Jul 1st, 12:00 AM

How to Compare Different Conceptual Approaches to Metapopulation Modelling

Frank M. Hilker

Martin Hinsch

Hans Joachim Poethke

How to Compare Different Conceptual Approaches to Metapopulation Modelling

Frank M. Hilker[a]**, Martin Hinsch**[b] **and Hans Joachim Poethke**[b]

[a]*Institute of Environmental Systems Research, Department of Mathematics and Computer Science, University of Osnabrück, Germany, e-mail:* `fhilker@uos.de`

[b]*Ecological Research Station, Bavarian Julius-Maximilians-University of Würzburg, Germany*

Abstract: Models are essential tools in understanding population dynamics and deriving management measures in the context of population viability analysis. However, very often the question arises which type of model architecture is appropriate for a given situation. Mostly this situation is characterized by a shortage of data for model parameterization. In this study, an approach is presented to overcome this lack of real-world data by using the output of long-term simulation runs of specific individual-based models. Thus, it is possible to evaluate the quality of macroscopic model predictions. Furthermore, this setting allows to compare totally different types of metapopulation models. As an exemplary case study, this approach is applied to generic grasshopper species in highly fragmented habitat landscapes, assessing on the one hand the well-known incidence function model and on the other hand a grid-based approach. The results show that predictions of both models have substantial biases. Nonetheless, recommendations can be derived how to obtain more accurate model estimates. Finally, the patch-matrix model proves to be more adequate than the grid-based approach.

Keywords: Metapopulation models; incidence function model; patch-matrix model; grid-based model; individual-based model

1. INTRODUCTION

In the last decades, a variety of modelling approaches has been developed in order to understand population dynamics as well as to be able to derive appropriate management measures in the context of population viability analysis. These models differ in the degree in which they take space, time and state variables into account. For example, there are ordinary and partial differential equations, (integro-) difference and integrodifferential equations, cellular automata, coupled map lattices, interacting particle systems or individual-based models (e.g., see Czárán [1998] and references therein).

Hence, modelleres are often confronted with the question, which conceptual approach seems to be most appropriate for a certain problem. This paper is concerned with a comparison between the well-known "practical model of metapopulation dynamics" of Hanski [1994] and a grid-based approach proposed by Settele [1998]. Both are incidence function models (IFMs), which can relatively easily be parameterized with the species' occupancy data. They differ in the spatial representation of habitats

(cf. Figure 1). In the Hanski model, which shall be referred to as patch-matrix model (PMM), the species is assumed to inhabit circular patches of different sizes within a hostile matrix. In distinction to the spatial implicit model of Levins [1969] and to spatially explicit models, the PMM is called a spatially realistic model [Hanski, 1999]. The PMM has frequently been used in population viability analyses of endangered species (see Hanski [2001] and references therein). In Settele's approach, which shall be referred to as grid-based model (GBM), space is sub-divided into equally sized cells with different carrying capacities. Each cell is assumed to be a possible habitat which may be occupied by the species. Please note the differentiation between the general habitat, patch (PMM) and cell (GBM) throughout this paper.

There is an increasing demand for grid-based models, since data on the distribution of various species are often available in a grid-based format, as this can easily be handled (for example with Geographic Information Systems). Regarding Settele's approach, however, there is some severe scepticism about the
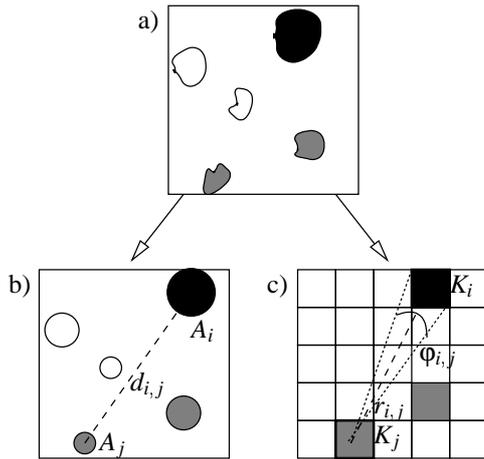
Figure 1. Representation of habitat configurations (a) in the patch matrix model (b) and in the grid-based approach (c). Habitats are coloured with grey values corresponding to their average occupancies.

biological realism in the underlying assumptions, as will be pointed out in the model description in the following section. The aim of this paper is to demonstrate the application of a general method for the comparison of conceptually different modelling approaches. Especially in conservation biology, this is of increasing interest, as generally little is known about the species under focus and often few quantitative data are available. This study makes use of an approach which has recently been presented by Hilker [2002], cf. Figure 2. As there is a lack of real-world data of sufficient resolution, an individual-based model (IBM) is used to simulate complex population dynamics of "virtual", generic species. The simulation runs generate extensive long-term data sets. On the one hand, the PMM and the GBM, which are both highly aggregated models, can now be parameterized with various short-term data samples. In this study, snapshot data of two or five consecutive years are used as it is typical for field campagains. On the other hand, the "real" parameter values describing the metapopulations dynamics can be extracted directly from the full amount of IBM-data (which consist in this study of 400 years). Thus, one obtains parameter *estimators* from the highly aggregated models (based on short-term snapshot data) and *"real" values* from the specific model (based on long-term data), which can be compared with each other, especially with respect to the (dis-)advantages of different space representation.

This paper is organized as follows. Firstly, all three models (PMM, GBM, IBM) are introduced. As a result of the PMM and GBM, the estimators of
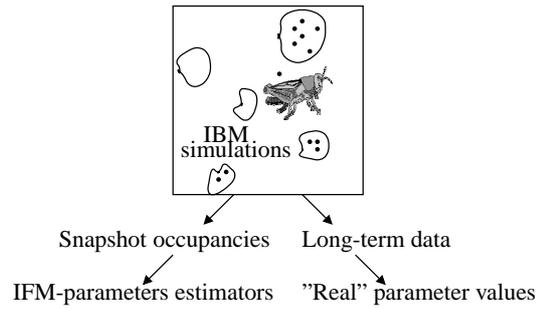
Figure 2. An IBM simulates a species' dynamics in fragmented habitats. From the available long-term data, the "real" parameter values can be extracted. With several snapshot data both the patch-matrix and the grid-based IFM are parameterized. Their parameter estimators can then be compared with the "real" values.

the metapopulation dynamic parameters are yielded (Subsection "Parameter estimators"). The "real" values are determined each from the full amount of IBM-data including dispersal events (Subsection "Extraction of 'real' values"). Please note, that all settings of the simulated species, habitat configuration and snapshot sampling are the same as in Hilker [2002]. Next, the results are given with a focus on the accuracy of the parameter estimators. Finally, potential reasons for the resulting deviations in the estimators are discussed.

2. INCIDENCE FUNCTION MODELS (IFMs)

IFMs rely on presence-absence data of a species in a set of habitats. In typical field campaigns, these occupancy data are collected over a single or (better) a few generations, which have not to be consecutive. Because of that, these data are often referred to as patch occupancy pattern or snapshot data. Henceforth, the latter notation will be used throughout this paper. The observed snapshot data are 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 metapopulation dynamic parameter estimators. Once these parameters are estimated, the IFMs can be used to predict habitat-specific colonization and extinction probabilites for a particular habitat configuration. Thus, occupancies, transient dynamics, and regional population persistence may be predicted.

If habitat i is extinct (respectively occupied), it has the colonization probability C_i (respectively extinction probability E_i) of becoming occupied (respectively extinct) at the next time step. These transi-

tions are assumed to occur at random for each habitat. The probability that habitat i will be occupied tends toward the stationary probability

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

which is called the incidence and assumes a quasi-steady state of metapopulation dynamics conditional on non-metapopulation extinction. In (1), the rescue effect is included. Mathematically, IFMs are time homogeneous, discrete time first order finite state Markov chains.

2.1 Patch-matrix model (PMM)

The PMM is described in detail by Hanski [1994, 1999], or see references therein. The extinction probability E_i is assumed to vary with the patch area A_i (in ha): $E_i = \min\{e_0 A_i^{-x}, 1\}$, where e_0 and x are extinction parameters. Next, the colonization probability C_i is approximated by the number of immigrants M_i arriving at patch i: $C_i = \frac{M_i^2}{M_i^2 + y^2}$, where y is a colonization parameter. M_i itself depends on the connectivity S_i through $S_i = \beta M_i = \beta \sum_{j \neq i} p_j A_j \exp(-\alpha d_{ij})$, with d_{ij} being the distance between patches i and j (in km), p_j the relative frequency of patch occupancy. β is assumed to equal unity and α is a migration parameter.

Finally, one can combine the parameters $y' = y\beta^{-1}$ and $e' = e_0 y'^2$ and then incorporate C_i and E_i in (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.

2.2 Grid-based model (GBM)

The GBM has been suggested by Settele [1998]. Space is represented by a grid, whose cells 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

$$E_i = \exp(-\kappa K_i) , \qquad (2)$$

where κ is an extinction parameter and K_i 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

$$C_i = \frac{M_i^2}{M_i^2 + \mu^2} , \qquad (3)$$

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

$$M_{ij} = p_j K_j \exp(-\rho r_{ij}) \varphi_{ij} . \qquad (4)$$

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} (in km) between the source cell j and the target cell i is determined by the migration parameter ρ. $\varphi_{ij} = \frac{1}{\pi} \arctan(\frac{D}{2r_{ij}})$ is the maximum angle of a circle-segment from the midpoint of the source patch to the ends of the target patch, cf. Figure 1. D is the cell length.

Principally, the GBM resembles the PMM in being a stochastic patch occupancy model based on a regression model. However, by dividing the landscape in a grid, local populations inhabiting an area greater than a single cell are also subdivided. Hence, the assumption of panmixia for local populations (patches) is relaxed. Or, contrariwise, two or more small habitats might be subsumed in one cell.

2.3 Parameter estimators

The PMM as well as the GBM are characterized by an initially unknown set of species-specific metapopulation dynamic parameters $\theta = (\alpha, e', x)$ or $\theta = (\rho, \kappa, \mu)$, respectively. These are obtained by fitting (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' occurences 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. Note, that 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 ($e_0 = A_0^x$, $y' = \sqrt{e'/e_0}$).

3. SIMULATION OF LONG-TERM DATA

3.1 Individual-based model (IBM)

The IBM simulates stochastically the metapopulation dynamics of generic bush crickets (or any similar invertebrate species) with a one-year life-cycle

(egg – larva – adult) and non-overlapping generations in a highly fragmented, realistic landscape with a binary habitat distinction (habitat vs. non-habitat). Larvae and adults move with certain distances and turning angles, in the matrix much longer and more straight-forward than within the habitat. Adults have a detection radius for finding mating partners in their vicinity (Allee effect). The number of eggs per female is Poisson-distributed and the number of propagules additionally depends on available resources (density dependence). The number of available resources fluctuates because of overlapping local catastrophes (locally correlated environmental fluctuations, but global stochasticity).

For more details, please see Hilker [2002], where the emergence of metapopulation dynamics from the individual behaviour and the patchy distribution of habitats has been demonstrated.

3.2 Extraction of "real" parameter values

A method to extract the IFMs-parameters from the long-term IBM data has been developed in Hilker [2002]. Here, it shall be focused on the method regarding the GBM (which is principally analogous to the PMM). Contrary to the maximum-likelihood approach for yielding the parameters estimators (Subsection 2.3), each of the real values can be extracted by fitting the mechanistic functions of the GBM, i.e. Eq.s (2), (3), (4), to the long-term IBM-data. The IBM is run 200 years to let the metapopulation dynamics reach its quasi-equilibrium. Then, further 400 years are simulated, in which the occupancies of each cell and thus the transitions between being occupied or empty are recorded.

Let N_i^{kl} denote the number of transitions of cell i from state k to l ($k,l = 1$: cell occupied, $k,l = 0$: empty). Then one obtains as likelihood function for the recorded transitions: $P_i = (1 - C_i)^{N_i^{00}} C_i^{N_i^{01}} (E_i - E_i C_i)^{N_i^{10}} (1 - E_i + E_i C_i)^{N_i^{11}}$. Now, C_i and E_i can be approximated by maximizing P_i (which has been done with the Fletcher-Reeves conjugate gradient algorithm [Ueberhuber, 1997]).

Once the extinction and the colonization probabilities of each cell are known, the model equations can be fitted to them in nonlinear least-square fits, thus yielding the unknown parameter set θ. Firstly, the extinction parameter κ can be extracted from the relationship $K_i - E_i$, cf. (2). Next, consider the migration parameter ρ. Transforming (4) yields $\frac{M_{ij}}{p_i K_i \varphi_{ij}} = \exp(-\rho\, r_{ij})$. Since the values of the exponential function for negative arguments are always in the unit interval, the left-hand side is scaled by

Table 1. Mean "real" parameter values (standard deviations) of the GBM.

Species	ρ	μ	κ
1	28.3 (7.8)	0.08 (0.09)	20.9 (19.0)
2	25.8 (8.9)	0.10 (0.07)	2.1 (0.25)
3	20.4 (9.7)	0.10 (0.05)	1.4 (0.07)

dividing through the maximum number of recorded immigrants: $\frac{M_{ij}}{p_i K_i \varphi_{ij} \max M_{ij}} = \exp(-\rho\, r_{ij})$. Now, this equation can be fitted as well. Having determined ρ, the cell connectivities can be computed, which allows to fit (3), thus finally obtaining the colonization parameter μ.

4. SIMULATIONS

With the IBM, three different species have been simulated in varying habitat configurations. Table 1 shows the mean values of the "real" GBM parameter values extracted from the long-term IBM data. Since the parameters are assumed to be species-specific, they are averaged over all habitat configurations as well as replications. In all simulations, the cell length of the GBM has been set to $D = 100$ m.

How accurate and precise are the parameter estimators of the GBM parameterized with snapshot data of two and five consecutive years? In Table 2, the relative errors and variation coefficients are given, which are measures for the accuracy and the precision. If the relative error equals zero, this means a perfect match. If it is positive/negative, the parameter is over-/underestimated, respectively. As one can easily see, there are enormous deviations in the colonization parameter μ. By using more extensive snapshot data with five years, these deviations are reduced, but they are still huge.

In many studies which make use of the PMM, the migration parameter is estimated by independent data (cf. the survey in Hanski [1999]). Analogously, consider the situation in which the "real" value of ρ is known. Then the dimension of the search space in the parameter estimation process is lowered from three to two. The results are listed in Table 3. For snapshot data consisting of two years, the colonization parameter μ is still heavily overestimated. But with five years, the extreme deviations vanish. The same tendency holds for the extinction parameter κ. With two years, it is obviously overestimated for two of the three species. Using five years, κ can be determined more accurately.

In the lower rows of Table 2 and 3, the PMM-

Table 2. Relative errors [variation coefficients] of the parameter estimators. Upper row: GBM, lower row: PMM.

Sp.	2 years	5 years	2 years	5 years	2 years	5 years	2 years	5 years
	ρ		μ		κ			
1	$-0.55\,[0.35]$	$-0.72\,[0.20]$	$403.0\,[130.1]$	$159.3\,[94.3]$	$-0.23\,[0.17]$	$-0.59\,[0.07]$		
2	$-0.85\,[0.02]$	$-0.91\,[0.01]$	$390.3\,[55.4]$	$196.2\,[33.7]$	$4.37\,[0.39]$	$2.24\,[0.25]$		
3	$-0.85\,[0.01]$	$-0.83\,[0.21]$	$369.4\,[29.0]$	$148.8\,[55.7]$	$6.29\,[0.42]$	$3.26\,[0.74]$		
	α		y'		x		e_0	
1	$-0.37\,[0.17]$	$-0.42\,[0.19]$	$9.7\,[8.1]$	$12.6\,[12.9]$	$1.17\,[0.44]$	$0.89\,[0.38]$	$-0.68\,[0.20]$	$-0.50\,[0.28]$
2	$-0.26\,[0.11]$	$-0.24\,[0.09]$	$3.7\,[5.0]$	$3.0\,[1.30]$	$1.20\,[0.37]$	$1.26\,[0.40]$	$-0.81\,[0.17]$	$-0.83\,[0.12]$
3	$-0.33\,[0.13]$	$-0.33\,[0.11]$	$8.9\,[6.8]$	$8.7\,[3.61]$	$1.71\,[0.35]$	$1.69\,[0.41]$	$-0.93\,[0.05]$	$-0.93\,[0.06]$

Table 3. Relative errors [variation coefficients] of the parameter estimators resulting from a 2-dimensional estimation process with given "real" migration parameter values. Upper row: GBM, lower row: PMM.

Species	2 years	5 years	2 years	5 years	2 years	5 years
	μ		κ			
1	$77.7\,[96.17]$	$2.91\,[4.15]$	$-0.21\,[0.25]$	$-0.67\,[0.09]$		
2	$77.0\,[146.1]$	$-0.11\,[0.72]$	$5.11\,[3.17]$	$1.12\,[0.40]$		
3	$43.0\,[114.2]$	$1.12\,[1.41]$	$6.17\,[3.90]$	$1.91\,[0.75]$		
	y'		x		e_0	
1	$0.39\,[1.04]$	$0.46\,[0.79]$	$1.08\,[0.50]$	$1.00\,[0.35]$	$-0.61\,[0.22]$	$-0.60\,[0.22]$
2	$0.68\,[0.51]$	$0.92\,[0.46]$	$1.34\,[0.31]$	$1.43\,[0.29]$	$-0.87\,[0.08]$	$-0.89\,[0.06]$
3	$2.48\,[0.91]$	$2.43\,[0.75]$	$1.59\,[0.30]$	$1.60\,[0.22]$	$-0.92\,[0.04]$	$-0.93\,[0.03]$

estimators are considered as well. Note, that there is one more extinction parameter in the PMM. Only in the case of using five years and with a predetermined migration parameter, the GBM yields deviation ranges similar to those of the PMM. In all other settings, the PMM is more accurate.

5. DISCUSSION

There are enormous deviations in the estimators of the GBM. They decrease, if the migration parameter is predetermined from independent data. That is not surprising, because the dimension of the search space is reduced. However, the usage of the GBM seems to be applicable only if the migration parameter is known. Moreover, the GBM proves to be relatively accurate only in the case when five snapshot years are available. This can be explained as follows. With two snapshot years, the relative frequency of occupancy p_i 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.

So far the accuracy of the estimators has been considered. What happens, if the estimators are used in the IFM simulation process? Incorporating the parameters into the model equations, i.e. in the case of GBM (2) and (3), the predicted incidences
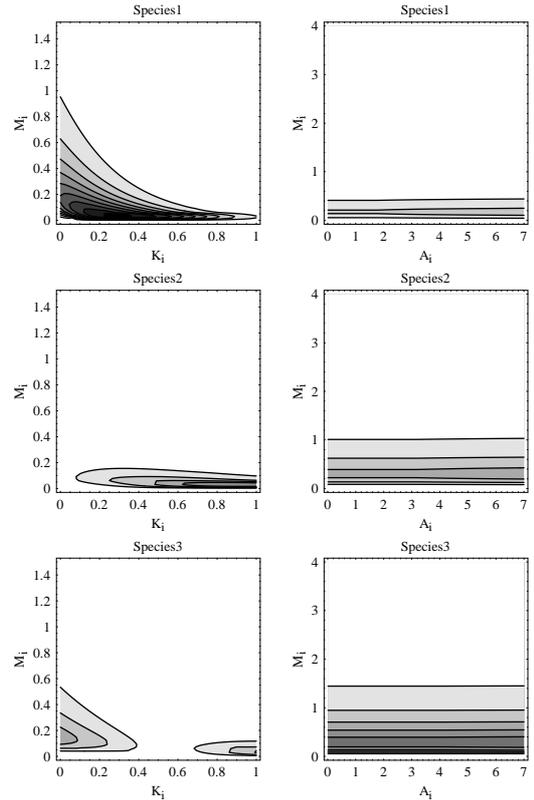
Figure 3. Contour plots of the residuals in J_i (left: GBM, right: PMM). White means a perfect match, and each contour line / darker shading corresponds to an increase in the residuals of 0.1. Details are explained in the text. Note the different scaling of the ordinate axes, i.e. the connectivities in the sense of the PMM and GBM.

J_i can be calculated as a function of the parameter estimators, M_i and K_i. If this process is repeated with the "real" values, the residual differences between estimated incidences can be calculated (Figure 3). These may be taken as a good measure for the relevance of errors in parameter estimators. In the PMM the patch areas have nearly no influence on the residuals. They are determined by the connectivity. In contrast, the residuals in the GBM are not only influenced by the connectivity, but much more by the carrying capacities.

Settele [1998] originally proposed only to consider cells which may be potential habitat. 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 of the IFM-approaches. When extracting the "real" value of the GBM extinction parameters, we scaled the number of immigrants in order to obtain a first approximation. This might be a problem, since the maximum number of recorded immigrants depends on the landscape. Alternatively, one could use for the fit a second extinction parameter (note that then the number of parameters would be equal to the PMM). However, these modifications could resolve the essential deviations (but much more field work would be necessary).

The cell length has been chosen in the size of the smallest habitat. A systematic investigation of the influence of the cell length would be of interest, of course. Nonetheless, the great deviations in the estimators elucidate severe disagreements in the underlying assumptions.

6. CONCLUSIONS

Testing the quality of metapopulation models is generally a difficult issue, because little or even no data are available. Highly specific models can be used to substitute missing "real-world" data. This allows not only to parameterize single models. Additionally, different model architectures can be compared. This has exemplarily been demonstrated to a grid-based approach vs. the well-known patch-matrix model.

It shows, that the GBM leads to worrying misestimations. However, conditions have been derived, under which the accuracy is in the same range as for the PMM. Apart from the possible dissection of natural habitats, the results of this study indicate another shortcoming of the grid-based approach, namely that there seems to be the need of a profound number of snapshot years to determine the carrying capacity of a cell. Regarding the PMM, the amount of snapshot years has been considered in the context whether the metapopulation has reached its quasi-equilibrium [Moilanen, 2000].

Concerning more general aspects, in many studies has been stated a gap between simple, highly aggregated models on the one hand and specific models on the other hand. The former are often analytically tractable due to their rather general assumptions about population dynamics (which are often simply ignored). Thus being parameter-sparse, they allow to give insight into elementary relationships of state variables. On the other hand, specific models need a lot of information about the species' life cycle. This paper is situated at the edge of these model types, utilizing the different conceptual approaches and trying to make them more comparable.

REFERENCES

Czárán, T. *Spatiotemporal models of population and community dynamics*. Chapman & Hall, London, 1998.

Hanski, I. A practical model of metapopulation dynamics. *Journal of Animal Ecology*, 63:151–162, 1994.

Hanski, I. *Metapopulation Ecoloy*. Oxford University Press, New York, 1999.

Hanski, I. Spatially realistic theory of metapopulation ecology. *Naturwissenschaften*, 88:372–381, 2001.

Hilker, F. M. *Parametrisierung von Metapopulationsmodellen*. Diploma thesis, Department of Mathematics and Computer Science, University of Osnabrück, 2002.

Levins, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15:237–240, 1969.

Moilanen, A. The equilibrium assumption in estimating the parameters of metapopulation models. *Journal of Animal Ecology*, 69:143–153, 2000.

Settele, J. *Metapopulationsanalyse auf Rasterdatenbasis*. Teubner Verlag, Leipzig, Stuttgart, 1998.

Ueberhuber, C. W. *Numerical Computation 2*. Springer, Berlin, 1997.