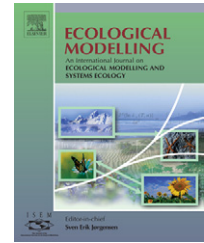


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

Parameterizing, evaluating and comparing metapopulation models with data from individual-based simulations

Frank M. Hilker^{a,*}, Martin Hirsch^b, Hans Joachim Poethke^c

^a Institute of Environmental Systems Research, Department of Mathematics and Computer Science, University of Osnabrück, 49069 Osnabrück, Germany

^b University of Groningen, CEES, Kercklaan 30, 9751NN Haren, The Netherlands

^c Field Station Fabrikschleichach, Julius-Maximilians-University of Würzburg, Glashüttenstr. 5, 96181 Rauhenbrach, Germany

ARTICLE INFO

Article history:

Published on line 5 September 2006

Keywords:

Model comparison
Parameterization
Metapopulation
Individual-based model
Incidence function model
Grid-based

ABSTRACT

Due to the lack of sufficient data and appropriate ecological information parameterizing predictive population dynamical models usually is a difficult task. The approach proposed in this study is meant to overcome this problem by using detailed individual-based simulations to generate artificial data. With short-term data samples, the models to be investigated can be parameterized and their predictions be compared. The flexibility of individual-based simulations as experimental tools also facilitates the evaluation and comparison of different (aggregated) model types. The presented approach is a step towards unifying models of different complexity. As an example we applied it to two metapopulation models of insect species in a highly fragmented landscape: the well-known incidence function model with a patch-based representation of space and a grid-based analogue. The models are tested with respect to their data requirement and recommendations for a better data sampling are derived.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Mathematical and computer-based models have become essential tools in understanding spatial population dynamics (Tilman and Kareiva, 1997; Bascompte and Solé, 1998; Czárán, 1998; Turchin, 1998). In the context of population viability analysis (Boyce, 1992), using highly specific model types such as spatially explicit or individual-based models (IBMs) (DeAngelis and Gross, 1992; Grimm, 1999; Grimm and Railsback, 2005) can be cumbersome and requires detailed biological information that is often lacking or difficult to obtain. On the other hand, more aggregated models of population dynamics in

fragmented habitats such as structured metapopulation models (e.g., Gyllenberg et al., 1997; Casagrandi and Gatto, 1999; Hanski and Ovaskainen, 2000; Ovaskainen and Hanski, 2001; Frank and Wissel, 2002) or the classic metapopulation model (Levins, 1969) neglect certain aspects which might be of special interest.

Spatially realistic metapopulation models (Hanski, 1999, 2001) seem to be promising tools, since they take spatial structure into account but still are analytically tractable. Within spatially realistic metapopulation theory, stochastic patch occupancy models (SPOMs) (Gosselin, 1998; Moilanen, 1999; Sjögren-Gulve and Hanski, 2000; Moilanen, 2004) are of

* Corresponding author. Present address: University of Alberta, Centre for Mathematical Biology, Department of Mathematical & Statistical Sciences, Edmonton Alberta T6G 2G1, Canada.

E-mail address: fhilker@uos.de (F.M. Hilker).

0304-3800/\$ – see front matter © 2006 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2005.06.011

special interest, because they simply classify habitat patches as either occupied or empty. They thus can be parameterized with patterns of presence–absence data, and are generally parameter-sparse. In this study, we investigate the incidence function model (IFM) of Hanski (1994), which has become a standard tool in quantitative metapopulation studies and has been applied to a broad range of animal species (Hanski, 2001, and references therein). However, there is some scepticism regarding the assumptions of quasi-equilibrium and environmental stochasticity or just the simplicity (Harrison, 1991, 1994; Thomas, 1994; Poethke et al., 1996; ter Braak et al., 1998; Gosselin, 1999; Moilanen, 2000).

Though the IFM in principle is easy to parameterize, in practice there are often situations where data are missing. The aim of this paper is to outline a conceptual approach how this lack of sufficient data can be overcome. The idea is as follows: instead of relying on rare real-world data, extensive artificial data-sets are generated by individual-based simulations. With samples from these artificial data, the aggregated model to be investigated can be parameterized. Moreover, the validation, evaluation and comparison of models is possible. The model predictions can be validated by checking their consistency with the underlying IBM-dynamics. IBMs allow to simulate various experimental settings (cf. Peck, 2004). Thus, the model can be evaluated regarding certain questions. In this paper, the data requirement of the IFM shall be investigated. Furthermore, different model types can be compared.

This shall exemplarily be demonstrated by means of a grid-based analogue of the IFM, which has been proposed by Settele (1998). There are two reasons to study this model. First, there is an increasing need of grid-based model types since data on species distributions are often available in a grid-based format, which additionally can easily be handled, for example with Geographic Information Systems. Second, the underpinning of the grid-based approach is questionable. By projecting the landscape onto a grid, populations inhabiting an area greater than a single cell are subdivided, thus relaxing the assumption of panmixia for local populations. Or, contrariwise, two or more small habitats might be subsumed in one cell.

The framework of model parameterization, evaluation and comparison is closely related with an approach suggested by Hanski (1999), which he motivates as follows (p. 76): “In practice, unfortunately, the problem often is that there are not enough data to test model predictions rigorously. An alternative approach that has not been sufficiently explored is to generate simulated data with some sufficiently complex model, to obtain a realistic sample from the simulated data, and to arrange contests among competing models based on such data”.

The outline of this paper is as follows. In the next section, the three models used in this study are described, i.e., the IBM, the incidence function model (henceforth referred to as patch-matrix model, PMM) and the grid-based model (GBM). The IBM exemplarily simulates various grasshopper species (but could be applied to any other similar insect species) in a highly fragmented landscape. Both the PMM and the GBM are parameterized with snapshot data which are short-term samples of the IBM. The model output are

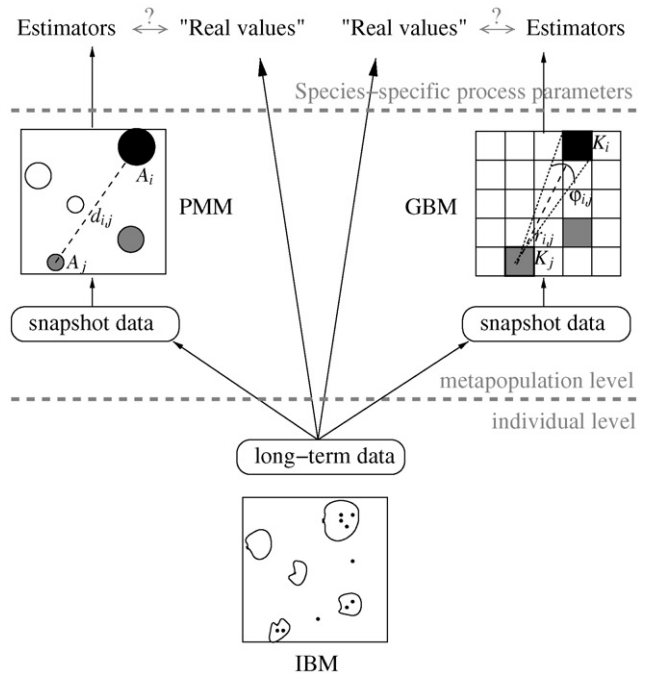


Fig. 1 – Approach to (i) parameterize the patch-matrix model (PMM) and the grid-based model (GBM) with snapshot samples from artificial long-term data generated by an individual-based model (IBM), (ii) evaluate the accuracy of the parameter estimators and (iii) compare the PMM and GBM themselves.

estimators of the species-specific process parameters, which shall be briefly referred to as *estimators*, describing the dispersal, colonization and survival abilities. On the other hand, the *real values* of the species-specific process parameters are directly extracted from the long-term IBM-data. This facilitates the evaluation and validation of the predicted estimators, cf. Fig. 1. The PMM and GBM are compared by assessing the accuracy with which their predictions match the real values. The more specific questions in mind are (i) whether increasing the amount of input data improves the model accuracy and (ii) whether the grid-based approach is suitable at all. Finally, the results are discussed and related to similar work.

2. Model descriptions

2.1. Individual-based model (IBM)

The IBM simulates a generic grasshopper species with non-overlapping generations. Three different species will be simulated in varying habitat landscapes. The details of these scenarios will be described in Section 3. Here, the general assumptions of the IBM are given.

The time step used in the model is one day. Fig. 2 shows the life cycle of an individual and the influence of both demographic and environmental stochasticity. Three life stages are distinguished: eggs, nymphs and adults. The hatching of eggs occurs stochastically between the middle and the end of May.

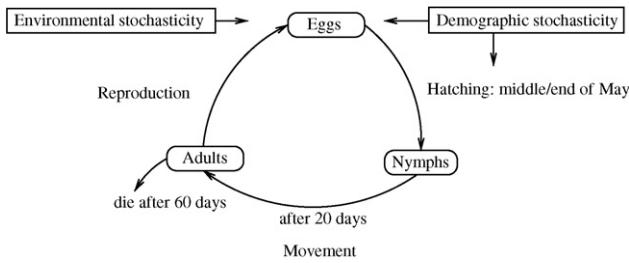


Fig. 2 – Life cycle of the simulated species.

A nymph passes to the adult stage after 20 days. The adults die after another 60 days.

Two adults mate if they are in the vicinity of *mateRadius* of each other. After mating, the female creates a number of eggs that is assumed to be Poisson-distributed with a fixed mean $\lambda = 10$. Ovipositing an egg requires an appropriate site, e.g., a soil gap or a blade of grass. A female detects all ovipositing sites in a distance of *nestRadius*. Once one egg is oviposited, the site is declared as ‘depleted’ until it becomes available again in the next year.

The movement of the nymphs and adults is modelled in the following way: all individuals are characterized by their actual position in space and by their movement direction. The spatial coordinates of consecutive days are calculated from an individual’s movement distance and turning angle which are drawn (daily) from uniform distributions (cf. Kindvall, 1999; Hein et al., 2004). There are two kinds of movement: when the individual is within a habitat patch, the movement distances are short and the turning angles vary widely, whereas in the matrix the movement is characterized by longer distances and few turns.

Environmental fluctuations are modelled by local catastrophes. They may occur each year up to three times with a probability of 0.67 each, and affect a circular area of up to one third of the total landscape. In this area, each egg is destroyed with a probability of 0.67 and the ovipositing sites are declared as non-available. The catastrophes may overlap, thus increasing the local extinction risk. Other combinations of the number of catastrophes, their area and their harm are possible, but this choice turned out to obey a good mixture between locally correlated environmental fluctuations and global stochasticity (due to overlapping local catastrophes). Other values may result in frequent metapopulation extinction or panmictic behaviour.

At the beginning of a simulation, 90,000 individuals with a 1:1 sex-ratio are generated and placed randomly in the landscape. The ovipositing sites are randomly distributed in the habitat patches, with 400 sites per ha of habitat.

Note that due to individuals having difficulties in finding mating partners at low densities an Allee effect can appear in the IBM. This effect is especially pronounced during colonization of an empty patch.

The basic “ingredients” of the IBM are extinction of subpopulations in local habitats (due to demographic stochasticity, catastrophes, Allee effect), exchange of individuals between habitats (movement) and recolonization of habitat patches (environmental stochasticity). During the creation of the IBM,

some of these mechanisms were varied in their intensity (or even skipped), but it turned out that all of them are necessary to generate persistent metapopulation dynamics, cf. also the first results in Section 3.

2.2. Patch-matrix model (PMM)

The PMM was described in detail by Hanski (1994, 1999). Belonging to the class of SPOMs, it is based on presence-absence data of a species in a set of habitats. These snapshot data are collected over a single or (better) several generations and assumed to represent the quasi-equilibrium of metapopulation dynamics. The modelling objective is to fit the incidence function to the observed snapshot data, thus obtaining estimators of the species-specific process parameters. Once these parameters are known, the model can be used to predict habitat-specific colonization and extinction probabilities for a particular habitat configuration. In this way occupancies, transient dynamics and regional population persistence can be predicted.

If habitat *i* is extinct (respectively occupied), it has the probability C_i (respectively E_i) of becoming recolonized (respectively extinct) at the next time step. These transitions are assumed to occur at random for each habitat. The probability that habitat *i* will be occupied tends toward the stationary probability

$$J_i = \frac{C_i}{C_i + E_i(1 - C_i)}, \tag{1}$$

which assumes a quasi-steady state of metapopulation dynamics conditional on metapopulation persistence. J_i is usually known as incidence. In Eq. (1), the rescue effect (Brown and Kodric-Brown, 1977) is included. Mathematically, patch occupancy models are time homogeneous, discrete time first order finite state Markov chains.

In the PMM, the extinction probability E_i is assumed to vary with the patch area A_i (in ha):

$$E_i = \min\{e_0 A_i^{-x}, 1\}, \tag{2}$$

where e_0 and x are extinction parameters. Next, the colonization probability C_i is approximated by the number of immigrants $M_i = \sum_{j \neq i} M_{ij}$ arriving at patch *i*:

$$C_i = \frac{M_i^2}{M_i^2 + y^2}, \tag{3}$$

where y' is a colonization parameter and

$$M_{ij} = p_j A_j \exp(-\alpha d_{ij}) \tag{4}$$

describes the number of immigrants from patch *j* to *i*. d_{ij} is the distance between patches *i* and *j* (in km), p_j gives the observed incidence (relative frequency of patch occupancy) of the patches around patch *i*, and α (km^{-1}) is a migration parameter.

Finally, one can reduce the number of parameters by setting $e' = e_0 y'^2$ and incorporate C_i and E_i in Eq. (1). Note that only patches with $A_i > A_0 := e_0^{1/x}$ are considered due to the minimum-operator in the extinction probability. A_0 is the critical patch area, below which the extinction probability E_i equals unity.

The initially unknown set of species-specific process parameters $\theta = (\alpha, e', x)$ is estimated by fitting Eq. (1) to the snapshot data. Using maximum pseudo-likelihood regression, the difference between the snapshot data p_i (approximating the quasi-steady state of the metapopulation) and the model-predicted incidences J_i is minimized. In the pseudo-likelihood function, a binomial distribution of the species' occurrences is assumed. Dealing with an optimization problem, the permutation term can be neglected and the likelihood be log-transformed, thus yielding

$$l(\theta) = \sum_i (p_i \log(J_i) + (1 - p_i) \log(1 - J_i)).$$

For maximization of this function, the simulated annealing algorithm is used, because it is able to escape from local optima in the search space and find global solutions (Kirkpatrick et al., 1983; Bounds, 1987; Moilanen et al., 1998). The PMM-parameters e_0 and y' can be separated from e' by defining A_0 as the area of the smallest occupied habitat patch ($e_0 = A_0^x, y' = \sqrt{e'/e_0}$).

2.3. Grid-based model (GBM)

The GBM has been suggested by Settele (1998). It resembles the PMM in being a SPOM based on a regression model. Space is represented by a grid, the cells of which may either be occupied by local populations or not. Since all cells are equally sized, the carrying capacity cannot be approximated by the area as in the PMM. Instead, the extinction probability is described by

$$E_i = \exp(-\kappa K_i), \tag{5}$$

where κ is an extinction parameter and K_i is a measure for the carrying capacity. K_i is set to the relative frequency p_i with which the cell is occupied in the snapshot data. In the case that a cell is always unoccupied and one can exclude that it is hostile to the species, one assigns the minimum capacity of all cells which have been occupied at least once.

The colonization probability is along the line of the PMM

$$C_i = \frac{M_i^2}{M_i^2 + \mu^2}, \tag{6}$$

with μ being a colonization parameter. The mean number of immigrants is approximated by $M_i = \sum_{j \neq i} M_{ij}$ with

$$M_{ij} = p_j K_j \exp(-\rho r_{ij}) \varphi_{ij}. \tag{7}$$

The term $p_j K_j$ is a measure for the population abundance in cell j . The fraction of individuals dispersing the Euclidean distance r_{ij} (km) between the source cell j and the target cell i is determined by the migration parameter ρ (km^{-1}). $\varphi_{ij} = \arctan(D/(2r_{ij}))/\pi$ is the maximum angle of a circle-segment from the midpoint of the source cell to the ends of the target cell, cf. Fig. 1. D is the cell length and throughout this paper set to 100 m, corresponding to the size of the smallest habitat. The set of species-specific parameters $\theta = (\rho, \kappa, \mu)$ is obtained analogously to the PMM.

2.4. Extraction of real values

The output of the PMM and GBM are estimators of the species-specific process parameters which are compared with the “real” values. Those can directly be extracted from the IBM-data, in a way which is described in this subsection.

The idea is as follows: the mechanistic equations of the PMM and GBM are simply fitted to the recorded IBM-data. In other words, the migration parameters α and ρ are obtained in nonlinear least-square fits of Eqs. (4) and (7) against d_{ij} and r_{ij} , respectively, to the number of immigrants M_{ij} , the colonization parameters y' and μ in fitting Eqs. (3) and (6) against the respective M_i to the colonization probabilities as well as the extinction parameters e_0, x and κ in fitting Eqs. (2) and (5) against A_i and K_i , respectively, to the extinction probabilities. Therefore, obviously (i) the number of immigrants as well as (ii) the extinction and colonization probabilities are needed.

The number of immigrants is approximated by tracking in the IBM all female individuals which oviposit their first egg in a different habitat than the one in which they mated.

The extinction and colonization probabilities are obtained by means of a simple maximum-likelihood estimation. Let N_i^{kl} denote the number of state transitions of patch (or grid cell, respectively) i with state being either 1 = occupied or 0 = empty. Then a likelihood function for the recorded transitions can be formulated:

$$P_i = (1 - C_i)^{N_i^{00}} C_i^{N_i^{01}} (E_i - E_i C_i)^{N_i^{10}} (1 - E_i + E_i C_i)^{N_i^{11}}$$

Now, C_i and E_i are approximated by maximizing the log-transformed P_i , which has been done with the Fletcher-Reeves conjugate gradient algorithm (Ueberhuber, 1997). Note that the case $N^{01} = N^{10} = N^{11} = 0$ is excluded, as with these values one of the partial derivatives would vanish. When fitting the number of immigrants, one has to be careful since the right-hand sides of Eqs. (4) and (7) might be much smaller. Here, they have been scaled by multiplying with the value of maximum observed immigrants over all patches. Alternatively, one could use a second fitting parameter.

3. Results

The IBM has been parameterized for three different grasshopper species mimicking low mobility (species LM), intermediate mobility (IM) and high mobility (HM). The respective parameter values are given in Table 1.

Table 1 – IBM-parameter values for the species with low, intermediate and high mobility (LM, IM and HM, respectively)

Species	LM	IM	HM
nestRadius (m)	4	5	6
mateRadius (m)	8	8	10
patchDist (m)	10	15	20
matrixDist (m)	70	140	210
patchTurn and matrixTurn have constantly been set to 1.57 and 0.785, respectively.			

Table 2 – Landscape characteristics of the habitat configurations

Habitat configuration	S	N ⁻	N ⁺	A ⁻	A ⁺
Number of patches	100	84	115	100	100
Mean patch area (ha)	1.48	1.48	1.45	1.29	1.65
Standard deviation of patch area (ha)	1.60	1.65	1.55	1.39	1.80
Fraction of habitat (%)	2.05	1.72	2.30	1.78	2.28
Minimum patch area (ha)	0.2	0.2	0.2	0.2	0.2
Maximum patch area (ha)	6.76	6.76	6.76	6.2	7.88

S: standard, N⁻/N⁺ and A⁻/A⁺: decreased/increased number of patches and mean area as well as standard deviation, respectively.

Parameter values have been chosen to guarantee metapopulation dynamics, i.e., they have been tuned by means of pattern-oriented modelling (Grimm et al., 1996; Wiegand et al., 2003) to let the model show typical metapopulation-like behaviour indicated by a significant but not too high turnover rate and a quasi-equilibrium of the fraction of occupied habitats. For the species LM, IM and HM, the turnover rates are 0.015, 0.08 and 0.112, the mean fractions of occupied habitats are 0.32, 0.62 and 0.75 and the mean of the standard deviations from these mean occupancies are 0.034, 0.042 and 0.033, respectively.

The standard habitat configuration (S) exhibits similar features (number of patches, minimum/maximim, mean and standard deviation of patch area, ratio habitat to matrix) as an exemplarily chosen habitat network of the bush cricket

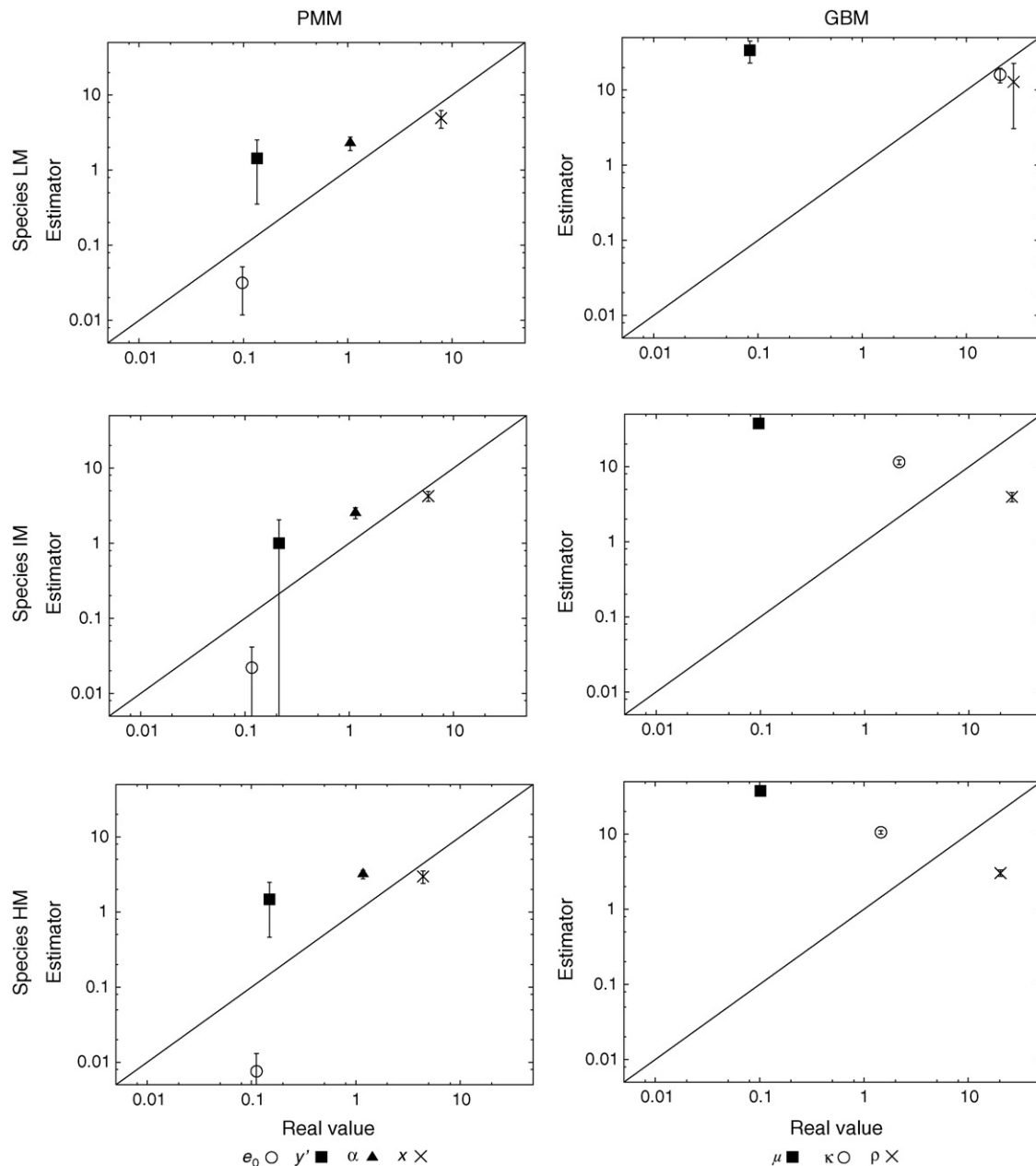


Fig. 3 – Mean estimators based on two consecutive snapshot years. The dotted vertical lines indicate the standard variance. Details are explained within the text.

Platycleis albopunctata in the Naturpark Hassberge in Northern Bavaria, Germany. The real-world landscape comprises 7.5 km × 12.92 km. The fraction of habitat is 1.4%. For modelling purposes, a quadratic artificial landscape (8.5 km × 8.5 km) has been assumed. Four other configurations have been created by decreasing/increasing the number of patches (N^- and N^+ , respectively) as well as mean and standard deviation of patch area (A^- and A^+ , respectively), cf. Table 2.

Each species has been simulated in all five habitat configurations with ten replicates. In each replicate, the spatial arrangement of the patches has been chosen stochastically. In the following, we shall focus on the species-specific process parameters. Since they are assumed to be habitat-independent, they are averaged over all habitat configurations and replicates. The simulations have been run 200 years to let the metapopulation dynamics reach quasi-equilibrium.

Then, further 400 years are simulated in which the data are recorded and analyzed.

The impact of the amount of parameterization data is investigated in three steps. The PMM and GBM are parameterized (i) with two consecutive snapshot years, (ii) with five consecutive snapshot years and (iii) with five consecutive snapshot years as well as the given migration parameter. Two and five snapshot years correspond to typical field studies (cf. Moilanen, 2000). In order to be independent of the particular years, the samples have been collected five times in each replicate beginning in different years. The migration parameter can be determined by independent additional data from, e.g., mark-recapture or genetic studies (Hanski, 1999, 2001 and references therein). The real values are extracted with the methods described in Subsection 2.4 from the full IBM-data set comprising 400 years. In the following Figs. 3–5, the result-

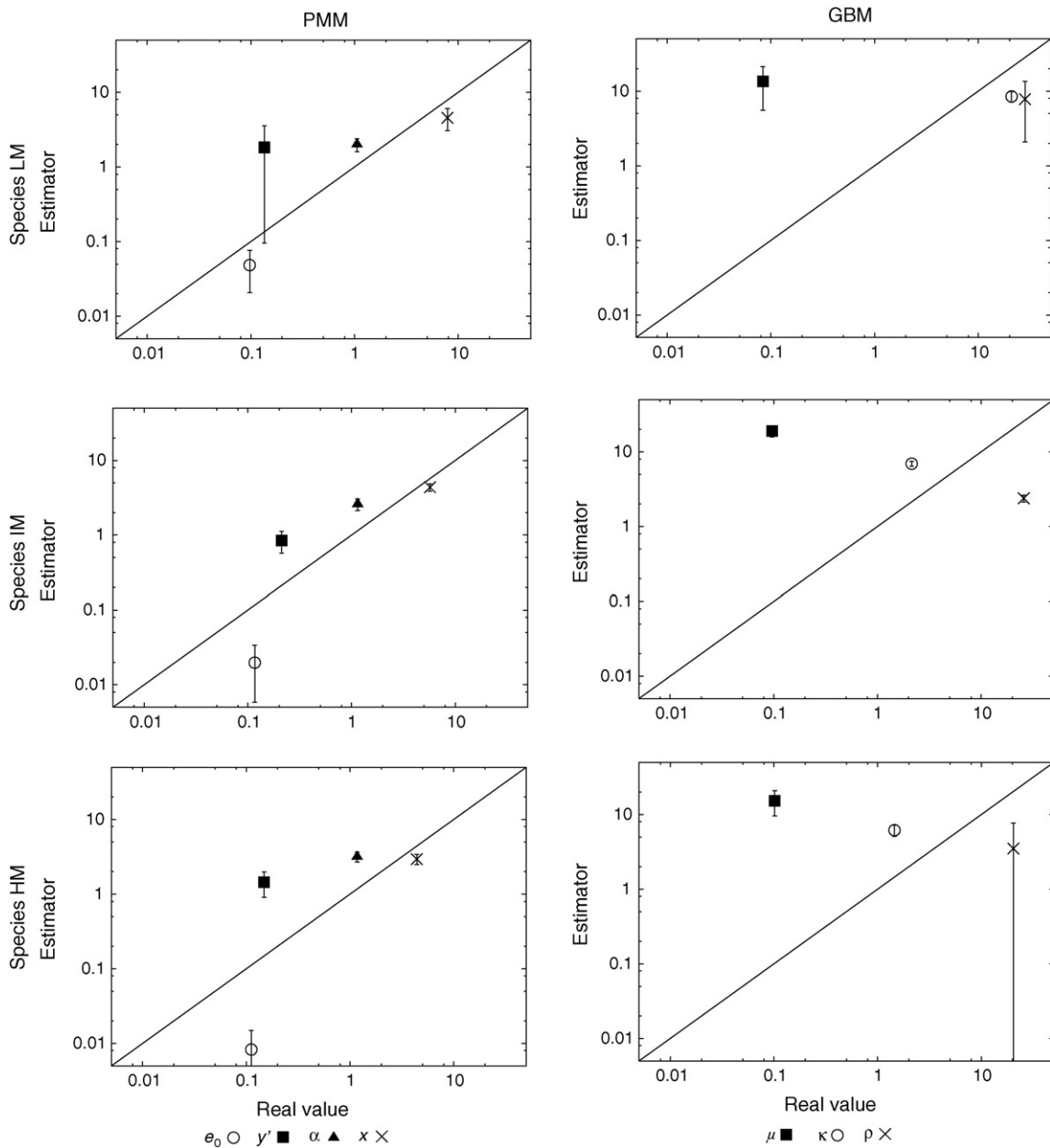


Fig. 4 – Estimators based on five consecutive snapshot years.

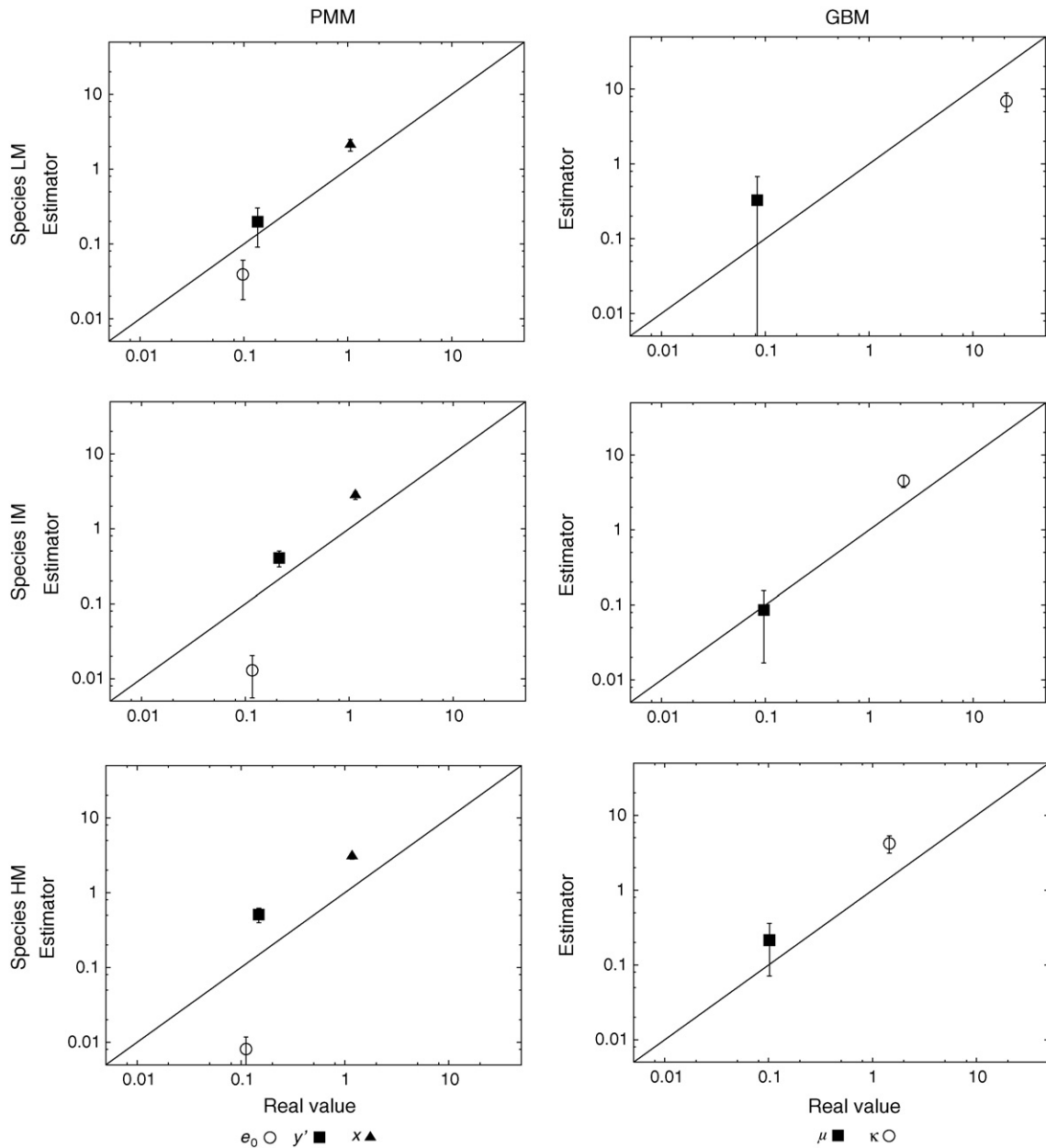


Fig. 5 – Estimators based on given migration parameters and five consecutive snapshot years.

ing mean estimators (averaged over all habitat configurations, replicates and collected snapshot samples) as well as their standard deviation are plotted against the mean real values.

First, Fig. 3 shows the estimators obtained from two snapshot years plotted against the real values. Regarding the PMM, the migration parameter α is estimated well, and the estimator of x coincides well with its real value, too. There are substantial biases, however, for y' and e_0 . Regarding the GBM, the migration parameter ρ and the extinction parameter κ are well estimated for species LM, whereas there are under- and overestimations, respectively, for species IM and HM. The estimators of the colonization parameter μ are far too large—for all three species more than two and a half orders of magnitude. The variance, however, is small. Thus, a systematic error can be suspected.

Second, both the PMM and GBM are parameterized with five snapshot years. The results are shown in Fig. 4. The accuracy of the PMM-estimators remains nearly the same. While the variance decreases for species IM and HM, there is an increased scattering for species LM. The estimators of μ in the GBM still deviate more than two orders of magnitude. Regarding these biases, the minor changes in the other estimators do not seem to be of relevance.

Third, the real values of the migration parameters are assumed to be given. Thus, the search space for the estimation problem reduces from three to two dimensions for both the PMM and the GBM (recall that two of the PMM parameters are combined during the estimation problem). The results are shown in Fig. 5. Now, the estimators of both x and y' are close to identity with the real value. The accuracy of the e_0 -estimators

for species LM is of a similar quality, but for species IM and HM there are deviations of one order of magnitude. In the GBM, κ is still estimated well. The enormous biases of μ vanish. The accuracy is quite good, especially for species IM and HM. There is a substantial variance in the μ -estimators of species LM, while for all other parameters and species the estimators are very precise.

4. Discussion and conclusions

Testing the quality of model predictions usually turns out to be a problem, because little or even no data are available. Highly specific models can be used to generate simulated data, or, in other words, to substitute missing “real-world” data. This allows not only the parameterization of single models. Moreover, these models can be investigated regarding certain aspects, for example the data requirement, or they can be compared with other modelling approaches.

In this study, an IBM of three grasshopper species in five different landscapes has been used to generate extensive long-term data sets. With several sampled snapshot years, two different SPOMs have been parameterized, which differ in their representation of space: once the classic patch-matrix model and once a grid-based approach. The parameterization has been done (i) with two snapshot years, (ii) with five snapshot years and (iii) with additionally given migration parameter, which can be estimated from independent data. The species-specific process parameter estimators have been compared with the real values which have been directly determined from the full amount of long-term IBM-data.

Surprisingly, the accuracy of the PMM predictions is not improved when the parameterization is done with five instead of two snapshot years. In both cases, there are misestimations of one order of magnitude. This bias is reduced, if additionally the migration parameter is known, but there still remain severe underestimations at least for two species.

The predictions of the GBM partly deviate more than two orders of magnitude. Only if the model is parameterized with given migration parameter, the GBM proves to be relatively accurate. In fact, the biases are less prominent than in the PMM, but there is some more variance in the predictions.

As a consequence, one cannot blindly rely on the model predictions. Each one should be taken very cautiously. While parameters can be estimated well for some species, this does not hold for others. As a kind of rule of thumb, however, the results suggest to put emphasis on determining the migration parameters α and ρ from additional, independent data rather than collecting more snapshot years.

Furthermore, the presented results indicate another possibility to improve the estimation. Since the estimated migration parameters match the real values well (in both cases with two and five snapshot years), their estimators can be used as the given values in another, repeated estimation process. Thus, the search space can again be reduced to two dimensions. In fact, there is empirical evidence that the migration parameter can be estimated well by patch occupancy models (Hanski et al., 1996; Appelt and Poethke, 1997; Moilanen et al., 1998).

While the amount of snapshot years has nearly no effect on the PMM, the large deviations of the GBM decrease at least to some part (but still are enormous). This can be explained as follows. With two snapshot years, the relative frequency of occupancy may either be 0.0, 0.5 or 1.0. Remember the usage of a minimum carrying capacity, which will be in this case at least ≥ 0.5 . Hence, there is an implicit tendency to homogenization of space, because nearly all cells are possible habitats. Using five snapshot years, instead, the minimum carrying capacity can be as low as 0.2. Alternatively, some other appropriate value could be used. Settele (1998) originally proposed only to consider cells which may be potential habitats. Moreover, he suggested to approximate K_i by the mean number of observed individuals. In this study, the relative occupancy frequency has been used, in order to ensure the comparability with the PMM, since this modelling approach uses only presence-absence data, too. However, despite the scepticism regarding the assumption of panmixia, the GBM might be an efficient tool in certain circumstances and deserves further exploration. The indicated modifications, though requiring in practice much more field work, could resolve the essential deviations. Future work will also have to consider different cell lengths.

It should be mentioned that a couple of assumptions concerning the IBM are debatable and could be refined to be even more ‘realistic’. For instance, the way how dispersal (distance and number of immigrants) is tracked might be a source of bias and could explain the deviations especially in the colonization parameters. Nevertheless, the migration parameters can be estimated well. Moreover, grasshoppers oviposit their eggs in clutches and not as single eggs, what might increase the overall variability in metapopulation dynamics. Also the way how environmental stochasticity is incorporated or the rather fixed population growth limitation by ovipositing sites could be refined. However, we want to emphasize that the main point is that the IBM macroscopically exhibits ‘typical’ metapopulation dynamics and that the current model assumptions seem to be sufficient for this more conceptual approach.

What truly matters for practical purposes are the predicted incidences which are based on the estimators. A possible extension of this study is the comparison of observed occupancies in the IBM with the predicted ones, once based on the real values (to validate the real values themselves) and once based on the estimators. This way of comparison has been done by (Hanski, 1999, pp. 95). He generated artificial data using a coupled map lattice model. The parameters were chosen to model the Glanville fritillary butterfly in a realistic patch network. With samples of these simulated data, on the one hand the IFM and on the other hand a state transition model was parameterized. With the estimated parameters, occupancies were predicted in the current, in another and in a perturbed habitat network and finally compared with the simulated occupancies. This study is restricted to the parameter estimators, because they play the key role in patch occupancy models. Extrapolating them to other landscapes (Hanski et al., 1996) or similar species (Wahlberg et al., 1996) presumes that they are validated. Deviations of up to one order of magnitude, however, suggest that the mechanistic underpinning of the model equations are violated. Thus, concentrating on the

process parameters instead of the incidences is essential for evaluating patch occupancy models.

There is some more related work. Moilanen (2000) simulated data by iterating a patch network of the Glanville fritillary butterfly with a logistic regression model. With samples of these data, two variants of the IFM and again a logistic regression model were parameterized. The aim was to investigate the impact of the metapopulation quasi-equilibrium assumption on the predicted patch occupancies. Other comparisons of model types have been conducted by Jepsen et al. (2005) and Hokit et al. (2001). Jepsen et al. (2005) studied the effect of corridors and landscape heterogeneity on the dispersal probabilities in an individual-based model, a movement model and a variant of the IFM. Therefore, they first ran an individual-based model of the field vole. By means of simplification, they then parameterized the other two models and compared the model predictions. The study by Hokit et al. (2001) aimed at understanding different modelling approaches regarding their assumptions and requirements for data. They parameterized both the IFM and a stage-based matrix model with real data samples of the Florida scrub lizard. The estimated parameters were used for predictions in perturbed habitat configurations as well as in artificial landscapes. Furthermore, the obtained parameter estimators were individually increased by 25%, thus testing the model sensitivity. Gosselin (1999) evaluated the IFM with respect to the equilibrium assumption using Monte Carlo simulations in artificial landscapes with parameter values for the Glanville fritillary butterfly. McCarthy et al. (2003) used data generated by a stochastic Ricker model and tested the reliability of model predictions in response to parameter changes.

The approach presented in this paper is more general. First, various species in different landscapes and replicates have been taken into account. Thus, the results are less specific. Second, the comparison is based on the real values of the process parameters and not on predictions of the same or just another model. Third, the model generating the artificial data is really independent of the 'simpler' (aggregated) models. For example, it does not produce by way of its construction the quasi-stability of the metapopulation. In fact, an IBM can be seen as the most general model possible. The emergence of metapopulation dynamics is just a result of individual interactions in a fragmented landscape. By this means, it is also possible to test the convergence of more complex (or structured) models to models on a more aggregated level or to relate IBM-parameters to emergent behaviour on metapopulation level (see, e.g., Adler and Nuernberger, 1994; Fahse et al., 1998; Wilson, 1998; Casagrandi and Gatto, 1999, 2002; Hanski and Ovaskainen, 2000; Keeling, 2002; Ovaskainen and Hanski, 2004; Revilla et al., 2004).

The use of an IBM also allows to investigate species whose biology is not known in detail by exploring a range of possible IBM-parameters. Though IBMs are hard to develop, to communicate and to analyze, they allow to incorporate a lot of biological detail and realism (Grimm et al., 1999). Furthermore, the impact of many other factors on the population or metapopulation level could be explored, e.g., continually varying habitat landscapes (instead of a binary distinction between habitat and non-habitat) or errors in the snapshot data (Berger et al., 1999; Moilanen, 2002)—just in the sense of

simulations as experiments (Peck, 2004). Generally, the modelling approach presented in this study can also be extended to other modelling types such as the classic Levins model or its variants (Ovaskainen, 2002). It thus provides a step towards a unifying framework for the parameterization, evaluation and comparison of different model types.

Acknowledgements

The authors thank Silke Hein and Thomas Hovestadt for helpful comments and discussions during earlier stages of this work.

REFERENCES

- Adler, F.R., Nuernberger, B., 1994. Persistence in patchy irregular landscapes. *Theor. Popul. Biol.* 45, 41–75.
- Appelt, M., Poethke, H.J., 1997. Metapopulation dynamics in a regional population of the bluewinged grasshopper (*Oedipoda caerulescens*; Linnaeus, 1758). *J. Insect Conserv.* 1, 205–214.
- Bascompte, J., Solé, R.V., 1998. *Modeling Spatiotemporal Dynamics in Ecology*. Springer-Verlag, Berlin.
- Berger, U., Wagner, G., Wolff, W.F., 1999. Virtual biologists observe virtual grasshoppers: an assessment of different mobility parameters for the analysis of movement patterns. *Ecol. Model.* 115, 119–127.
- Bounds, D.G., 1987. New optimization methods from physics and biology. *Nature* 329, 215–219.
- Boyce, M.S., 1992. Population viability analysis. *Annu. Rev. Ecol. Syst.* 23, 481–506.
- Brown, J.H., Kodric-Brown, A., 1977. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* 58, 445–449.
- Casagrandi, R., Gatto, M., 1999. A mesoscale approach to extinction risk in fragmented habitats. *Nature* 400, 560–562.
- Casagrandi, R., Gatto, M., 2002. Habitat destruction, environmental catastrophes, and metapopulation extinction. *Theor. Popul. Biol.* 61, 127–140.
- Czárán, T., 1998. *Spatiotemporal Models of Population and Community Dynamics*. Chapman & Hall, London.
- DeAngelis, D.L., Gross, L.J., 1992. *Individual-based Models and Approaches in Ecology. Populations, Communities and Ecosystems*. Chapman & Hall, New York.
- Fahse, L., Wissel, C., Grimm, V., 1998. Reconciling classical and individual-based approaches in theoretical population ecology: a protocol for extracting population parameters from individual-based models. *Am. Nat.* 152 (6), 838–852.
- Frank, K., Wissel, C., 2002. A formula for the mean lifetime of metapopulations in heterogeneous landscapes. *Am. Nat.* 159 (5), 530–552.
- Gosselin, F., 1998. Reconciling theoretical approaches to stochastic patch-occupancy metapopulation models. *Bull. Math. Biol.* 60, 955–971.
- Gosselin, F., 1999. Test of mathematical assumptions behind the 'incidence function' estimation process of metapopulations' dynamic parameters. *Math. Biosci.* 159, 21–32.
- Grimm, V., 1999. Ten years of individual-based modelling in ecology: what have we learned, and what could we learn in the future? *Ecol. Model.* 115, 129–148.

- Grimm, V., Frank, K., Jeltsch, F., Brandl, R., Uchmański, J., Wissel, C., 1996. Pattern-oriented modelling in population ecology. *Sci. Total Environ.* 183, 151–166.
- Grimm, V., Railsback, S.F., 2005. *Individual-based Modeling and Ecology*. Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton and Oxford.
- Grimm, V., Wyszomirski, T., Aikman, D., Uchmański, J., 1999. Individual-based modelling and ecological theory: synthesis of a workshop. *Ecol. Model.* 115, 275–282.
- Gyllenberg, M., Hanski, I., Hastings, A., 1997. Structured metapopulation models. In: Hanski, I., Gilpin, M.E. (Eds.), *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press, San Diego 93–122.
- Hanski, I., 1994. A practical model of metapopulation dynamics. *J. Anim. Ecol.* 63, 151–162.
- Hanski, I., 1999. *Metapopulation Ecology*. Oxford University Press, New York.
- Hanski, I., 2001. Spatially realistic theory of metapopulation ecology. *Naturwissenschaften* 88, 372–381.
- Hanski, I., Moilanen, A., Pakkala, T., Kuussaari, M., 1996. The quantitative incidence function model and persistence of an endangered butterfly metapopulation. *Conserv. Biol.* 10, 578–590.
- Hanski, I., Ovaskainen, O., 2000. The metapopulation capacity of a fragmented landscape. *Nature* 404, 755–758.
- Harrison, S., 1991. Local extinction in a metapopulation context: an empirical evaluation. *Biol. J. Linnean Soc.* 42, 73–88.
- Harrison, S., 1994. Metapopulations and conservations. In: Edwards, P.J., May, R.M., Webb, N.R. (Eds.), *Large-scale Ecology and Conservation Biology*. Blackwell, Oxford 111–128.
- Hein, S., Pfenning, B., Hovestadt, T., Poethke, H.-J., 2004. Patch density, movement pattern, and realised dispersal distances in a patch-matrix landscape—a simulation study. *Ecol. Model.* 174, 411–420.
- Hokit, D.G., Stith, B.M., Branch, L.C., 2001. Comparison of two types of metapopulation models in real and artificial landscapes. *Conserv. Biol.* 15 (4), 1102–1113.
- Jepsen, J.U., Baveco, J.M., Topping, C.J., Verboom, J., Vos, C.C., 2005. Evaluating the effect of corridors and landscape heterogeneity on dispersal probability: a comparison of three spatially explicit modelling approaches. *Ecol. Model.* 181, 445–459.
- Keeling, M.J., 2002. Using individual-based simulations to test the Levins metapopulation paradigm. *J. Anim. Ecol.* 71, 270–279.
- Kindvall, O., 1999. Dispersal in a metapopulation of the bush cricket, *Metroptera bicolor* (Orthoptera: Tettigoniidae). *J. Anim. Ecol.* 68, 172–185.
- Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220, 671–680.
- Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* 15, 237–240.
- McCarthy, M.A., Andelman, S.J., Possingham, H.P., 2003. Reliability of relative predictions in population viability analysis. *Conserv. Biol.* 17 (4), 982–989.
- Moilanen, A., 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. *Ecology* 80 (3), 1031–1043.
- Moilanen, A., 2000. The equilibrium assumption in estimating the parameters of metapopulation models. *J. Anim. Ecol.* 69, 143–153.
- Moilanen, A., 2002. Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos* 96 (3), 516–530.
- Moilanen, A., 2004. SPOMSIM: software for stochastic patch occupancy models of metapopulation dynamics. *Ecol. Model.* 179, 533–550.
- Moilanen, A., Smith, A.T., Hanski, I., 1998. Long-term dynamics in a metapopulation of American pika. *Am. Nat.* 152 (4), 530–542.
- Ovaskainen, O., 2002. The effective size of a metapopulation living in a heterogeneous patch network. *Am. Nat.* 160 (5), 612–628.
- Ovaskainen, O., Hanski, I., 2001. Spatially structured metapopulation models: global and local assessment of metapopulation capacity. *Theor. Popul. Biol.* 60, 281–302.
- Ovaskainen, O., Hanski, I., 2004. From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. *Am. Nat.* 164 (3), 364–377.
- Peck, S.L., 2004. Simulation as experiment: a philosophical reassessment for biological modelling. *Trends Ecol. Evol.* 19 (10), 530–534.
- Poethke, H.J., Gottschalk, E., Seitz, A., 1996. Gefährdungsgradanalyse einer räumlich strukturierten Population der Westlichen Beißschrecke (*Platycleis albopunctata*): Ein Beispiel für den Einsatz des Metapopulationskonzepts im Artenschutz. *Z. Ökologie Naturschutz* 5, 229–242.
- Revilla, E., Wiegand, T., Palomares, F., Ferreras, P., Delibes, M., 2004. Effects of matrix heterogeneity on animal dispersal: from individual behavior to metapopulation-level parameters. *Am. Nat.* 164 (5), E130–E153.
- Settele, J., 1998. *Metapopulationsanalyse auf Rasterdatenbasis*. Teubner, Leipzig, Stuttgart.
- Sjögren-Gulve, P., Hanski, I., 2000. Metapopulation viability analysis using occupancy models. *Ecol. Bull.* 48, 53–71.
- ter Braak, C.J.F., Hanski, I., Verboom, J., 1998. The incidence function approach to modeling of metapopulation dynamics. In: Bascompte, J., Solé, R.V. (Eds.), *Modeling Spatiotemporal Dynamics in Ecology*. Springer-Verlag, Cambridge 69–86.
- Thomas, C.D., 1994. Difficulties in deducing dynamics from static distributions. *Trends Ecol. Evol.* 9 (8), 300.
- Tilman, D., Kareiva, P., 1997. *Spatial Ecology. The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, NJ.
- Turchin, P., 1998. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer, Sunderland, MA.
- Ueberhuber, C.W., 1997. *Numerical Computation*, vol. 2. Springer-Verlag, Berlin.
- Wahlberg, N., Moilanen, A., Hanski, I., 1996. Predicting the occurrence of endangered species in fragmented landscapes. *Science* 273, 1536–1538.
- Wiegand, T., Jeltsch, F., Hanski, I., Wissel, C., 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory. *Oikos* 100, 209–222.
- Wilson, W.G., 1998. Resolving discrepancies between deterministic population models and individual-based simulations. *Am. Nat.* 151 (2), 116–134.