality and number of observations by generating simulated data from different alternative patch networks and then estimating parameters from these simulated data.

If the probability of false zeros (z) is included in parameter estimation, it is reasonable to perform sensitivity analysis in respect of z in order to check that other parameters are not highly sensitive to small errors in the estimation of z. The quantitative effects caused by false zeros depend on assumptions made about their occurrence. Here it was assumed that z is independent of patch size and isolation, which will not apply to all systems. If population density is the same for small and large patches, and assuming that a constant time per area unit is used for the survey, z will be smaller for large patches. In contrast, if the population size would be the same in a large and a small patch, z may be greater in a large patch. However, the effects caused by false zeros will not change qualitatively. Unfortunately for real studies, is z likely to be larger for large patches, because it may be very difficult to reliably check the occupancy status of a large patch with a limited workforce in a reasonable time. This is true especially for small animals that are difficult to observe. The effects of false zeros are bound to be more serious in parameter estimation if z is correlated positively with patch

size. This is because large patches are the ones that have most impact on metapopulation dynamics and which are most important for metapopulation persistence, and frequent false zeros observed for large patches will seriously undermine the predictive ability the model has considering the dynamics of large patches. This can be understood by considering relative error rates. It does not matter very much if a small patch has an observed extinction rate of 51% or 60% instead of a true rate of 50%. However, considering a large patch with a true extinction rate of 1%, it makes a huge relative difference if the observed rate is 2% ($z = 0.01$) or 11% ($z = 0.1$). Consequently, biases in parameter estimates and predictions will be more serious when z is positively correlated with patch area.

Analogously to false zeros, it may also be possible to have 'false presences', that is, a patch is considered occupied when it is not, but this error is likely to be uncommon for most species, because generally an observation of the presence of individuals on a patch can be interpreted as a presence of a population (there is more scope for error if indirect ways of observing species presence are used). Effects of false presences could have been studied similarly as the effects of false zeros, but this was not done here, because the effects of false ones are qualitatively similar to the effects of false zeros as both generate erroneous extinction and colonization events into the data.

False zeros may have relevance for the choice of the dispersal kernel of the metapopulation model. In this work a negative exponential dispersal kernel was used; another alternative form would be $1/(1 + \alpha d_{ij}^{\beta})$, which allows for a fatter tail in the dispersal kernel (Shaw 1994a, b). If habitat density in the metapopulation is high or the area of the metapopulation is small compared to the migration distances of the species, it is unlikely that the functional form of the dispersal kernel would make a big difference to the model fit and use. The situation changes when the metapopulation includes also isolated patches or small groups of patches, as the tail of the dispersal kernel may significantly affect the connectivity and dynamics of these isolated patches. The results obtained here about false zeros imply that the fit of a dispersal kernel may be strongly affected by false zeros. A false zero observed for an isolated patch may make the difference between whether a thin- or fat-tailed dispersal kernel fits the data best. It follows that data quality is of great significance if the shape of the dispersal kernel is to be deduced from data. If false zeros cannot be avoided, it may be best to rely on direct observations of dispersal events (e.g. mark-release-recapture data) in the estimation of the dispersal kernel.

It would be convenient to be able to dispense with the data quality problem by just assuming that data quality is adequate. However, considering the substantial logistic problems associated with large-scale ecolog-

528

ical studies, it is likely that all large metapopulation data sets contain some error. Metapopulation studies may have to cover large geographical areas, and given limited resources, a trade-off between study size and data quality may exist. I will conclude by describing known errors in three metapopulation data sets with which I have worked. There is no reason to assume that these data sets are exceptionally bad or exceptionally good in terms of data quality.

The Glanville fritillary butterfly (*Melitaea cinxia*) metapopulation on the Åland islands, Finland, is probably the largest and most thoroughly studied metapopulation anywhere (see e.g. Hanski 1998a, b, 1999 for details and references). Here I summarize only a few details pertinent to data quality. During the first survey in 1993, 1500 habitat patches suitable for *M. cinxia* were found. In 1998 and 1999 the entire study area was searched again and 2900 new habitat patches were found. Most of these patches are small, peripheral and located outside the previously known patch aggregates (Nieminen, Pöyry and Hanski unpubl.), but nonetheless the number of new patches is high enough to affect parameter estimation (fortunately a good-quality subset of the data can be selected for estimation purposes). Errors in patch areas and the probability of observing false zeros have been estimated with independent control studies (Nieminen, Pöyry and Hanski unpubl.). The probability of observing false zeros is approximately 10%. When different pairs of students independently estimated the areas of test patches, the estimates generally varied from -50% to $+100\%$ around the average. Patch area can be measured accurately using GIS, but GIS does not necessarily help delineate the patch in the first place, nor does it help account for variability in patch quality.

A data set for the false heath fritillary butterfly (*Melitaea diamina*; Wahlberg et al. 1996, Wahlberg 1997) also contains various errors, which affect parameter estimation (Moilanen and Cabeza 2002). The original survey of the *M. diamina* patch network was conducted in 1994 and patch occupancy was resurveyed in 1999. Between 1994 and 1999, 14 patches had disappeared due to agriculture, construction or forestry, and 36 previously unknown patches were discovered. When patch areas were reestimated in 1999 using GIS methods, an average difference of 400% between 1994 and 1999 area estimates was discovered. This difference is partly due to habitat loss, which has changed some patch areas greatly, but differences of several hundred percent were observed even for patches that had not changed in the field. The worst potential source of errors in the *M. diamina* data set is deterministic succession. *Melitaea diamina* inhabits moist meadows that become overgrown in approximately 20 yr if the vegetation is not kept low. The 1994 and 1999 data sets contain 14 and 35

patches that are being overgrown by spruce plantations or willow, respectively. These patches are in the final stages of succession still suitable for *M. diamina*. Field experts have observed only small numbers of butterflies in these patches and it is doubtful whether the species is actually able to reproduce in some of them (J. Heliölä and N. Wahlberg pers. comm.). Many of the late-successional patches happen to be among the largest patches in the *M. diamina* patch network. If these patches are included in IFM parameter estimation, parameter x is estimated to be 0.3, but when these patches are excluded from estimation $x = 1.08$ (Moilanen and Cabeza 2002). This is actually a question of patch definition; the obligate larval food plant still grows on some late-successional patches, but otherwise these patches are shady and apparently unsuitable for the butterfly. Including these almost always unoccupied patches into parameter estimation produces an effect that is analogous to having large patches with false zeros; deterministic extinction caused by patch quality degradation is confused with the stochastic extinction assumed by the IFM. Given enough information it would be possible to correct patch areas by patch quality and to arrive to an effective patch area, which would be smaller than geographic area for bad-quality patches. This kind of correction would be perfectly natural as patch area in the metapopulation model is often only an easily measurable surrogate for population size.

The best quality data set that I have worked with is one on the American pika (Moilanen et al. 1998). In this data set patches are mine tailings, which clearly stand out from the surrounding desert landscape. There is little patch quality variation, patch size is accurately determined by measuring the perimeter of the heaps of rock (pikas establish territories in the margins of tailings) and territory occupancy is extremely easy to determine (Smith 1974). The only known source of error was a couple of patches which could not be surveyed because they were located on private land. The absence of these patches from the data set was reflected in an unexpectedly high observed occupancy for patches that were near the missing patches (Moilanen et al. 1998).

In summary, errors in metapopulation data can induce large biases into metapopulation model parameter estimates and subsequent model predictions. Understanding these biases may help to avoid errors in the execution of an empirical metapopulation study. Especially maintaining a low proportion of false zeros, and estimating the rate of error, should be regarded as important elements in designing empirical metapopulation studies. The data quality issue is most important when the parameterized metapopulation model is used to produce predictions that will be used as a basis for conservation decisions in landscapes other than the one used for parameter estimation.

References

ter Braak, C. J. F., Hanski, I. and Verboom, J. 1998. The incidence function approach to the modeling of metapopulation dynamics. – In: Bascompte, J. and Solé, R. V. (eds), Modeling spatiotemporal dynamics in ecology. Springer-Verlag, pp. 167–188.

Brown, J. H. and Kodric-Brown, A. 1977. Turnover rates in insular biogeography: effect of immigration on extinction. – Ecology 58: 445–449.

Burgman, M. A., Ferson, S. and Akçakaya, H. R. 1993. In: Risk assessment in conservation biology. – Chapman & Hall, pp. 169–217.

Conroy, M. J., Cohen, Y., James, F. C. et al. 1995. Parameter estimation, reliability and model improvement for spatially explicit models of animal populations. – Ecol. Appl. 5: 17–19.

Hanski, I. 1994. A practical model of metapopulation dynamics. – J. Anim. Ecol. 63: 151–162.

Hanski, I. 1997. Predictive and practical metapopulation models: the incidence function approach. – In: Tilman, D. and Kareiva, P. (eds), Spatial ecology. The role of space in population dynamics and interspecific interactions. Princeton Univ. Press, pp. 21–45.

Hanski, I. 1998a. Connecting the parameters of local extinction and metapopulation dynamics. – Oikos 83: 390–396.

Hanski, I. 1998b. Metapopulation dynamics. – Nature 396: 41–49.

Hanski, I. 1999. Metapopulation ecology. – Oxford Univ. Press.

Moilanen, A. 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. – Ecology 80: 1031–1043.

Moilanen, A. 2000. The equilibrium assumption in estimating the parameters of metapopulation models. – J. Anim. Ecol. 69: 143–153.

Moilanen, A. and Cabeza, M. 2002. Single-species dynamic site selection. – Ecol. Appl. (in press).

Moilanen, A., Smith, A. T. and Hanski, I. 1998. Long-term dynamics in a metapopulation of the American pika. – Am. Nat. 152: 530–542.

Ruckelhaus, M., Hartway, C. and Kareiva, P. 1997. Assessing the data requirements of spatially explicit dispersal models. – Conserv. Biol. 11: 1298–1306.

Shaw, M. W. 1994a. Modeling stochastic processes in plant pathology. – Annu. Rev. Phytopathol. 32: 523–544.

Shaw, M. W. 1994b. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. – Proc. R. Soc. Lond. B 259: 243–248.

Sjögren Gulve, P. and Ray, C. 1996. Large scale forestry extirpates the pool frog: Using logistic regression to model metapopulation dynamics. – In: McCullough, D. R. (ed.), Metapopulations and wildlife conservation and management. Island Press, pp. 111–137.

Smith, A. T. 1974. The distribution and dispersal of pikas: consequences of insular population structure. – Ecology 55: 1112–1119.

Steinberg, E. K and Kareiva, P. 1997. Challenges and opportunities for empirical evaluation of "Spatial Theory. – In: Tilman, D. and Kareiva, P. (eds), Spatial ecology. The role of space in population dynamics and interspecific interactions. Princeton Univ. Press, pp. 318–332.

Turner, M. G., Arthaud, G. J., Engstrom, R. T. et al. 1995. Usefulness of spatially explicit population models in land management. – Ecol. Appl. 5: 12–16.

Wahlberg, N. 1997. Suomen uhanalaisia eläimiä: Tummaverkkoperhonen (Endangered animals in Finland: the Glanville fritillary, in Finnish). – Oy Edita Ab, Helsinki, Finland.

Wahlberg, N., Moilanen, A. and Hanski, I. 1996. Predicting the occurrence of species in fragmented landscapes. – Science 273: 1536–1538.