



Research

Cite this article: Song C, Fortin M-J, Gonzalez A. 2022 Metapopulation persistence can be inferred from incomplete surveys. *Proc. R. Soc. B* **289**: 20222029. <https://doi.org/10.1098/rspb.2022.2029>

Received: 9 October 2022

Accepted: 17 November 2022

Subject Category:

Global change and conservation

Subject Areas:

ecology, ecosystems, theoretical biology

Keywords:

metapopulation capacity, patch importance, connectivity matrix, landscape matrix, incomplete survey

Author for correspondence:

Chuliang Song

e-mail: chuliang.song@mcgill.ca

[†]Present address: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6326400>.

Metapopulation persistence can be inferred from incomplete surveys

Chuliang Song^{1,2,†}, Marie-Josée Fortin² and Andrew Gonzalez¹

¹Department of Biology, Quebec Centre for Biodiversity Science, McGill University, Montreal, Canada H3A 1B1

²Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada M5S 3B2

CS, 0000-0001-7490-8626; M-JF, 0000-0002-9935-1366; AG, 0000-0001-6075-8081

Habitat destruction and fragmentation are principal causes of species loss. While a local population might go extinct, a metapopulation—populations inhabiting habitat patches connected by dispersal—can persist regionally by recolonizing empty patches. To assess metapopulation persistence, two widely adopted indicators in conservation management are metapopulation capacity and patch importance. However, we face a fundamental limitation in that assessing metapopulation persistence requires that we survey or sample all the patches in a landscape: often these surveys are logistically challenging to conduct and repeat, which raises the question whether we can learn enough about the metapopulation persistence from an incomplete survey. Here, we provide a robust statistical approach to infer metapopulation capacity and patch importance by sampling a portion of all patches. We provided analytic arguments on why the metapopulation capacity and patch importance can be well predicted from sub-samples of habitat patches. Full-factorial simulations with more complex models corroborate our analytic predictions. We applied our model to an empirical metapopulation of mangrove hummingbirds (*Amazilia boucardi*). On the basis of our statistical framework, we provide some sampling suggestion for monitoring metapopulation persistence. Our approach allows for rapid and effective inference of metapopulation persistence from incomplete patch surveys.

1. Introduction

Populations in nature are rarely isolated. Instead, populations are often distributed in discrete habitat patches, where these patches are linked by dispersal and colonization events [1,2]. These metapopulations [3] shift the focus from local persistence to regional persistence. That is, even when a patch is not suitable for the persistence of a population *locally*, the metapopulation can still persist *regionally* via the dispersal process [4]. In the current era with the increasing habitat loss and fragmentation [5,6], we need to be able to predict the persistence of metapopulations for conservation management.

To predict metapopulation persistence in a patchy landscape, a fundamental indicator is the metapopulation capacity [7–9]. Rooted within population dynamics theory, the metapopulation capacity estimates the ability of any arbitrary landscape structure to support metapopulation persistence. A metapopulation would persist regionally as long as its metapopulation capacity is greater than the extinction rate. Furthermore, based on the contribution of each patch to metapopulation capacity, we can quantify the conservation importance of each patch [10]. Beyond applicability for metapopulations, metapopulation capacity and patch importance are also useful for the study of metacommunities [11,12] and in conservation planning and management [13–16].

However, we face a fundamental limitation to assessing metapopulation persistence empirically in landscapes supporting many populations [17]. As all patches have some non-zero influence on metapopulation capacity, to quantify metapopulation persistence would seem to require surveying or sampling all the habitat patches in a landscape for the relevant parameters, such as

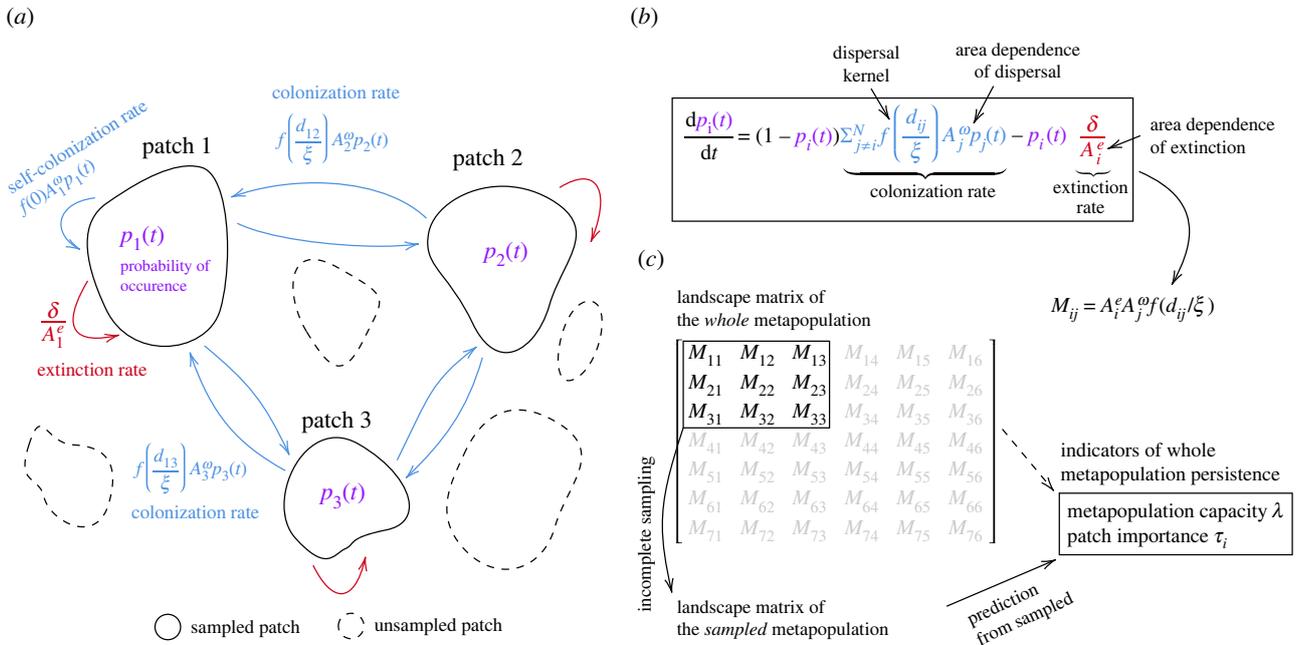


Figure 1. Fundamental limitation in quantifying metapopulation persistence. (a) A metapopulation of seven habitat patches in a landscape. Each patch has probability p_i of occurrence. This occurrence probability is affected by the colonization rate and extinction rate. The colonization rate includes the dispersal from other patches and potentially self-colonization. In this example, three patches are sampled, while the other four patches are unsampled. (b) The model metapopulation dynamics. The colonization rate on patch i is quantified as $\sum_{j=1}^N f(d_{ij}/\xi) A_j^\omega p_j(t)$, which summarizes the contributions of patch distance and area: $f(d_{ij}/\xi)$ represents the effect of distance on dispersal rate, where d_{ij} denotes patch distance between patch i and j , ξ denotes the characteristic dispersal distance that transforms distance unitless and f denotes the dispersal kernel that scales the distance to appropriate rate; A_j^ω represents the effects of patch area, where A_j denotes the area of patch j and ω denotes the extent of area dependence of dispersal (larger area disperses more if $\omega > 0$, and vice versa). The extinction rate on patch i is quantified as δ/A_i^e , where δ denotes the general extinction rate (a property of the species and of the habitat network as a whole) and e denotes the extent of area dependence of extinction (larger area has a lower extinction rate). With this general spatially explicit formulation of metapopulation dynamics, the connectivity matrix is given by $M_{ij} = A_i^e A_j^\omega f(d_{ij}/\xi)$. (c) The problem in quantifying metapopulation persistence. Two important indications are metapopulation capacity λ and patch importance τ_i . Metapopulation theory proves that metapopulation capacity λ is given by the leading eigenvalue of the connectivity matrix, and patch importance τ_i is given by the associated eigenvector. As the metapopulation has seven patches in total, the connectivity matrix of the whole metapopulation is of dimension 7×7 . With the incomplete sampling, only a small part of the connectivity matrix is estimated, which has dimension 3×3 . In other words, although we have sampled three of seven patches, we only have information on 9 out of 49 elements of the connectivity matrix. This quadratic decline of information with sampling effort underscores the importance of a statistical framework to facilitate empirical sampling and monitoring of metapopulation persistence. (Online version in colour.)

patch quality, pairwise patch distance and so on. Assessment of some features can be done with remote sensing, but typically patch quality must be done with the fieldwork. Yet, it is often logistically infeasible to survey all the patches, so we are left with incomplete information. Similarly, this fundamental limitation also applies to the quantification of patch importance.

To address this crucial constraint, we provide a new statistical approach to infer metapopulation capacity and patch importance by sampling a small subset of all the patches. Technically, metapopulation capacity is the leading eigenvalue of the connectivity matrix, a matrix summarizing the structure of connections among habitat patches. For an arbitrary matrix, the eigenvalue of the whole matrix is not related to the eigenvalues of sub-matrices [18]. However, thanks to the ecological constraints on the connectivity matrix, the eigenvalue of the whole matrix (i.e. true metapopulation capacity) is inherently constrained by the eigenvalues of sub-matrices (i.e. sampled metapopulation capacity). The inference for patch importance is similar. Patch importance is technically measured as the eigenvector corresponding to the leading eigenvalue of the connectivity matrix. Again, thanks to the ecological constraints, the eigenvector can be inferred from eigenvalues [19]. In the following, we justify our statistical framework with analytic arguments

and validate it with extensive simulations with the state-of-the-art model complexity. As a proof of concept, we applied our result to an empirical metapopulation of the mangrove hummingbird (*Amazilia boucardi*) [15]. Following our statistical analysis, we highlight the implications of our method for conservation.

2. Method

(a) Metapopulation dynamics

We adopt a general model of metapopulation dynamics. Suppose we have n discrete habitat patches. The probability of finding species in patch i is governed by:

$$\frac{dp_i(t)}{dt} = (1 - p_i(t)) \underbrace{\sum_{j=1}^N f\left(\frac{d_{ij}}{\xi}\right) A_j^\omega p_j(t)}_{\text{Colonization rate}} - p_i(t) \underbrace{\frac{\delta}{A_i^e}}_{\text{Extinction rate}}, \quad i = 1, \dots, n, \quad (2.1)$$

where d_{ij} denotes the distance between patch i and j , ξ is the characteristic distance, f is the dispersal kernel, A_i is the 'quality' of patch i , ω is the area dependence of dispersal, e is the area dependence of extinction, and δ is the extinction rate. Figure 1 illustrates this metapopulation dynamics. Importantly, the patch quality A_i was originally measured by patch area [20–22]

and has been widely used in empirical studies. However, a number of more suitable modern measures have been proposed to better account for species fitness and landscape suitability, measuring quality as local environmental conditions instead of area [1,23,24]. For simplicity and in accordance with the literature on metapopulation capacity, we refer to patch quality A_i as patch area throughout.

As a historical note, the basic formulation of equation (2.1) comes from [7]. We have incorporated many recent modifications: negative extinction-area parameter [25], self-colonization $f(d_{ii}) > 0$ [26], non-Gaussian dispersal kernel [27] and a non-monotonic dispersal kernel [28].

Following the metapopulation dynamics (equation (2.1)), the connectivity matrix M is given by

$$M_{ij} = A_i^e A_j^o f\left(\frac{d_{ij}}{\xi}\right), \quad (2.2)$$

which summarizes the structural information defining how patches are connected by dispersal. The leading eigenvalue λ of the connectivity matrix M is the metapopulation capacity. Reference [7] proved a simple criterion that a metapopulation will persist as long as

$$\lambda > \delta. \quad (2.3)$$

The patch importance A_i of patch i measures how much the metapopulation capacity decreases if we remove this patch. Mathematically, the patch importance is [10],

$$A_i = \lambda \tau_i, \quad (2.4)$$

where τ_i is the i th element in the leading eigenvector of matrix M . The eigenvector is normalized such that τ_i sums to 1. Thus, the relative patch importance is given by τ_i . Importantly, given the locality of this measure, it does not necessarily imply what happens if we remove two or more patches, or what happens if we add a new patch.

(b) Analytic argument on metapopulation capacity

All elements M_{ij} in the connectivity matrix M (equation (2.2)) are clearly non-negative. If we assume that the kernel is strictly positive ($f(d_{ij}) > 0$), then the Perron–Frobenius theorem guarantees that the metapopulation capacity is positive (i.e. leading eigenvalue $\lambda > 0$). More concretely, the metapopulation capacity is bounded below by the minimum row sum and bounded above by the maximum row sum (i.e. $\min_i \sum_j M_{ij} \leq \lambda \leq \max_i \sum_j M_{ij}$). This bound has been previously noted in studying metapopulation capacity [9]. This is the best general estimate we have for positive matrix (see ch. 8 of [29]). However, this estimate is not ideal for inference, as we need to infer the ‘outlier’ behaviour of row sums.

If we further assume that the area dependence rate of dispersal ω and extinction rate e are equal (most studies assume both to be 1), then, in addition to positivity, M is also symmetric (i.e. $M_{ij} = M_{ji}$). Then the Cauchy Interlace Theorem guarantees that M always has a larger leading eigenvalue than any of its principal sub-matrices [30]. In ecological language, this theorem means that a metapopulation always has a higher capacity than any of its sub-metapopulations.

Furthermore, for a symmetric positive matrix, metapopulation capacity can be equivalently expressed as $\lambda = \max_{x \neq 0} (x^T M x / x^T x)$ [31]. We can consider the vector x as the weights of matrix elements. For example, if we assign equal weight to each element (i.e. $x = (1, \dots, 1)$), we see the metapopulation capacity is bounded below by $\lambda \geq n\bar{M}$, where \bar{M} represents the average of the connectivity matrix. This lower bound has been previously noted in studies of metapopulation capacity [28]. This lower bound ($n\bar{M}$) is better for inference compared to the previous bound ($\min_i \sum_j M_{ij}$), as we only need to estimate the

‘average’ behaviour of row sums. The powerful central limit theorem makes it easy and robust to estimate the lower bound ($n\bar{M}$). Specifically, as long as M_{ij} are from some probability distribution, any distribution, we can estimate its mean.

We can further improve the bound if we view M_{ij} as a random symmetric matrix. That is, we assume that the off-diagonal elements M_{ij} in the upper triangle are i.i.d. with variance σ^2 , and the lower triangle is identical to the upper triangle. Note that this assumption is not valid when we have heterogeneity of the patch area A_i (as $A_i^e A_j^o$ generally does not equal to $A_j^e A_i^o$). With this assumption, the expectation of metapopulation capacity follows [32]:

$$\lambda = n\bar{M} + \frac{\sigma^2}{\bar{M}} + O\left(\frac{1}{\sqrt{n}}\right), \quad (2.5)$$

where $O(1/\sqrt{n})$ denotes that the error of the estimate decreases fast with the number of total patches. In addition, the variance of the mean metapopulation capacity is bounded by $2\sigma^2$, which is independent of the number of patches. These features make the inference of metapopulation capacity from an incomplete survey feasible.

(c) Analytic argument on patch importance

Recall that the patch importance is given by the eigenvector corresponding to the largest eigenvalue (i.e. metapopulation capacity) of the connectivity matrix. We denote the eigenvector as τ and the importance of patch i as τ_i . At first glance, predicting patch importance appears to be more difficult than predicting metapopulation capacity. One obvious obstacle is, while metapopulation capacity is a global summary of all patches, patch importance is different for each patch. Thus, instead of the quantitative value, we focus on the rank of patch importance among all patches (i.e. whether patch i has a higher importance than patch j).

We assume that the connectivity matrix M is symmetric. Importantly, the leading eigenvector is related to the eigenvalues of the connectivity matrix (whole metapopulation) and its submatrix (sub-metapopulation). Specifically, we have the eigenvector-eigenvalue identity [19]:

$$\tau_i \propto \prod_{k=1}^{n-1} (\lambda(M) - \lambda_k(M_i)), \quad (2.6)$$

where $\lambda(M)$ is the metapopulation capacity, M_i denotes the connectivity matrix corresponding to the metapopulation without patch i and $\lambda_k(M_i)$ denotes the k th highest eigenvalue of the matrix M_i .

We then additionally assume that the connectivity matrix M is a random matrix. While we are not aware of any analytic result on the full distribution of $\lambda_k(M_i)$, the distribution is almost centred around its median. Specifically, the probability of $\lambda_k(M_i)$ deviates from the median decreases super-exponentially [33]. This provides a heuristic justification of why the relative rank of patch importance τ_i is conservative. Importantly, the conserved structure renders the inference of patch importance easier than that of metapopulation capacity, as we do not need to know the total number of patches.

(d) Statistical inference

We provide two different methods to infer the population capacity from sample or survey representing a subset of habitat patches. The first method is regression-based inference. Suppose that we have sampled k patches. We bootstrap resampling with different number of patches (e.g. from 2 to k). We calculate the sub-metapopulation capacity for each sample. We then make use of regression methods, including linear regression and generalized additive models, to predict the metapopulation capacity with the number of patches.

Table 1. This table summarizes the parameters for the full-factorial simulations. For patch area, we have considered two probability distributions, normal and uniform. For patch location, we considered six distributions [34]. The first distribution is a regular grid. The second and third are Poisson processes on homogeneous and inhomogeneous landscapes, respectively. These two distributions describe random locations. The fourth and fifth are Neyman–Scott point process on homogeneous and inhomogeneous landscapes, respectively. These two distributions describe clustered locations. The last one is hard clusters, where the borders between patches are evident and patches are strictly located within a boundary. Electronic supplementary material, appendix B, illustrates these six distributions of patch locations. For dispersal dependency ω of area, we consider positive, null and negative values. All scenarios are observed in empirical studies [25]. For extinction dependency e of area, we consider only null and positive values. For dispersal kernel $f(d_{ij})$ of non-zero patch distance, we consider three different kernels: exponential kernel [7], Gaussian kernel, and non-monotonic kernel [28]. Electronic supplementary material, appendix C, illustrates these three kernels. For self-colonization $f(0)$, we considered both existence of self-colonization or not, as both are present in the literature [7,26].

patch area A_i	detail	parameter value
normal distribution	normal(1, σ)	$\sigma = 0.1, 0.3$
uniform distribution	uniform(0, b)	$b = 1, 10$
patch location		
regular grid	regular grid points	
random location	Poisson point process	
inhomogeneous random location	inhomogeneous Poisson point process	
homogeneous clusters	Neyman–Scott point process	
inhomogeneous clusters	inhomogeneous Neyman–Scott point process	
hard cluster		
area dependence parameter		
dispersal dependency ω		$\omega = -1, 0, 1$
extinction dependency e		$e = 0, 0.2, 1$
dispersal kernel $f(d_{ij})$		
Gaussian kernel for $d_{ij} > 0$	$\exp(-d_{ij}^2/2\xi^2)$	$\xi = 0.1, 1$
exponential kernel for $d_{ij} > 0$	$\exp(-\frac{d_{ij}}{\xi})$	$\xi = 0.1, 1$
non-monotonic kernel for $d_{ij} > 0$	$(\Gamma(\alpha + \beta)/\Gamma(\alpha)\Gamma(\beta))(d_{ij}/\xi)^{\alpha-1}(1 - (d_{ij}/\xi))^{\beta-1}$	$\xi = 0.1, 1, \alpha = \beta = 2$
self-colonization $f(0)$		0,1

The second method is analytic-based inference. This method is inspired by the analytical argument for metapopulation capacity. We can decompose the connectivity matrix as $M = A \odot D$, where $A_{ij} = A_i^c A_j^o$ and $D_{ij} = f(d_{ij}/\xi)$, and \odot is the Hadamard product of the matrix. The matrix D closely satisfies the assumption of a random matrix (off-diagonal elements D_{ij} are independent). We can thus use the asymptotic estimate (equation (2.5)) with correlation for dependency. The effect of dependency caused by A can be corrected by the ratio between the true and inferred estimate of the largest eigenvalue of A . Mathematically, we have the following analytically derived estimate:

$$\hat{\lambda}(M) = \underbrace{\left(n\bar{D} \cdot \bar{A} + \frac{\sigma_A^2}{\bar{A}} + \frac{\sigma_D^2}{\bar{D}} + \frac{\sigma_D^2 \sigma_A^2}{\bar{D} \cdot \bar{A}} \right)}_{\text{estimate under independency}} \cdot \underbrace{\frac{\lambda(A)}{k\bar{A} + \sigma_A^2/\bar{A}}}_{\text{correct for dependency}}, \quad (2.7)$$

where n is the total number of patches, k is the number of sampled patches, \bar{D}^2 and σ_D^2 are the sample mean and variance of the off-diagonal elements of D , \bar{A}^2 and σ_A^2 are the sample mean and variance of the off-diagonal elements of A and $\lambda(A)$ is the largest eigenvalue of the matrix A . Derivation of the estimate (equation (2.7)) can be found in electronic supplementary material, appendix A.

To infer the patch importance, we bootstrap resampling with different number of patches from the sampled patches. We calculate the patch importance τ_i for each patch. We normalize the patch importance by the number of patches k in a resample (i.e. $\tau \rightarrow k\tau$). The rationale is that the null expectation of patch

importance with k patches would be $1/k$. We then compute the average patch importance $\bar{k}\tau$ from the resamples. Details on the regression-based method can be found in electronic supplementary material, appendix A.

(e) Full-factorial simulation

To validate our statistical framework, we consider a factorial combination of metapopulation configurations. Specifically, we consider combinations of different dispersal kernels, distributions of patch area, distributions of patch location, the existence of self-colonization, rates of extinction rates and rates of area dependence of dispersal. This allows us to explore the state-of-the-art modelled complexity of metapopulation configurations. Table 1 presents a summary of the simulation combinations. Electronic supplementary material, appendix B, illustrates the different distributions of patch location, and electronic supplementary material, appendix C, illustrates the different dispersal kernels. Importantly, most simulation combinations would violate the simplifying assumptions in our analytic arguments. For example, when patch areas are heterogeneous, or when the dispersal dependency rate ω does not equal to the extinction dependency rate e , or when patches are not distributed with a homogeneous point process.

We then quantify the prediction accuracy of our statistical approach. For metapopulation capacity, we calculate the quantitative difference between the predicted and the true capacities $(\hat{\lambda} - \lambda)/\lambda$. For patch importance, to assess how well the relative order of patch importance is preserved, we calculate the

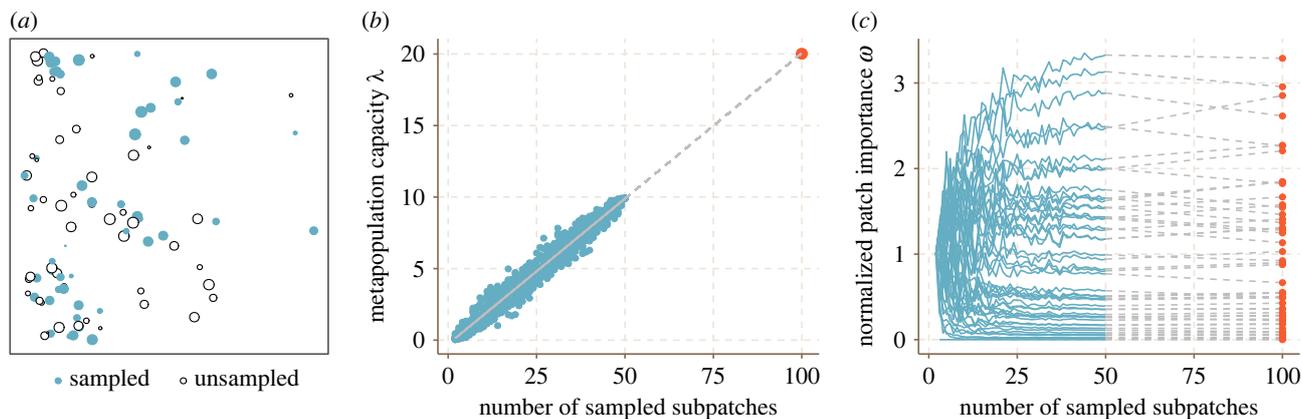


Figure 2. Illustration of inference of metapopulation capacity. (a) A simulated metapopulation with patches located on an inhomogeneous landscape. It has in total 100 patches, where 50 patches are sampled (blue circles) while the other 50 patches are unsampled (white circles). We bootstrap the sampled 50 patches to get an ensemble of sub-metapopulations with different patch numbers. (b) The sampled and the predicted metapopulation capacities. From the sampled metapopulation capacities (blue solid points), we predict the capacity of the whole metapopulation (regression line). The predicted capacity (the end of the regression line) is close to the true capacity of the whole metapopulation (orange point). (c) The sampled patch importance and their relative ranking as increasing numbers of patches are sampled up to 50 patches (blue lines) and the final patch ranking for 100 patches (red points). (Online version in colour.)

Spearman's rank correlation between the predicted and true importance $\text{cor}[(\hat{\tau}_1, \dots, \hat{\tau}_k), (\tau_1, \dots, \tau_k)]$.

(f) Empirical data

We used a dataset of a metapopulation of the endangered mangrove hummingbird (*Amazilia boucardi*) [15]. The metapopulation has 403 patches, and the total habitat area is 351 km². The patches are identified with a 30 m resolution map of global mangrove habitat [35]. The areas and locations of the patches are measured with a high precision. Appendix D in the electronic supplementary material provides the distribution of patch area and pairwise patch distances in this metapopulation.

Huang *et al.* [15] provided additionally the values of the parameters of the metapopulation dynamics. The area dependence of dispersal ω is set to be 1, and the area dependence of extinction ω is set to be 0.5. There exists self-colonization within a patch. The empirically inspired dispersal kernel is the log-sech distribution:

$$f\left(\frac{d_{ij}}{\xi}\right) = \frac{2}{\pi} \arctan\left[\left(\frac{\xi}{d_{ij}}\right)^{\beta-1}\right], \quad (2.8)$$

where the characteristic dispersal distance ξ is set to be 317 m, and the distribution tail thickness β is set to be 1.77. However, there are uncertainties on these parameters. For example, the values of ξ and β are derived from field data on Amazonian passerine birds instead of directly from the mangrove hummingbird [36]. To account for the uncertainty on the parameters [15,37], we additionally consider the factorial combination of the parameter values: ξ ranges from 100 m to 5000 m, β ranges from 1 to 2, ω ranges from 0.5 to 1.5 and ω ranges from 0.5 to 1.5, and whether self-colonization exists.

We then vary the number of sampled patches to test how the sampling effort affects the predictability. Specifically, the percentage of sampled patches from 0.1 (i.e. one tenth of the patches are sampled) to 0.5 (i.e. half of the patches are sampled).

3. Results

We first show an example to illustrate why inference is possible (figure 2). We simulate a metapopulation with 100 patches in a given landscape. We sample only 50 of all patches. We resampled the sampled 50 patches, which provided an ensemble of sub-populations with patch number ranging from 2 to

50. Focusing on metapopulation capacity, we observe a linear structure of how capacity grows with increasing patch numbers sampled. This linear structure is consistent with our analytic arguments. Then by focusing on patch importance, we observe a conserved structure of relative rank with increasing patch numbers sampled. This conserved structure is also consistent with our analytic arguments. These two structures are the inherent constraints of metapopulation dynamics, and they allow us to infer metapopulation persistence with a substantial proportion of unsampled patches.

We then validate our statistical approach using the various metapopulation configurations (figure 3). We first focus on metapopulation capacity. The analytic-based estimator has an average prediction error of -0.025 with 95% confidence interval $(-0.17, 0.08)$ (figure 3a). In comparison, the regression-based estimator has an average error of 0.005 with 95% confidence interval $(-0.21, 0.24)$ (figure 3b). Thus, the analytic-based estimator provides a conservative prediction with lower variance, while the regression-based estimator provides an unbiased prediction with higher variance. We then focus on patch importance (figure 3c). The predicted rank of patch importance has a high correlation with the true rank. The median of the correlation is 0.94 with 95% confidence interval $(0.33, 1)$.

We finally applied our statistical approach to the empirical metapopulation of the mangrove hummingbird (figure 4). We predict metapopulation capacity with different sampling efforts using the analytic-based estimator (figure 4a). The prediction error of metapopulation capacity decreases and saturates with the sampling effort. Specifically, the mean prediction error decreases from -0.05 when sampling 10% patches to -0.007 when sampling 50% patches. This is consistent with simulation data. Results are qualitatively similar when using the regression-based estimator (electronic supplementary material, appendix D). We then predict patch importance using different sampling efforts (figure 4b). In contrast to the normal-like distribution of prediction errors on metapopulation capacity, we observe a two-mode distribution on patch importance. One mode is centred close to a correlation of 1. Specifically, the proportion of samples with correlation higher than 95% increases from 20% when sampling 10% patches to 30%

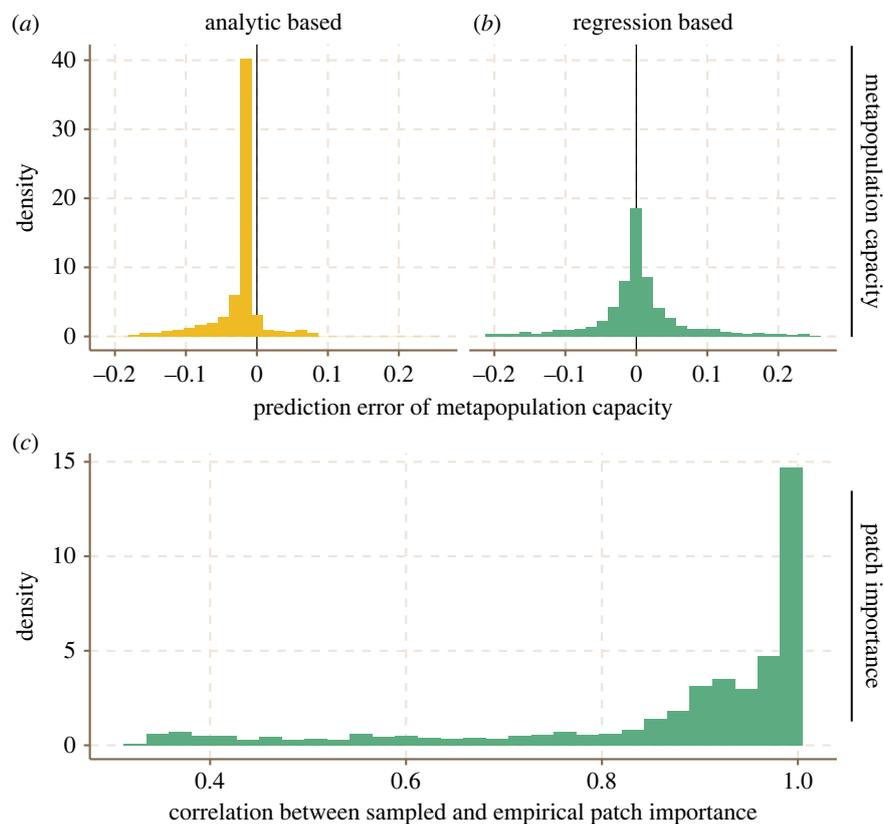


Figure 3. Prediction of metapopulation capacity in full-factorial simulations. (a) and (b) The prediction of metapopulation capacity from incomplete sampling. (a) The analytic-based estimation (equation (2.7)) is used, and (b) the regression-based estimation is used. The vertical axis shows the density of the distribution. The horizontal axis shows the prediction error of the estimation from the true value $((\hat{\lambda} - \lambda)/\lambda)$. If the error is larger than 0, then we have overestimated the capacity and vice versa. The analytic-based estimator generally underestimates the capacity, which can be used as a conservative estimate. In contrast, the regression-based estimator, which can be used as an unbiased estimate. (c) The prediction of relative rank of patch importance from incomplete sampling. The vertical axis shows the density of the distribution. The horizontal axis shows the Spearman's rank correlation between the predicted and true importance $(\text{cor}[(\hat{\tau}_1, \dots, \hat{\tau}_k), (\tau_1, \dots, \tau_k)])$. The majority of the correlation is close to 1, and the correlation is almost never negative. (Online version in colour.)

when sampling 50% patches. Another mode is a normal-like distribution. Specifically, the mean of samples with correlation less than 95% increases from 0.64 to 0.81. In summary, to predict metapopulation capacity requires less sampling effort than to predict patch importance.

4. Discussion

Sampling patches to assess the state of a metapopulation is costly in the field. This logistical constraint prevents a wider adoption of the elegant and rigorous framework of metapopulation persistence in conservation biology. To effectively monitor metapopulation persistence, we need a quick and yet rigorous approach to assist with incomplete surveys of patches. Our study makes a first step in filling this gap between theory and empirical studies. We provide a robust statistical approach to infer metapopulation persistence when only a small proportion of patches are sampled. This statistical approach is justified with analytic arguments and validated with full-factorial simulations and an analysis of an empirical metapopulation.

That we can infer metapopulation persistence from an incomplete survey is not a trivial result. What makes the inference possible is that the special structures in the connectivity matrix are constrained by the metapopulation dynamics. This property is known as *coarse-grainability* in physics [38]. Specifically, the inherent linear structure of metapopulation capacity and conserved structure of patch importance (figure 2). In

contrast, other seemingly similar questions in ecology do not have this property, which renders inference with incomplete data difficult, if not impossible. Technically, to infer metapopulation capacity is to infer the leading eigenvalue from the connectivity matrix. By replacing the connectivity matrix with other matrices, the leading eigenvalue would have other fundamental ecological interpretations. For example, the leading eigenvalue of the Jacobian matrix of community dynamics determines its resilience to perturbations [39,40], while that of the Leslie matrix of age distribution determines the long-term population growth [41]. However, the leading eigenvalue is not coarse-grainable for neither Jacobian nor for Leslie matrix. For Jacobian matrix, while many works have derived the stability criteria for random Jacobian matrix [42–44], the system is not coarse-grainable. To see this, adding a new species can increase or decrease community stability (as there is no sign constraint on the Jacobian matrix), while adding a new patch into a connectivity matrix never decreases metapopulation capacity (as all elements are non-negative in the connectivity matrix).

Our statistical approach offers some guideline of patch sampling in the field. The success of our approach fundamentally requires that we conduct an unbiased survey of patches or a survey with a known bias. The bias here refers to surveys of larger and/or occupied patches, which are sampled with higher probability than other patches. This requirement is not a unique condition of our approach, but general for almost all prediction purposes [45,46]. For example, the empirical sampling bias in a large-scale survey is systematic but quite

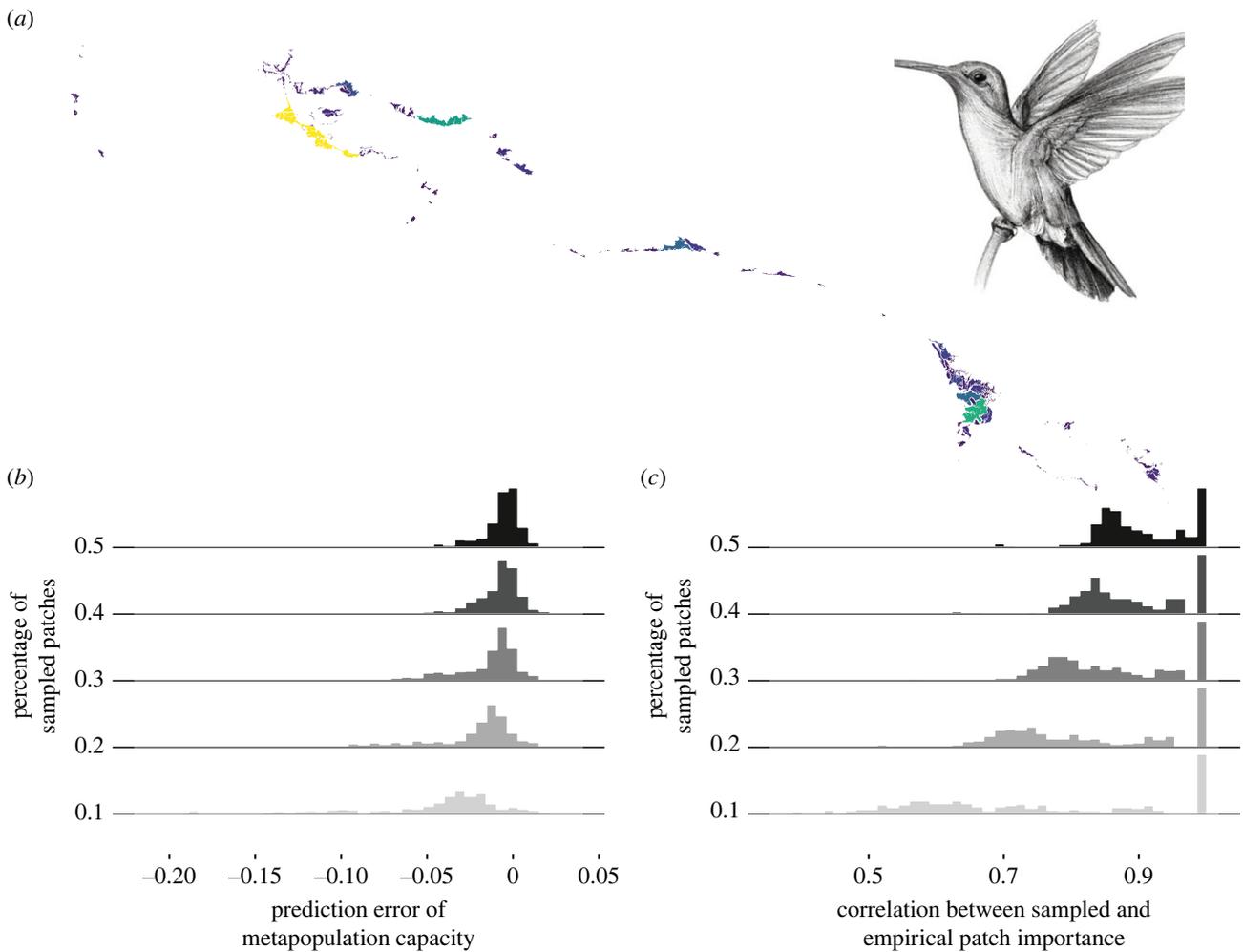


Figure 4. Predicting metapopulation persistence in empirical data. We applied our statistical approach to an empirical metapopulation of the endangered mangrove hummingbird *Amazilia boucardi* [15]. (a) The geographical distributions of the patches. The colour denotes the area size of each patch, where green is the smallest, while yellow is the largest. The clip-art of the hummingbird was made with DALL-E. (b) The prediction error of metapopulation capacity with different sampling efforts. The sampling effort ranges from 10% of all patches to 50% of all patches. The prediction ability increases and saturates with the sampling effort. We use the analytic-based estimator here. Consistent with our simulation results, it generally provides conservative estimations of metapopulation capacity. (c) The prediction of relative rank of patch importance with different sampling efforts. The prediction of the ranks increases with sampling effort, and it does not saturate with sampling of 50% of all patches. (Online version in colour.)

small; however, 2.3 million samples with this small bias would have equivalent statistical power of 400 unbiased samples [47]. Unfortunately, biased sampling is common in empirical studies [48,49]. We often have no idea how biased the sampling is, as they were not originally collected for the purpose of inference we describe here. A silver lining, though, is that sampling bias is more problematic in larger datasets (an empirical metapopulation usually has less than a couple of hundred patches). We simulated how the level of sampling bias affects our prediction ability in electronic supplementary material, appendix F. While this can provide an ad hoc estimation of prediction uncertainty, we believe that a better solution is to adopt a protocol of random sampling or survey. After all, sampling quality trumps sampling quantity for inference purposes. A simple protocol is: for a homogeneous landscape, several small areas are randomly surveyed thoroughly and extrapolated to the whole landscape; and for a heterogeneous landscape, survey locations are randomly chosen within the whole landscape. A mix of the two sampling protocols may be preferred in landscapes with unknown heterogeneity. We suggest further studies on more sophisticated sampling protocols [50].

To link our study to practical metapopulation conservation, another key issue is the practical difficulty in locating patches in

metapopulation surveys. Metapopulation models typically assume that patches have 'hard' boundaries, which separate habitable and non-habitable terrains. While some methods are available to detect such boundaries in empirical landscape [51–53], finding such boundary is still a daunting issue [54–57]. An alternative approach is to consider individual pixels instead of a collection of patches [58–61]. Nonetheless, the lack of a clear boundary renders the patch area, an often ill-defined concept in practice. The silver lining, though, is that the patch area is just a proxy for how habitable the patch is (which we called patch quality when we introduced the general model of metapopulation dynamics; equation (2.1)). Patch quality can be inferred without the need of detecting the patch boundaries. For example, we infer the patch quality from the counted number of individuals per patch [62].

Our study has the opposite focus to many other studies on metapopulation capacity. We focus on inference with unsampled patches, while previous works have mostly focused on habitat destruction or degradation by removing sampled patches [10,11,63,64]. Our analytic arguments apply as well to removing patches (a caveat though is that we would expect a larger prediction error due to a smaller patch number; see electronic supplementary material, appendix E). However, in this

case, as we already have the information (i.e. the whole connectivity matrix), direct simulations might be better suited for this task than our analytic arguments.

Our results increase our understanding on the effects of heterogeneity on metapopulation persistence. For example, equation (2.7) shows that a higher variance in patch distance always increases mean metapopulation capacity, while a higher variance in patch area or quality has context-dependent effect on mean metapopulation capacity. Interestingly, the effects of variance in patch distance agree with the previous results on the lower bound of metapopulation capacity, while the effects of variance in the patch area or quality do not [28].

A potential extension of our approach is to consider other models of metapopulation dynamics. In our study, we have adopted the deterministic metapopulation dynamics (equation (2.1)). An alternative approach is individual-based models of spatially structured populations [65–67]. Earlier studies have suggested strong links between these two different modelling approaches [68]. Another potential extension is to consider metapopulation dynamics with habitat modification [69]. Habitat modification has the potential to increase metapopulation persistence, although it is yet unclear how to quantify metapopulation capacity and patch importance in this model setting.

5. Conclusion

The metapopulation concept is central to conservation science. It is used to predict the consequences of habitat

loss for species persistence in highly fragmented landscapes [14,70,71]. It is also used in the assessment of threatened species and to guide planning for habitat connectivity and restoration [26,72,73].

Our findings have significant implications for conservation organizations needing to rapidly survey metapopulations and assess their long-term persistence with limited financial and logistical means. Being able to make robust inference about persistence from a fraction of the total habitat network offers the potential for rapid monitoring for metapopulation status. We recommend widespread testing of our methods in the lab and in the field to examine their robustness.

Data accessibility. The dataset of the endangered mangrove hummingbirds (*Amazilia boucardi*) is available from <https://doi.org/10.1111/cobi.13364>. The source code to produce the results is available on GitHub at <https://github.com/clsong/ReproduceMetaCap>.

The data are provided in electronic supplementary material [74].

Authors' contributions. C.S.: conceptualization, data curation, formal analysis, investigation and methodology; M.-J.F.: funding acquisition and project administration; A.G.: conceptualization, funding acquisition and project administration.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. M.-J.F. acknowledges the funding of the CRC in Spatial Ecology. A.G. was supported by the Liber Ero Chair in Biodiversity Conservation.

Acknowledgements. The authors thank the editor and two reviewers for their suggestions that improved our paper.

References

- Hanski I. 1999 *Metapopulation ecology*. Oxford, UK: Oxford University Press.
- Levins R. 1969 Some demographic and genetic consequences of environmental heterogeneity for biological control. *Am. Entomol.* **15**, 237–240. (doi:10.1093/besa/15.3.237)
- Hanski I, Gilpin M. 1991 Metapopulation dynamics: brief history and conceptual domain. *Biol. J. Linnean Soc.* **42**, 3–16. (doi:10.1111/j.1095-8312.1991.tb00548.x)
- Gonzalez A, Lawton JH, Gilbert F, Blackburn TM, Evans-Freke I. 1998 Metapopulation dynamics, abundance, and distribution in a microecosystem. *Science* **281**, 2045–2047. (doi:10.1126/science.281.5385.2045)
- Fahrig L. 2003 Effects of habitat fragmentation on biodiversity. *Ann. Rev. Ecol. Evol. Syst.* **34**, 487–515. (doi:10.1146/annurev.ecolsys.34.011802.132419)
- Haddad NM *et al.* 2015 Habitat fragmentation and its lasting impact on earth's ecosystems. *Sci. Adv.* **1**, e1500052. (doi:10.1126/sciadv.1500052)
- Hanski I, Ovaskainen O. 2000 The metapopulation capacity of a fragmented landscape. *Nature* **404**, 755–758. (doi:10.1038/35008063)
- Hanski I, Ovaskainen O. 2003 Metapopulation theory for fragmented landscapes. *Theor. Popul. Biol.* **64**, 119–127. (doi:10.1016/S0040-5809(03)00022-4)
- Ovaskainen O, Hanski I. 2001 Spatially structured metapopulation models: global and local assessment of metapopulation capacity. *Theor. Popul. Biol.* **60**, 281–302. (doi:10.1006/tpbi.2001.1548)
- Ovaskainen O. 2003 Habitat destruction, habitat restoration and eigenvector–eigenvalue relations. *Math. Biosci.* **181**, 165–176. (doi:10.1016/S0025-5564(02)00150-5)
- Häussler J, Barabás G, Eklöf A. 2020 A Bayesian network approach to trophic metacommunities shows that habitat loss accelerates top species extinctions. *Ecol. Lett.* **23**, 1849–1861. (doi:10.1111/ele.13607)
- Wang S, Brose U, van Nouhuys S, Holt RD, Loreau M. 2021 Metapopulation capacity determines food chain length in fragmented landscapes. *Proc. Natl Acad. Sci. USA* **118**, e2102733118. (doi:10.1073/pnas.2102733118)
- Dallas TA, Saastamoinen M, Schulz T, Ovaskainen O. 2020 The relative importance of local and regional processes to metapopulation dynamics. *J. Anim. Ecol.* **89**, 884–896. (doi:10.1111/1365-2656.13141)
- Hanski I, Schulz T, Wong SC, Ahola V, Ruokolainen A, Ojanen SP. 2017 Ecological and genetic basis of metapopulation persistence of the glanville fritillary butterfly in fragmented landscapes. *Nat. Commun.* **8**, 1–11. (doi:10.1038/ncomms14504)
- Huang R, Pimm SL, Giri C. 2020 Using metapopulation theory for practical conservation of mangrove endemic birds. *Conserv. Biol.* **34**, 266–275. (doi:10.1111/cobi.13364)
- Matter SF, Goff J, Keyghobadi N, Roland J. 2020 Direct estimates of metapopulation capacity from dispersal show high interannual variability, but little effect of recent forest encroachment on network persistence. *Landsc. Ecol.* **35**, 675–688. (doi:10.1007/s10980-020-00972-3)
- Van Moorter B, Kivimäki I, Noack A, Devooght R, Panzacchi M, Hall KR, Leleux P, Sauerens M. In press. Accelerating advances in landscape connectivity modelling with the conscape library. *Methods Ecol. Evol.*
- Horn RA, Johnson CR 2012 *Matrix analysis*. Cambridge, UK: Cambridge University Press.
- Denton P, Parke S, Tao T, Zhang X. 2022 Eigenvectors from eigenvalues: a survey of a basic identity in linear algebra. *Bull. Am. Math. Soc.* **59**, 31–58. (doi:10.1090/bull/1722)
- Albert EM, Fortuna MA, Godoy JA, Bascompte J. 2013 Assessing the robustness of networks of spatial genetic variation. *Ecol. Lett.* **16**, 86–93. (doi:10.1111/ele.12061)
- Hanski I. 1998 Metapopulation dynamics. *Nature* **396**, 41–49. (doi:10.1038/23876)
- Hanski I, Gyllenberg M. 1997 Uniting two general patterns in the distribution of species. *Science* **275**, 397–400. (doi:10.1126/science.275.5298.397)
- Gyllenberg M, Hanski I. 1997 Habitat deterioration, habitat destruction, and metapopulation persistence in a heterogenous landscape. *Theor. Popul. Biol.* **52**, 198–215. (doi:10.1006/tpbi.1997.1333)
- Moilanen A, Hanski I. 1998 Metapopulation dynamics: effects of habitat quality and landscape

- structure. *Ecology* **79**, 2503–2515. (doi:10.1890/0012-9658(1998)079[2503:MDEOHQ]2.0.CO;2)
25. Wang S, Altermatt F. 2019 Metapopulations revisited: the area-dependence of dispersal matters. *Ecology* **100**, e02792. (doi:10.1002/ecy.2792)
 26. Schnell JK, Harris GM, Pimm SL, Russell GJ. 2013 Estimating extinction risk with metapopulation models of large-scale fragmentation. *Conserv. Biol.* **27**, 520–530. (doi:10.1111/cobi.12047)
 27. Hastings A *et al.* 2005 The spatial spread of invasions: new developments in theory and evidence. *Ecol. Lett.* **8**, 91–101. (doi:10.1111/j.1461-0248.2004.00687.x)
 28. Grilli J, Barabás G, Allesina S. 2015 Metapopulation persistence in random fragmented landscapes. *PLoS Comput. Biol.* **11**, e1004251. (doi:10.1371/journal.pcbi.1004251)
 29. Meyer CD 2000 *Matrix analysis and applied linear algebra*, vol. 71. Philadelphia, PA: SIAM.
 30. Hwang S-G. 2004 Cauchy's interlace theorem for eigenvalues of hermitian matrices. *Am. Math. Mon.* **111**, 157–159. (doi:10.1080/00029890.2004.11920060)
 31. Walker SG, Van Mieghem P. 2008 On lower bounds for the largest eigenvalue of a symmetric matrix. *Linear Algebra and its Appl.* **429**, 519–526. (doi:10.1016/j.laa.2008.03.007)
 32. Füredi Z, Komlós J. 1981 The eigenvalues of random symmetric matrices. *Combinatorica* **1**, 233–241. (doi:10.1007/BF02579329)
 33. Alon N, Krivelevich M, Vu VH. 2002 On the concentration of eigenvalues of random symmetric matrices. *Israel J. Math.* **131**, 259–267. (doi:10.1007/BF02785860)
 34. Dale MR, Fortin M-J 2014 *Spatial analysis: a guide for ecologists*. Cambridge, MA: Cambridge University Press.
 35. Giri C, Ochieng E, Tieszen LL, Zhu Z, Singh A, Loveland T, Masek J, Duke N. 2011 Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecol. Biogeogr.* **20**, 154–159. (doi:10.1111/j.1466-8238.2010.00584.x)
 36. Van Houtan KS, Pimm SL, Halley JM, Bierregaard Jr RO, Lovejoy TE. 2007 Dispersal of amazonian birds in continuous and fragmented forest. *Ecol. Lett.* **10**, 219–229. (doi:10.1111/j.1461-0248.2007.01004.x)
 37. Martin AE, Fahrig L. 2018 Habitat specialist birds disperse farther and are more migratory than habitat generalist birds. *Ecology* **99**, 2058–2066. (doi:10.1002/ecy.2428)
 38. Moran J, Tikhonov M. 2022 Defining coarse-grainability in a model of structured microbial ecosystems. *Phys. Rev. X* **12**, 021038. (doi:10.1103/PhysRevX.12.021038)
 39. Case TJ. 1999 Illustrated guide to theoretical ecology. *Ecology* **80**, 2848–2848.
 40. Song C, Saavedra S. 2021 Bridging parametric and nonparametric measures of species interactions unveils new insights of non-equilibrium dynamics. *Oikos* **130**, 1027–1034. (doi:10.1111/oik.08060)
 41. Caswell H 2000 *Matrix population models*, vol. 1. Sunderland, MA: Sinauer.
 42. Allesina S, Tang S. 2015 The stability–complexity relationship at age 40: a random matrix perspective. *Popul. Ecol.* **57**, 63–75. (doi:10.1007/s10144-014-0471-0)
 43. May RM. 1972 Will a large complex system be stable? *Nature* **238**, 413–414. (doi:10.1038/238413a0)
 44. Song C, Saavedra S. 2018 Will a small randomly assembled community be feasible and stable? *Ecology* **99**, 743–751. (doi:10.1002/ecy.2125)
 45. Clark AT *et al.* 2021 General statistical scaling laws for stability in ecological systems. *Ecol. Lett.* **24**, 1474–1486. (doi:10.1111/ele.13760)
 46. Thompson SK 2012 *Sampling*, vol. 755. Hoboken, NJ: John Wiley & Sons.
 47. Meng X-L. 2018 Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 US presidential election. *Ann. Appl. Stat.* **12**, 685–726. (doi:10.1214/18-AOS11615F)
 48. Fahrig L. 2020 Why do several small patches hold more species than few large patches? *Global Ecol. Biogeogr.* **29**, 615–628. (doi:10.1111/geb.13059)
 49. Riva F, Fahrig L. 2022 The disproportionately high value of small patches for biodiversity conservation. *Conserv. Lett.* **15**, e12881. (doi:10.1111/conl.12881)
 50. Wang S, Loreau M, Arnoldi J-F, Fang J, Rahman KA, Tao S, de Mazancourt C. 2017 An invariability–area relationship sheds new light on the spatial scaling of ecological stability. *Nat. Commun.* **8**, 1–8. (doi:10.1038/ncomms15211)
 51. Akçakaya HR. 2000 Viability analyses with habitat-based metapopulation models. *Popul. Ecol.* **42**, 45–53. (doi:10.1007/s101440050043)
 52. Akçakaya HR, McCarthy MA, Pearce JL. 1995 Linking landscape data with population viability analysis: management options for the helmeted honeyeater lichenostomus melanops cassidix. *Biol. Conserv.* **73**, 169–176. (doi:10.1016/0006-3207(95)00054-8)
 53. Urban D, Keitt T. 2001 Landscape connectivity: a graph-theoretic perspective. *Ecology* **82**, 1205–1218. (doi:10.1890/0012-9658(2001)082[1205:LCAGTP]2.0.CO;2)
 54. Fahrig L. 2007 Landscape heterogeneity and metapopulation dynamics. *Key topics and perspectives in landscape ecology* (eds J Wu, RJ Hobbs), pp. 78–89. Cambridge, UK: Cambridge University Press.
 55. Fahrig L, Nuttle WK. 2005 Population ecology in spatially heterogeneous environments. In *Ecosystem function in heterogeneous landscapes* (eds GM Lovett, MG Turner, CG Jones, KC Weathers), pp. 95–118. New York, NY: Academic Press.
 56. With KA. 2004 Metapopulation dynamics: perspectives from landscape ecology. In *Ecology, genetics and evolution of metapopulations* (ed. I Hanski), pp. 23–44. New York, NY: Elsevier.
 57. Wu J, Hobbs R. 2002 Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. *Landscape Ecol.* **17**, 355–365. (doi:10.1023/A:1020561630963)
 58. Bertuzzo E, Rodriguez-Iturbe I, Rinaldo A. 2015 Metapopulation capacity of evolving fluvial landscapes. *Water Resour. Res.* **51**, 2696–2706. (doi:10.1002/2015WR016946)
 59. García-Valdés R, Zavala MA, Araujo MB, Purves DW. 2013 Chasing a moving target: projecting climate change-induced shifts in non-equilibrium tree species distributions. *J. Ecol.* **101**, 441–453.
 60. Giezdanner J, Bertuzzo E, Pasetto D, Guisan A, Rinaldo A. 2019 A minimalist model of extinction and range dynamics of virtual mountain species driven by warming temperatures. *PLoS ONE* **14**, e0213775. (doi:10.1371/journal.pone.0213775)
 61. Purves DW, Zavala MA, Ogle K, Prieto F, Benayas JMR. 2007 Environmental heterogeneity, bird-mediated directed dispersal, and oak woodland dynamics in mediterranean spain. *Ecol. Monogr.* **77**, 77–97. (doi:10.1890/05-1923)
 62. Giezdanner J, Pasetto D, Perez-Saez J, Cerrato C, Viterbi R, Terzagio S, Palazzi E, Rinaldo A. 2020 Earth and field observations underpin metapopulation dynamics in complex landscapes: near-term study on carabids. *Proc. Natl Acad. Sci. USA* **117**, 12 877–12 884. (doi:10.1073/pnas.1919580117)
 63. DeWoody YD, Feng Z, Swihart RK. 2005 Merging spatial and temporal structure within a metapopulation model. *Am. Nat.* **166**, 42–55. (doi:10.1086/430639)
 64. Ovaskainen O, Sato K, Bascompte J, Hanski I. 2002 Metapopulation models for extinction threshold in spatially correlated landscapes. *J. Theor. Biol.* **215**, 95–108. (doi:10.1006/jtbi.2001.2502)
 65. Gaona P, Ferreras P, Delibes M. 1998 Dynamics and viability of a metapopulation of the endangered iberian lynx (*lynx pardinus*). *Ecol. Monogr.* **68**, 349–370. (doi:10.1890/0012-9615(1998)068[0349:DAVOAM]2.0.CO;2)
 66. King AW, With KA. 2002 Dispersal success on spatially structured landscapes: when do spatial pattern and dispersal behavior really matter? *Ecol. Modell.* **147**, 23–39. (doi:10.1016/S0304-3800(01)00400-8)
 67. Pettifor RA, Caldow RW, Rowcliffe J, Goss-Custard J, Black JM, Hodder KH, Houston A, Lang A, Webb J. 2000 Spatially explicit, individual-based, behavioural models of the annual cycle of two migratory goose populations. *J. Appl. Ecol.* **37**, 103–135. (doi:10.1046/j.1365-2664.2000.00536.x)
 68. Ovaskainen O, Hanski I. 2004 From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. *Am. Nat.* **164**, 364–377. (doi:10.1086/423151)
 69. Miller ZR, Allesina S. 2021 Metapopulations with habitat modification. *Proc. Natl Acad. Sci. USA* **118**, e2109896118. (doi:10.1073/pnas.2109896118)
 70. Earn DJ, Levin SA, Rohani P. 2000 Coherence and conservation. *Science* **290**, 1360–1364. (doi:10.1126/science.290.5495.1360)
 71. Hanski I, Zurita GA, Bellocq MI, Rybicki J. 2013 Species–fragmented area relationship. *Proc. Natl Acad. Sci. USA* **110**, 12 715–12 720. (doi:10.1073/pnas.1311491110)
 72. Huang R. 2019 Conservation through population assessments across variable landscapes. PhD thesis, Duke University.
 73. Schnell J, Russell G, Harris G, Pimm S. 2010 Metapopulation capacity with self-colonization: finding the best patches in fragmented habitats. *Nat. Prec.* (doi:10.1038/npre.2010.5356.1)
 74. Song C, Fortin M-J, Gonzalez A. 2022 Metapopulation persistence can be inferred from incomplete surveys. Figshare. (doi:10.6084/m9.figshare.c.6326400)