Testing metapopulation models with stream-fish assemblages

Nicholas J. Gotelli1* and Christopher M. Taylor2

1Department of Biology, University of Vermont, Burlington, VT 05405 and 2Department of Biological Sciences, PO Drawer GY, Mississippi State University, Mississippi State, MS 39762, USA

ABSTRACT
We assessed the utility of simple metapopulation models in an analysis of a decade (1976–86) of stream-fish censuses at 10 sites on the Cimarron River, Oklahoma. This assemblage is a good candidate for metapopulation analysis because there was substantial annual turnover of population, with frequent local extinctions and recolonizations. However, annual probabilities of colonization and extinction were usually not correlated with the fraction of sites occupied, a key prediction of many Levins-style metapopulation models. Instead, these probabilities were related to position in the stream gradient. For most species, colonization was less likely and extinction more likely in upstream than downstream sites. Consequently, the simple Levins-style metapopulation models failed to accurately predict the dynamics of most species. Metapopulation models that incorporate spatial variability in colonization and extinction probabilities may be more successful for populations that are distributed across environmental gradients.

Keywords: colonization, extinction, freshwater fishes, metapopulation.

INTRODUCTION
Metapopulations represent a major research focus in both basic (Hanski and Gilpin, 1991) and applied (Doak and Mills, 1994) population biology. Whereas classic demographic models treat local populations as closed systems (Hastings, 1997), metapopulation models describe an open system, in which extinction and colonization depend on movement of individuals among a set of patchy sites (Levins, 1969, 1970). These models may be relevant to the dynamics of endangered species, which often persist in highly fragmented landscapes (Edwards et al., 1994). A key prediction of many metapopulation models is that the fraction of sites occupied in the metapopulation affects the probability of local colonization and extinction (Gotelli and Kelley, 1993). The appropriate null hypothesis is that probabilities of colonization and extinction are independent of landscape occupancy, which corresponds to an ‘island–mainland’ model (Harrison et al., 1988).

*Author to whom all correspondence should be addressed. e-mail: ngotelli@zoo.uvm.edu

© 1999 Nicholas J. Gotelli
In spite of the popularity of metapopulation models, they have proven difficult to test with field data. Documenting extinctions is difficult and controversial because of complications caused by sampling effort (Clark and Rosenzweig, 1994), residency and extinction criteria (Haila and Hanski, 1993), and the lengthy time frame necessary to document colonization and extinction of entire populations (Harrison et al., 1988). What is needed is long-term monitoring of many sites in which resident populations exhibit frequent demographic extinctions (e.g. Pfister, 1998).

In this study, we analysed a remarkable data set – a decade of annual censuses of 46 fish species at 10 sites on the Cimarron River, Oklahoma. From these data we extracted annual probabilities of colonization and extinction for each species. We then tested whether those probabilities were correlated with the fraction of sites occupied in each yearly census. We also analysed whether colonization and extinction probabilities could be predicted on the basis of site location or year. Whereas most tests of metapopulation models have analysed static patterns that are predicted by the models (e.g. Scheiner and Rey-Benayas, 1997), our analyses address the underlying dynamic assumptions of the models.

MATERIALS AND METHODS

Study sites
Our analyses are based on a decade-long census of fishes of the Cimarron River, graciously provided by Jimmie Pigg. Detailed descriptions of the river, the terrestrial landscape and the fish fauna can be found in Pigg (1988) and are briefly summarized here. Censuses were conducted at 10 permanent study sites on the Cimarron River in Oklahoma, covering a distance of 950 km (Fig. 1). The median distance between adjacent sampling sites is 44 km (range 19–347 km). These sites span an ecological gradient from west to east (Omernik, 1987). Western sites are characterized by ephemeral river flows, low precipitation (45 cm·year⁻¹), high evaporation rates and high summer chloride levels. Eastern sites are characterized by more stable river flows, higher precipitation (91 cm·year⁻¹), and lower evaporation rates and chloride levels. Although the census sites do not represent true islands or discrete habitat patches, there is enough habitat heterogeneity (pools, riffles, runs) in the Cimarron River and enough distance between the sites to allow us to treat them as discrete patches in a metapopulation model.

Censuses
Fish collections were made by J. Pigg from May 1976 to November 1986. At each of the 10 sites, he sampled fishes two to three times a year with a 3.3 × 1.3 m heavy-leded seine with a 3.00 mm mesh. At each site, 200 m of stream were sampled each visit, with an effort to sample the same area with the same effort each time. In addition to seining, a 33.0 × 1.6 m monofilament gill net (2.5 cm mesh) was placed across the stream for approximately 2 h during each census visit. All sampling methods contain some biases; however, Pigg’s methods were designed to thoroughly sample the fish fauna, so that collections would be directly comparable among years and among sites. A species was defined as locally ‘present’ in a particular year if it occurred in one or more of the local collections from that year. A species was defined as locally ‘absent’ if it did not occur in any of the collections from that year. Average abundance was calculated for each species across all years by averaging its
total abundance in all collections in which it was present. Forty-one native species were recorded and used in our analyses. We discarded data on five introduced species, although the results were similar with and without the inclusion of non-natives. A species list and discussion of phylogenetic relationships is provided in a companion paper on macroecological analysis (Gotelli and Taylor, 1999).

**Metapopulation models**

Levins (1969) first introduced metapopulation models of the form:

\[
df/dt = \text{immigration rate} - \text{extinction rate}
\]

where \( f \) is the fraction of homogeneous sites occupied in a metapopulation (0.0 \( \leq f \leq 1.0 \)). These models ingeniously captured metapopulation dynamics without resorting to more complex equations that explicitly describe individual dispersal between patches and population dynamics within patches. Gotelli and Kelley (1993) constructed a general metapopulation model that captures the dynamics of many of Levins-style models and generates testable predictions:

\[
df/dt = (1 - f)(a + bf) - (f)(c - df)
\]

In this model, the immigration rate is defined as the product of the fraction of unoccupied sites (1 - \( f \)) and the probability of local colonization (\( a + bf \)). The extinction rate is defined as the product of the fraction of occupied sites (\( f \)) and the probability of local extinction (\( c - df \)). The metapopulation models of Levins (1969), Hanski (1982) and Gotelli (1991), as well as MacArthur and Wilson’s (1967) equilibrium model for a single species, can all be derived as special cases of this general model (Gotelli and Kelley, 1993).

The general model makes qualitative predictions about the relationship between the fraction of sites occupied and the probability of local colonization and extinction (Fig. 2). If colonization can occur from occupied patches within the metapopulation, the coefficient \( b \) must be greater than zero. In contrast, if colonization can occur only from a permanent mainland area (‘propogule rain’), the coefficient \( b \) equals zero. The coefficient \( a \) describes
the probability that a solitary patch is colonized in an empty landscape. If $a$ and $b$ are both greater than zero, colonization depends on propagules derived from both internal and external sources.

Similarly, the term $(c - df)$ describes the extinction function. If there is a ‘rescue effect’ (Brown and Kodric-Brown, 1977), the coefficient $d$ must be greater than zero, because the probability of extinction diminishes as more sites are occupied. The coefficient $c$ describes the probability of extinction for an occupied patch in an empty landscape. If $c$ and $d$ are both greater than zero, then extinction depends on intrinsic properties of local populations as well as influences of migrating individuals from other patches.

**Statistical tests**

We used a Monte Carlo simulation to evaluate the relationship between the fraction of sites occupied and the probability of colonization or extinction. Our procedure was to construct, for each species, an occurrence matrix in which each row was a year and each column was a site. In this matrix, an entry of 1 indicates the presence of a population at a site in a particular year, and a zero indicates that the species was absent. Sites that were uncensused in a particular year were not used in any of our calculations. Next, we calculated $p_e$, the probability of extinction between each pair of consecutive years, as:

$$p_e = \frac{\text{number of sites occupied in year } (t) \text{ that were unoccupied in year } (t+1)}{\text{number of sites occupied in year } (t)}$$

Thus, for these census data of 11 years, there were 10 annual probabilities of extinction that could be calculated. Next, we calculated $f$, the fraction of sites occupied, as:

$$f = \frac{\text{number of sites occupied in year } (t)}{\text{number of sites that were censused in year } (t)}$$

Figure 3 illustrates these calculations. Clark and Rosenzweig (1994) present alternative formulae that can be used when censuses are at irregular intervals, or when there is a chance...
Testing metapopulation models

of an intervening colonization and extinction between censuses. In this case, because the sites were censused two or three times annually, it is less likely that there were intervening extinctions or recolonizations, so we have retained the simpler formulae above.

For each species, we calculated the least squares regression slope (Model I) of the probability of extinction versus the fraction of sites occupied. To evaluate the statistical significance of this regression slope, we randomized the occurrence matrix for each species. The occurrences and absences of each species were randomly resampled, with the proviso that uncensused sites were excluded and could not receive either a 1 or a zero. Neither row nor column totals were constrained, so the different matrix arrangements were equiprobable. This randomization eliminated any relationship between the fraction of sites occupied and the subsequent probability of extinction.
From this randomized matrix, we also calculated a least squares regression slope. We repeated the simulation 1000 times, and generated a histogram of 1000 slope values. We used this histogram to calculate the tail probability that the observed slope was less than or equal to the expected slope. We used a one-tailed test of the hypothesis that the observed slope was less than expected. In other words, we tested whether the probability of local extinction decreased significantly as more sites were occupied. This is the key prediction of metapopulation models that include a rescue effect.

We also carried out simulations to test for metapopulation effects on colonization. The probability of colonization $p_i$ was defined as:

$$p_i = \frac{\text{number of unoccupied sites in year }(t) \text{ that were occupied in year }(t + 1)}{\text{the number of sites that were censused in year }(t)}$$

We carried out an identical simulation procedure to test for correlations of colonization probability and fraction of sites occupied. In these simulations, we used a one-tailed test for whether the observed slope was significantly greater than expected. A large, positive slope would mean that, as more sites were occupied, the probability of local colonization of empty sites increased. This would be the case in which colonization is internal, and depends on migrants from other occupied sites. In the absence of metapopulation dynamics, propagules could arrive from a large constant, ‘mainland’ source, in which case the probability of colonization would be independent of site occupancy (the island–mainland model).

**Testing the premises**

We used this same simulation approach to test two of the basic premises of metapopulation models. One premise is that the metapopulation has reached an equilibrium state, in which local extinctions are balanced by local colonizations. If equilibrium has not been reached – for example, if the environment is changing – probabilities of colonization and extinction will change with time. To test this premise, we first computed, across sites, the probabilities of colonization and extinction for each census year. Next, we measured the slope of the relationship between colonization and extinction probabilities, and the year of the census. The null hypothesis is that the slope does not differ significantly from zero. We tested this hypothesis by comparing the observed slope to the expected slope calculated from the randomized data set. We used this same simulation protocol to test whether the fraction of sites occupied by a species increased or decreased significantly with time.

The second premise of Levins-style metapopulation models is that the sites are homogeneous, so that probabilities of local colonization and extinction are the same across all sites. In this study, the sites occur along a linear geographic gradient, from the headwaters of the Cimarron River to its confluence with the Arkansas River. There are major changes in physical, chemical and biotic conditions along this gradient that are likely to influence the probability of colonization and extinction. Therefore, we calculated these probabilities for each site, across all times, and assessed whether these probabilities were significantly correlated with the position of a site in the river continuum (from 1 = headwaters to 10 = river mouth). As before, we compared observed regression slopes to the histogram of slopes calculated from the randomized data set.

We then used a second simulation procedure to control for spatial effects due to the linear geographic gradient. As before, we calculated the observed slope of the regression of the
probability of colonization or extinction against the fraction of sites occupied. However, we restricted the randomization so that the ones and zeros in the matrix were randomized only within each column. This restriction preserved the pattern of occurrence of species at each site, so that sites that were sparsely or frequently occupied retained that status in the simulated matrix. This test allowed us to search for metapopulation patterns while accounting for differences among sites that might affect the probability of colonization and the probability of extinction.

We carried out our simulations on all species, except those with two or fewer occurrences in the census matrix, or those in which extinction or colonization probabilities could not be estimated. For each simulation, we created the histogram of tail probability values. Treating each species as an independent observation, we used a simple chi-square goodness-of-fit test to decide whether the histogram of species tail probabilities matched the expected number of species in the left tail ($P \leq 0.05$), in the right tail ($P \geq 0.95$) and in the random fraction ($0.05 < P < 0.95$) of the distribution.

**RESULTS**

Local extinctions

The relationship between the probability of local extinction and the fraction of sites occupied was non-random (Table 1A), with a slight excess of species occurring in both tails of the distribution. Specifically, there were five species for which the probability of extinction decreased as more sites were occupied. This pattern is consistent with the ‘rescue effect’ of metapopulation models (Fig. 2b). There were four species that deviated in the opposite direction, with the probability of extinction increasing significantly as more sites were occupied. These test results are based on an unrestricted randomization, which assumes that sites are equiprobable. We repeated the randomization with a constraint on column totals, which preserves site differences, and found that the observed and expected distributions did not differ significantly. The number of species exhibiting a significant rescue effect fell from five to one (Table 1B).

For most species, the probability of extinction neither increased nor decreased significantly with time (Table 1C). However, there were strong effects of site position (Table 1D). For 17 of 36 species, the probability of extinction decreased significantly from the headwaters (site 1) to the mouth of the Cimarron River (site 10). Four species showed the opposite pattern. For the remaining 15 species, the probability of extinction was unrelated to the position of the site.

Local colonizations

The relationship between the probability of colonization and the fraction of sites occupied was non-random (Table 1E). For eight species, the probability of colonization was greater when fewer sites were occupied, which goes against the predictions of metapopulation models. Only one species showed the predicted pattern of a significant increase in the probability of colonization with increasing site occupancy. A similar pattern held when the randomization was constrained by column totals (Table 1F).

Colonization patterns were also non-random with respect to time and space. Five species showed a significant decrease in colonization probabilities over time, and no species showed
Table 1. Monte Carlo significance tests of slope values

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Randomization</th>
<th>Number of species with a significant negative slope ($P \leq 0.05$)</th>
<th>Number of species with a non-significant slope ($0.05 &lt; P &lt; 0.95$)</th>
<th>Number of species with a significant positive slope ($P \leq 0.95$)</th>
<th>$\chi^2$</th>
<th>Overall $P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$p_e$</td>
<td>Occupancy ($f$)</td>
<td>Full</td>
<td>5</td>
<td>27</td>
<td>4</td>
<td>9.3</td>
<td>0.01</td>
</tr>
<tr>
<td>B</td>
<td>$p_e$</td>
<td>Occupancy ($f$)</td>
<td>Restricted</td>
<td>1</td>
<td>32</td>
<td>3</td>
<td>1.2</td>
<td>0.56</td>
</tr>
<tr>
<td>C</td>
<td>$p_e$</td>
<td>Time</td>
<td>Full</td>
<td>3</td>
<td>31</td>
<td>2</td>
<td>0.9</td>
<td>0.64</td>
</tr>
<tr>
<td>D</td>
<td>$p_e$</td>
<td>Spatial position</td>
<td>Full</td>
<td>17</td>
<td>15</td>
<td>4</td>
<td>140.4</td>
<td>$3.3 \times 10^{-31}$</td>
</tr>
<tr>
<td>E</td>
<td>$p_i$</td>
<td>Occupancy ($f$)</td>
<td>Full</td>
<td>8</td>
<td>27</td>
<td>1</td>
<td>23.6</td>
<td>$7.5 \times 10^{-6}$</td>
</tr>
<tr>
<td>F</td>
<td>$p_i$</td>
<td>Occupancy ($f$)</td>
<td>Restricted</td>
<td>5</td>
<td>30</td>
<td>1</td>
<td>6.2</td>
<td>0.04</td>
</tr>
<tr>
<td>G</td>
<td>$p_i$</td>
<td>Time</td>
<td>Full</td>
<td>5</td>
<td>31</td>
<td>0</td>
<td>7.5</td>
<td>0.02</td>
</tr>
<tr>
<td>H</td>
<td>$p_i$</td>
<td>Spatial position</td>
<td>Full</td>
<td>2</td>
<td>14</td>
<td>20</td>
<td>194.5</td>
<td>$5.8 \times 10^{-41}$</td>
</tr>
<tr>
<td>I</td>
<td>$f$</td>
<td>Spatial position</td>
<td>Full</td>
<td>6</td>
<td>6</td>
<td>25</td>
<td>321.4</td>
<td>$1.6 \times 10^{-70}$</td>
</tr>
</tbody>
</table>

* The letters in the first column indicate hypotheses referred to in the text. The second and third columns give the dependent and the independent variable in the analysis, respectively. The fourth column indicates whether the entire data set was randomized (full), or whether the randomizations were constrained within sites (restricted). The next three columns give the number of species with slopes that were significantly less than, equal to, or greater than expected. A chi-square test is used to assess the fit of this distribution to that expected by chance. $f$ = fraction of sites occupied, $p_e$ = probability of extinction, $p_i$ = probability of colonization, time = year of census (1976 . . . 1980), spatial position = location of census site (1 . . . 10).
a significant increase (Table 1G). Twenty species had greater colonization probabilities in downstream sites, whereas only two species showed significantly greater colonization probabilities in upstream sites (Table 1H).

**DISCUSSION**

As with the original model of MacArthur and Wilson, metapopulation models were quickly embraced by the conservation community (Doak and Mills, 1994), perhaps because they generated simple conservation prescriptions about the minimum amount of habitat necessary to prevent extinctions (Nee, 1994; Tilman et al., 1994). Although it has been popular to fit metapopulation models to static spatial data, empirical evidence supporting the assumptions and predictions of these models is weak (Harrison, 1991; Murdoch et al., 1996; Scheiner and Rey-Benayas, 1997; Pfister, 1998).

We think it is important to test for the dynamic relationships in Fig. 2, which embody the assumptions of Levins-style metapopulation models. Similar issues have arisen in the evaluation of the MacArthur and Wilson (1967) equilibrium model of island biogeography. Although many studies have used the model of MacArthur and Wilson (1967) to interpret species-area relationships, very few have addressed the model's premise of a dynamic balance between ongoing colonization and extinction (Williamson, 1981; Strong and Rey, 1982).

We know of only one other study that has tested directly the metapopulation predictions in Fig. 2. Pfister (1998) conducted a 3 year study of colonization and extinction by three species of north Pacific tidepool fishes. As in our analyses, she found that the fraction of sites occupied did not affect extinction. For one species, the probability of colonization was negatively correlated with site occupancy, a pattern that we also found for eight species in the Cimarron River assemblage. For the tidepool populations, tide pool volume was a better predictor of colonization and extinction than site occupancy. This again parallels our finding that, for most species, position in the linear stream gradient was more important than site occupancy in determining colonization and extinction patterns. Perhaps these similar results for marine and freshwater fishes suggest that metapopulation models are more appropriate for certain other taxa, such as butterflies (Hanski et al., 1994, 1995) and shrews (Peltonen and Hanski, 1991).

Although many populations occur in fragmented landscapes, they must be linked by the proper amount of dispersal to qualify as true metapopulations. Indeed, many fragmented populations have not been shown to even undergo local extinction and recolonization, a minimum prerequisite for a dynamic metapopulation model. Fish populations of the Cimarron River do show considerable turnover and local recolonization on an annual scale (Fig. 3), so they are good candidates for a test of metapopulation dynamics. However, few species display colonization or extinction dynamics that are consistent with the predictions of a large class of Levins-style metapopulation models (Fig. 2).

Classic metapopulation models assume a set of homogeneous patches with similar colonization and extinction probabilities. To control for differences in site quality, we re-analysed the effect of site occupancy by restricting our randomizations within sites. The result was that even fewer species showed non-random distributions (Table 1B, 1F). A small, but significant fraction of species in this assemblage also showed evidence of temporal changes in colonization or extinction probabilities. For these species, occupancy patterns are highly dynamic with time, and probably do not represent equilibrium conditions. Such species may
be undergoing geographic range expansions or contractions, and their dynamics are best described by other kinds of models (Veit and Lewis, 1996).

At the spatial scale of our analyses, the best explanation for the dynamics of most species of Cimarron River fishes is the ‘island–mainland model’ (Harrison et al., 1988), in which colonizations are from sites other than those in the metapopulation, and extinctions are independent of one another. This model is a single-species analog of the MacArthur and Wilson equilibrium model (Gotelli, 1991). This is not to say that local extinctions and colonizations are unpredictable. Indeed, our analyses suggest that colonizations and extinctions for most species vary predictably by their position in the river gradient. Extinction probabilities increase upstream, and colonization probabilities increase downstream. These dynamics lead to highly non-random spatial distributions for 31 of 37 species analysed (Table 1I). Thus, a metapopulation model that incorporates spatial variability in extinction and colonization probabilities (Verboom et al., 1991; Hanski, 1996) may be more appropriate for this system than the classic Levins-style models.

Our analysis is the first that we are aware of to demonstrate directly that non-random patterns of stream-fish distribution reflect spatial gradients in colonization and extinction probabilities. Other studies have also implicated position within a watershed as an important determinant of stream-fish colonization and extinction dynamics (Osborne and Wiley, 1992; Taylor, 1997). Although we found little evidence for the predictions of classic metapopulation models, extinction, colonization and the occurrence of most species were predictable on the basis of site location (Table 1D, 1H, 1I). As Pfister (1998) has recently emphasized, more long-term studies of this type are needed to address directly the key assumptions of metapopulation theory.

ACKNOWLEDGEMENTS

We thank the late J. Pigg for generous access to his long-term data set. The manuscript benefited from comments by A. Brody, G. Entsminger, J. Brown and C. Pfister.

REFERENCES

Testing metapopulation models


