An Exact Coarse-Graining Consistent Metapopulation Model

Damiano Pasetto[1,*], Jonathan Giezendanner[2,*], Andrea Rinaldo[3]

[1]Ca' Foscari University of Venice, IT
[2]Massachusetts Institute of Technology (MIT), USA
[3]Laboratory of Ecohydrology (ECHO), Swiss Federal Institute of Technology Lausanne (EPFL), CH

[*]Authors were at EPFL when the work was carried out

Corresponding author: Damiano Pasetto (damiano.pasetto@unive.it)

Different levels of coarse-graining are of paramount importance to understand how conclusions drawn from local studies can be made general and extrapolated to larger regions. We here investigate how consist metapopulation model are when considering different resolutions of the landscape matrix, i.e. different levels of coarse-graining. A formulation of the metapopulation model, taking into consideration a dispersal volume depending on the grid-cell area is proposed. This first derivation from the classical model is then extended in order to account for intra-cellular colonization (called self-colonization), which permits an aggregation of cells, as typically done with coarse-graining. The consistency of the proposed model formulations is then investigated, first, in a flat homogeneous domain, where the results are independent of the landscape structure. In a second step, a complex landscape structure is introduced in the form of an elevation field, the Gran Paradiso National Park. The results suggest that the proposed formulations of the model are independent of coarse-graining in the flat domain. This appears to be the case in the park as well, up to a certain level of coarse-graining, where loss of information on the landscape matrix becomes too severe. In all considered levels of coarse-graining the proposed formulation of the model outperforms the previously proposed models in terms of convergence of the results.

1 Introduction

Spatially-explicit ecological models, such as species distribution and metapopulation models, are nowadays frequently adopted to investigate the relationship between habitat properties and species diversity and richness (Giezendanner, 2020). In fact, the recently increased availability of high-resolution remote sensing maps is allowing to characterize in detail the habitat features that might determine the suitability of a species to a certain environment. The grain-size at which a species is meant to be modeled has to be carefully chosen, as the resulting species distribution may vary greatly (Kirchheimer et al., 2016). This is especially true for models calibrated at a certain grain and extrapolated to different grain sizes and extends (Keil et al., 2013).

Issues on how one should theoretically address the problem have long been considered. For example, whenever assessments of risk to biodiversity need to rely on spatial distributions of species and ecosystems, range-size metrics, quite sensitive to measurement scale, must be used extensively (Keith et al., 2018). One of the key measures in these assessments is the area of occupancy (AOO). Scale issues emerge therefore, prompting proposals to measure them at different scales based on the shape of the distribution or ecological characteristics of the biota. Despite their dominant role in the description of endangered species, decades, appropriate spatial scales of AOO for predicting risks of species' extinction or ecosystem collapse remain an active subject of research (Keith et al., 2018). By using stochastic simulation models to explore risks to ecosystems and species in complex landscapes, Keith et al. (2018) showed that area of occupancy proves an accurate predictor of risk in most cases only when measured with grid cells 0.1–1.0 times the largest plausible area threatened by an event. Estimates of AOO at relatively coarse scales thus prove better predictors of risk than finer-scale estimates. Because the optimal scale for modeling the relevant dynamics depend on the spatial scales of threats more than the shape or size of biotic distributions, appreciable potential for grid-measurement errors exist (Keith et al., 2018). Neutralizing geometric uncertainty embedding effective scaling procedures for assessing risks posed by landscape-scale threats to species and ecosystems is thus definitely relevant.

Hanski's metapopulation model (Hanski and Gilpin, 1991) lends itself to the study of the above issue. It is based on the interplay between colonization and extinction processes, where the colonization of an empty patch depends on the occupancy probability and distance of surrounding patches. Such a model has been used in literature to simulate the distribution of different species, from butterflies to birds (e.g., Schnell et al. (2012)). The connectivity among patches is the main driver of the model and the landscape structure thus plays a fundamental role. Due to computational limitations, it is not always feasible to calibrate the model, say by comparing with data on habitat gathered by a digital terrain model over large-scales Giezendanner et al. (2019), at the finest possible grid resolution. It is therefore important to understand how modeling results, such as species presence, vary depending on grain and, in particular, if these results are due to numerical errors introduced by the model or to the loss of information due to the upscale of the landscape structure.

While with different levels of coarse-graining certain quantities, such as the mean elevation, are preserved, Rodriguez-Iturbe and Rinaldo (2001) demonstrate that the connectivity properties of a landscape may drastically change under progressive coarse-graining. In particular, drainage directions (i.e. the directions of topographic gradients) change dramatically with the level of coarsening, even if the mean elevation is preserved (Rodriguez-Iturbe and Rinaldo, 2001). This has to be considered because gravity effects affect decisively species dispersal, potentially creating barriers and/or limitations (Colwell and Hurtt, 1994; Colwell and Lees, 2000). Additionally, Palmer and White (1994) showed that "*species-richness patterns were neither self-similar nor hierarchical*". In simpler terms, different regions might be connected or disconnected with different grid approximations of a given underlying landscape. Only scale-free distributions, epitomized by power laws, tare invariant under coarse-graining (i.e. $p(bx) = g(b)p(x)$ where b is a scale factor), and it cannot be postulated that such characters are universal in species distributions – on the contrary, they almost never are. Patches with previously non-existing connections might be averaged into one larger region, allowing for interactions between populations that were not possible at a different scale. Small patches of habitat suitable for a focus species might simply vanish with scaling. Overall, connections between patches might be neglected, rendering population migration harder, and therefor changing the outcome of the simulation.

Here, we study how different levels of coarse-graining influence the spatial occupancy of a set of virtual species invading the Gran Paradiso National Park region when simulated by a gridded metapopulation model (Hanski, 1998; Fahrig and Nuttle, 2005; Fahrig, 2007; Purves et al., 2007; Rybicki and Hanski, 2013; Giezendanner et al., 2020). In particular, we propose a formulation of the metapopulation model where the numerical errors resulting from coarse-graining are to most parts limited to the spatial discretization of the dispersal process. This guarantees that the differences in the results are mostly due to the information on the landscape structure lost because of the coarse-graining, even though these are brought to a minimum as well.

Rybicki and Hanski (2013) and Giezendanner et al. (2019, 2021) have suggested that the landscape properties can be considered within the metapopulation model through a suitable fitness function, which re-scales the colonization and extinction parameters on the basis of how suitable a patch is for a given species. In this study, the elevation field is considered as the sole landscape feature relevant to highlight the scaling properties of the metapopulation model. This is also justified by the recent works of Bertuzzo et al. (2016), which show how terrain topology and elevation constitutes fundamental indicators of habitat suitability that can be considered as a proxy of other habitat properties. In fact, topology and elevation have an important contribution in shaping the habitat-connectivity and, thus, the population dynamics. By subdividing the landscape into cells, Bertuzzo et al. (2016) show that cells with low connectivity to cells with similar altitude - or available in a suitable range for given species - tend to be less populated, hence display a low species diversity. Connectivity is, in this case, defined as a function of the path between cells, specifically focusing on elevation gradients and distance. Peaks, as well as valleys, seem to prevent population movement, as species most suited to inhabit these regions would have to cross through unfavorable terrain to propagate. Connectivity, as defined here, is strongly dependent on the underlying terrain, which needs to be sampled from the environment. In a raster-like approach, the size of the cells is thus a crucial variable to capture these connectivities (Palmer and White, 1994).

2 Materials and Methods

2.1 Metapopulation model

The model considered in this study is based on Hanski's metapopulation model (Hanski and Gyllenberg, 1997), adapted to a raster-like spatial structure (Purves et al., 2007) ideal to incorporate Earth Observation (EO) data (Pasetto et al., 2018; Zaoli et al., 2019; Giezendanner et al., 2020). In the metapopulation model, given an initial spatial distribution of probabilities of occupancy $p_{i,0}$ (i =1,. . . ,n, with n number of cells), the variations in the occupancy are governed by the following system of ordinary differential equations (ODEs):

$$\begin{cases} \frac{dp_i(t)}{dt} = C_i(t)\,(1 - p_i(t)) - E_i p_i(t) \\ p_i(0) = p_{i,0} \end{cases} \tag{1}$$

where E_i and $C_i(t)$ are the extinction and colonization affecting the cell i. The colonization pressure of a species from cell j to cell i is expressed by the product of four quantities: i) the colonization rate c [T^{-1}]; ii) the probability of presence in cell j, $p_j(t)$; iii) a term representing the dispersal kernel, K_{ij}; iv) and a term representing the quality of cell j, q_j^c [-]. The colonization term $C_i(t)$ in 1 is thus obtained by the sum of the colonization contributions from all cells:

$$C_i(t) = c \sum_{\substack{j=1 \\ j \neq i}}^{n} K_{ij} q_j^c p_j(t) \tag{2}$$

Similarly to the approach in Rybicki and Hanski (2013), the dispersal kernel is described using an exponential function:

$$K(d) = r \, \exp\left(-\frac{d}{D}\right) \tag{3}$$

where D [L] is the species dispersion, d [L] is the distance between two points, and r [L^2] is a normalization factor, typically computed imposing that the integral of the dispersal kernel is 1 on an infinite domain:

$$\int_{\mathbb{R}^2} K(d) = 1 \tag{4}$$

which results in $r = \frac{1}{2\pi D^2}$ (Rybicki and Hanski, 2013).

There exist several ways to discretize the dispersal kernel on the given domain (see, e.g., Keith et al. (2018)). Given two cells i and j and a distance d_{ij}, a straightforward approach consists in assigning $K_{ij} = K(d_{ij})$. In this way, however, the sum of K_{ij} on the domain cells changes with the grid resolution, thus implying a different evaluation of the colonization pressure in each cell (because C_i in equation 2 depends upon the sum of K_{ij}). A different approach descends by imposing that the sum of the discretized kernel values approximate the integral of the kernel, thus satisfying the property in equation 4. A simple numerical approximation consists in setting $K_{ij} = A_j K(d_{ij})$. In the metapopulation model this means that the single contributions of cell j to the colonization of cell i depends on the area of the source cell j. Section 2.3 analyses more in detail the advantages of this choice.

The extinction term in equation 1 is represented by an extinction rate e [T^{-1}] scaled over a term q_i^e [-] representing the quality of cell i: $E_i = \frac{e}{q_i^e}$. In the original metapopulation model by Hanski (1998), q_i^c and q_i^e depend on the area A_i [L^2] of patch i:

$$q_i^c = A_i^{\alpha_c} \quad ; \quad q_i^e = A_i^{\alpha_e}, \tag{5}$$

where $\alpha_c \geq 0$ and $\alpha_e \geq 0$. In more complex formulations (see e.g., Rybicki and Hanski (2013); Bertuzzo et al. (2015, 2016); Giezendanner et al. (2019)) q_i^c and q_i^e are obtained through a fitness function $f(x)$ describing the suitability of the landscape features for the species: $q_i^c = q_i^e = f_i$, with f_i being an approximation of the fitness on cell i. For example, considering only the impact of elevation as a proxy of temperature (Giezendanner et al., 2019), one may have:

$$f_i \propto \exp\left(\frac{z_{opt} - z_i}{\sigma}\right), \tag{6}$$

where z_i is the elevation of cell i (obtained, for example, from a DEM), and z_{opt} and σ are species-dependent parameters that represent the optimal elevation for the species and the niche width around the optimal elevation.

2.2 Graining problem

The described metapopulation model was designed to represent the species dynamics on an ensemble of disjointed patches of different sizes, representing the focus species' suitable areas. Changes in the model domain mainly consisted in different number of patches and in different spatial configurations.

The application of the metapopulation model on a regular grid is advantageous to well represent the landscape features in all the domain and to approximate the space-continuity of the dispersal process of the offsprings. In this framework, it is important to identify the suitable grain size at which modeling the extinction, colonization and dispersal processes of the species of interest. Intuitively, results obtained on more refined grid (i.e. the grains) should yield a higher level of accuracy than those on coarse grids, because the dispersion process and the landscape features are better described when using finer discretization. However, this approach has two main limitations: i) the computational cost associated with the numerical solution of the model equation rapidly increases with the cell numbers, and ii) the data available to describe the landscape features have a fixed resolution (e.g., the resolution of remote sensing products), and using a finer grid in the metapopulation model without improving the resolution of the inputs might not add any improvement in the accuracy of the solution. For these reasons, coarser grains are frequently adopted in applications, seeking for a compromise between computational complexity and solution accuracy.

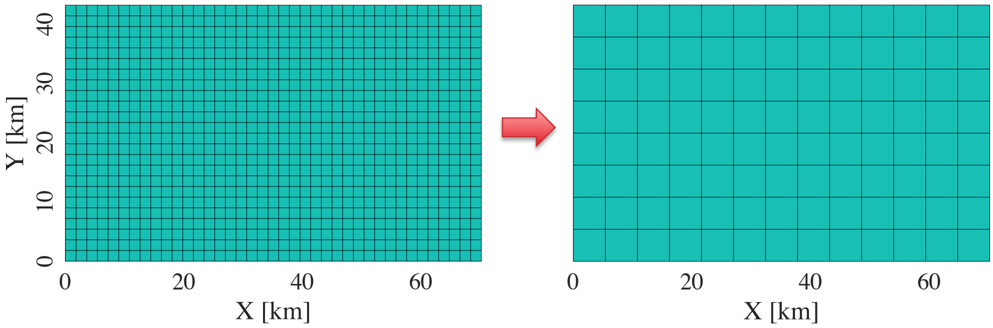

Figure 1: Example of coarse-graining a landscape. In the coarse-graining problem, we are interested to assess the changes in the solution of the metapopulation model when changing the grain size in the domain, as depicted, for example, in this figure.

However, it is fundamental to quantify the error in the metapopulation solution introduced by the process of coarse-graining. Here, we distinguish two type of errors introduced by coarse-graining (see Fig. 1):

1. numerical errors due to the discretization of the dispersal process on coarser grids;
2. numerical errors due to the loss of information in the input data, in particular in those used for the characterization of the landscape (e.g., the DEM in the fitness function 6, as depicted in figure 2).

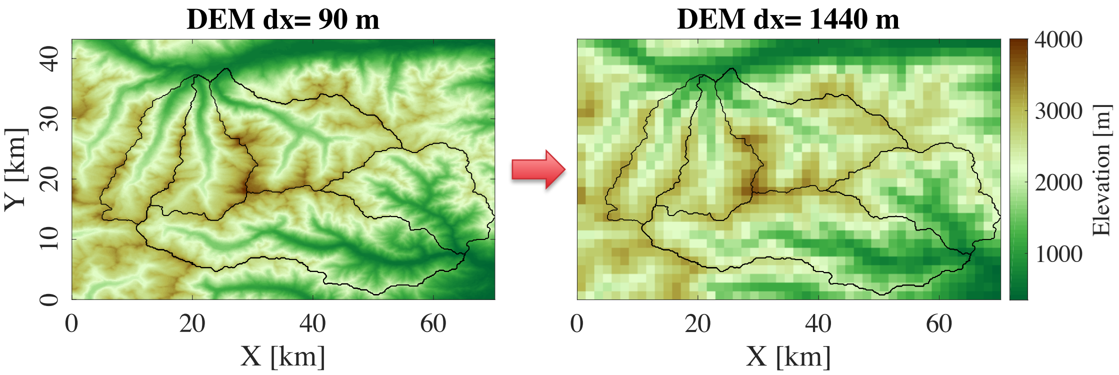

Figure 2: Loss in information in landscape structure with coarse-graining: example on the GPNP. Changes in the metapopulation model solution after changing the grain size might be introduced by numerical errors and/or by the upscaling of the data describing the landscape features, as depicted in this example for the DEM of the GPNP passing from a grain size of 90 m to 1440 m.

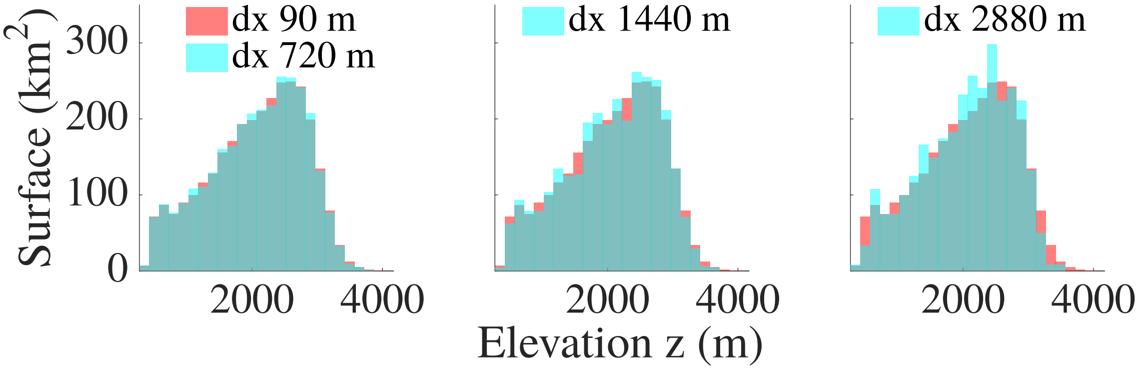

Figure 3: Behavior of the hypsographic curves with coarse-graining. Hypsographic curve of three rescaled resolution and compared to the finest resolution (dx = 90 m). Note that the frequency of the quantity of area available at each elevation changes with coarse-graining, and that the average remains the same.

2.3 A metapopulation model consistent with coarse-graining

Before quantifying the impact of coarse-graining on the accuracy of the numerical solution, it is important to understand the properties of the model solution that should be preserved when using two different grain sizes.

For simplicity, we consider here the rectangular domain depicted in Fig. 4. We want to use the metapopulation model to describe the probabilities of occupancy in the finer cells, termed p_i, $i = 1, \dots, 8$, and those in the coarser cells, termed p_{B_1} and p_{B_2}. Intuitively, at every computational time the probabilities at the coarser grains, should correspond to the average of the probabilities at the finer grains:

$$p_{B_1}(t) = \frac{1}{4}\sum_{i=1}^{4} p_i(t) \quad ; \quad p_{B_2}(t) = \frac{1}{4}\sum_{i=5}^{8} p_i(t). \tag{7}$$

It is natural to question whether this property is satisfied by the solution of two metapopulation models that correspond to the two grain sizes.

Adapting equation 1 to the finer grid we have:

$$\frac{dp_i}{dt} = \left(c\sum_{j=1}^{8} K_{ij}(A_1) q^c(A_1) p_j(t) \right) (1 - p_i(t)) - \frac{e}{q^e(A_1)} p_i(t) \quad , \quad i = 1, \dots, 8 \tag{8}$$

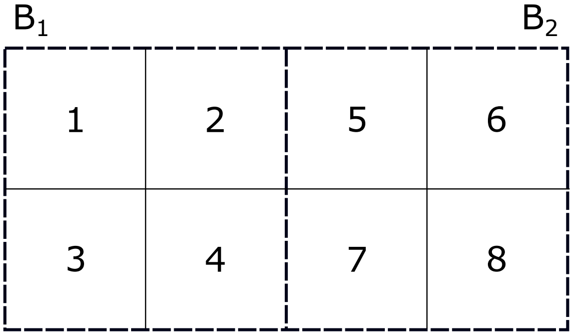

Figure 4: Close up on coarse-graining. Example of a domain having eight squared cells (called cells $1, \dots, 8$) having area A_1, upscaled to the two wider cells called B_1, B_2, of area $A_2 = 4A_1$.

where we explicitly represented the possible dependence of q^c, q^e, and K_{ij} by the area. In an analogous way, adapting equation 1 to the coarser grid yields:

$$\begin{cases} \frac{dp_{B_1}(t)}{dt} = \left(cK_{B_1B_1}(A_2)q^c(A_2)p_{B_1}(t) + cK_{B_1B_2}(A_2)q^c(A_2)p_{B_2}(t)\right)(1 - p_{B_1}(t)) - \frac{e}{q^e(A_2)}p_{B_1}(t) \\ \frac{dp_{B_2}(t)}{dt} = \left(cK_{B_1B_2}(A_2)q^c(A_2)p_{B_1}(t) + cK_{B_2B_2}(A_2)q^c(A_2)p_{B_2}(t)\right)(1 - p_{B_2}(t)) - \frac{e}{q^e(A_2)}p_{B_2}(t) \end{cases} \tag{9}$$

Note that the parameters defining the species, i.e. the rates of extinction e, colonization c and the dispersal D have the same values in equations 8 and 9, since they do not depend on the particular grain size at which we discretize the problem. The same property holds true for the parameters in the fitness, such as z_{opt} and σ in equation 6.

In the following it is shown that, under these hypotheses, the terms q^c and q^e should be independent of the area of the cells in order to satisfy the scaling property in equation 7. Considering a species that persists in this domain (which exists for suitable parameters e, c and D), the stationary solution has a homogeneous probability of occupancy in each cell, due to the homogeneous area of the cells. We call $\tilde{p}$ this probability of occupancy at steady state, $\tilde{p} \neq 0$, and, for (7), we have $\tilde{p} = p_A = p_B = p_i$, $i = 1, \dots, 8$.

We consider equation 8 at the stationary condition, which is constant in time, and thus the temporal derivatives in (8) are null. The sum of the first four equations results in:

$$cq^c(A_1)\left(\sum_{i=1}^{4}\sum_{j=1}^{4}K_{ij}(A_1) + \sum_{i=1}^{4}\sum_{j=5}^{8}K_{ij}(A_1)\right)(1 - \tilde{p}) = 4\frac{e}{q^e(A_1)}, \tag{10}$$

which is strictly connected to the stationary condition of the first equation in 9:

$$cq^c(A_2)\left(K_{B_1B_1}(A_2) + K_{B_1B_2}(A_2)\right)(1 - \tilde{p}) = \frac{e}{q^e(A_2)}. \tag{11}$$

In particular the following equation must be verified:

$$q^c(A_1)q^e(A_1)\frac{\sum_{i=1}^{4}\sum_{j=1}^{4}K_{ij}(A_1) + \sum_{i=1}^{4}\sum_{j=5}^{8}K_{ij}(A_1)}{4} = q^c(A_2)q^e(A_2)\left(K_{B_1B_1}(A_2) + K_{B_1B_2}(A_2)\right). \tag{12}$$

The term $\frac{1}{4}\sum_{i=1}^{4}\sum_{j=1}^{4}K_{ij}(A_1)$ provides the mean dispersion within cells in B_1, and thus should be an approximation of the term $K_{B_1B_1}(A_2)$. Similarly, the term $\frac{1}{4}\sum_{i=1}^{4}\sum_{j=5}^{8}K_{ij}(A_2)$ represents the dispersion from cells in B_1 to cells in B_2, and should be an approximation of $K_{B_1B_1}(A_2)$. Note that these properties descend directly when considering $K_{ij}(A_1) = A_1\,K(d_{ij})$ and $K_{B_1B_2}(A_2) = A_2\,K(d_{B_1B_2})$.

Under this assumption we have that

$$q^c(A_1)q^e(A_1) = q^c(A_2)q^e(A_2)$$

which, under representation in equation 5, yields:

$$A_1^{\alpha_c} A_1^{\alpha_e} = (4A_1)^{\alpha_c}(4A_1)^{\alpha_e}, \implies (4)^{\alpha_c}(4)^{\alpha_e} = 1.$$

Since $\alpha_c \geq 0$ and $\alpha_e \geq 0$, this is possible only if q^c and q^e are area-independent ($\alpha_e = \alpha_c = 0$).

Note that in this example a self-colonization term is considered in the model equations, i.e., $cK_{ii}(A_1)q^c$ in equation 8 and $cK_{B_1B_1}(A_2)q^c$ in equation 9. The approach proposed here shows that this self-colonization term must be considered in metapopulation models when dealing with coarse-graining problems. In fact, this feature is essential because the intra-cellular dispersal mimics the dispersal between the corresponding internal cells at finer resolutions. The model obtained is thus made consistent. i.e. invariant under coarse-graining. Note that including the self-colonization term in the equation was proposed by some authors (e.g. Schnell et al. (2012)), but the role of area when changing the grain size was not considered.

The final model formulation thus reads:

$$\begin{aligned}
\frac{dp}{dt} &= C_i(t)\,(1 - p_i(t)) - E_i(t)p_i(t), && (13)\\
C_i(t) &= c\sum_{j=1}^{n} K_{ij} q_j^c p_j(t), && (14)\\
K_{ij} &= \frac{A_i}{2\pi D^2}\exp\left(-\frac{d_{ij}}{D}\right), && (15)\\
E_i(t) &= e/q_j^e, && (16)
\end{aligned}$$

with the dispersal volume thus dependent on the area of the source pixel, the colonization of the focus pixel including self-colonization, and the extinction independent of the area.

2.4 Model comparison

In the following we want to highlight the importance of considering a model consistent with coarse-graining through numerical simulations. To achieve this goal, we compare the following three models:

- Model **M0**. This model corresponds to the original Hanski's model presented in equation 1, where we set:
  - the quality-properties of the cells equal to the area times a fitness function, $q_i^c = q_i^e = A_i\, f_i$;
  - the coefficients of the kernel discretization are set to be area-independent, $K_{ij} = K(d_{ij})$;
  - the self-colonization terms are not considered ($i \neq j$ in equation 2).
- Model **M1**. This model is a variation of the model presented in equation 13, where self-colonization is not considered. We set:
  - the quality-properties of the cells area-independent, $q_i^c = q_i^e = f_i$;
  - the coefficients of the kernel discretization are set to be area-dependent, $K_{ij} = A_j K(d_{ij})$;
  - the self-colonization terms are not considered ($i \neq j$ in equation 2).
- Model **M2**. This corresponds to the model in equation 13. Differently from **M1**, self-colonization is considered.

2.5 Simulations

To test the consistency of the model under coarse-graining, we simulate the occupancy dynamics of an invading species within an area of about 70×43 km^2 in two landscapes (Figs. 1 and 2):

- Landscape **L1**. A flat domain meant to investigate the consistency of the model with graining when not affected by the heterogeneous landscape matrix ($f_i = 1$). In this landscape we compare results for the three considered models: **M0**, **M1**, and **M2**.

- Landscape **L2**. A domain consisting of a rectangular area surrounding the Gran Paradiso National Park. In this case the quality of the cell expressed in terms of the cell-elevation (used as a proxy of temperature), i.e. f_i is computed using equation 6. Elevation is based on a digital elevation model (DEM) extracted from the Copernicus EU-DEM v1.1 (25 m resolution). In this landscape we compare results for models **M1**, and **M2**.

Both landscapes are considered with grain-sizes of 90 m, 180 m, 360 m, 720 m, 1440 m and 2880.

In landscape **L2**, the original DEM of 90 m resolution is upscaled to the coarser grains by bilinear interpolation, which preserves the mean elevations (Fig. 3). Fig. 2 shows the DEMs at resolutions of 90 m and 1440 m. We can see that the main landscape features are preserved in the coarser grid, however, the available surface at elevations between 1800–2300 m is slightly overestimated (Fig. 3).

The model comparison in the homogeneous domain (landscape **L1**) is meant to investigate the consistency of the dispersal approach with coarse-graining.

In landscape **L2**, the complex GPNP domain, we investigate how the added heterogeneity influences the consistency of the model with scaling. In this case, the process of coarse-graining causes loss of information in the habitat of the virtual species (the quality), which is set dependent on the elevation. This deterioration in the input data, might be more or less relevant in the model results depending on the particular properties characterizing the species, e.g. the dispersion D and the niche width σ.

Table 1: Metapopulation parameters of the species considered in the study

Species	S1	S2	S3	S4	S5	S6
c [y^{-1}]	1					
e [y^{-1}]	0.1					
z_{opt} [m]	150					
σ [m]	150			300		
D [m]	100	500	1000	100	500	1000

The invasion of six virtual species (called S1, . . . , S6) is simulated in both landscapes for a period of 100 years. The species are characterized by different values of the five parameters (c, e, D, z_opt, σ) described in Table 1 (note that σ has no influence of the species in the flat landscape which means only three species are considered there).

3 Results

3.1 Homogeneous landscape (L1)

Fig. 5 depicts the average occupancy over time for three invading species characterized by different values of dispersal. Occupancy results are computed using the three variations of the model and for the six different grain sizes, thus highlighting the different impacts of coarse-graining in the solutions. Differences in the occupancy results are also represented by relative errors with respect the reference solution (dx =90 m). The occupancy results depicted in Fig. 5 highlight that, for the three species considered in this landscape, the original formulation of the model (model **M0**) converges to different steady states depending on the grain size. Instead, the proposed new formulation but without self-colonization (model **M1**) leads to results more stable with coarse-graining, especially for species with large dispersal (species S2 and S3) and small differences in grains.

The self-colonization term considered in model **M2** has a strong positive impact when dealing with large grains. In fact, when dx =1440 m or 2880 m, **M1** is not able to obtain the same occupancy value that is reached with finer grains. This problem, instead, is not present in the results from **M2**, thanks to the added colonization pressure provided by the self-colonization term. This claim is corroborated by the lower relative error associate to **M2** results

M2 results for species S1 also show that very fine grains are necessary to well simulate the dynamics of the colonization process when dealing with species having short dispersal values. In

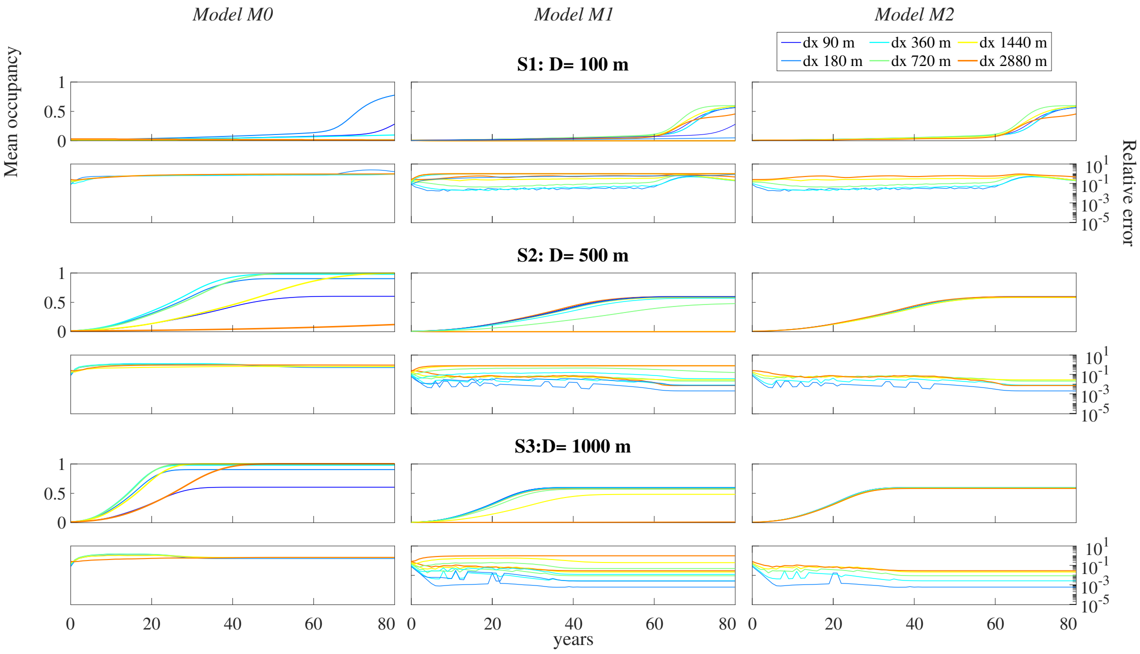

Figure 5: Behavior of the different model formulations under coarse-graining in a flat domain. Landscape **L1**, homogeneous domain. Temporal evolution of the probability of occupancy (averaged over the entire domain) and relative root mean squared error between an occupancy map and the corresponding reference solution obtained at the finest grain size (dx =90 m, dark blue line). Results are depicted for the three different dispersal distances of the species and the six grain sizes. The three columns show the results for the three different models (c.f. section 2.4): (Left) model **M0**; (Center) model **M1**; (Right) model **M2**.

fact, only for this species, **M1** occupancies associated to different grains show different patterns when the colonization process become faster (e.g. due to a larger part of the domain occupied). The relative errors between the occupancy maps also increase during this phase of fast occupancy of the domain.

These results suggest that, in a homogeneous landscape, the proposed metapopulation model **M2** is the most consistent with respect coarse-graining. However, there might still be numerical errors arising for species having short dispersal values (e.g., species S1), For these short dispersal values, fine graining is necessary in order to correctly approximate the transient states. The main reason for the presence of these errors is related to the technique used for the discretization of the dispersal kernel. When the dispersal length is significantly smaller than the cell size, the discrete kernel used in any simulation does not adequately approximate the dispersal process and more accurate discretization techniques should be employed, as discussed for example by Keith et al. (2018).

3.2 Complex GPNP landscape (L2)

Fig. 6 shows the time evolution of occupancy (and its associated error) of model **M1** and **M2** for the six invading species (c.f. table 1) in the GPNP region (**landscape L2**). The difference between **M1** and **M2** clearly shows the effect of considering self-colonization. Considering for instance species S1 or S4, without self-colonization (**M1**), only the most detailed landscape (dx = 90 m) converges to the intended steady state, whereas almost all grain sizes converge to the same steady state for model **M2**. Model **M1** still struggles to reach the same average occupancy for very coarse grains (orange line in Fig. 6, **M2**, S1), but the error is not comparable to what is experienced in model **M1**.

To a certain extent, the observed differences in occupancy in model **M2** can still be related

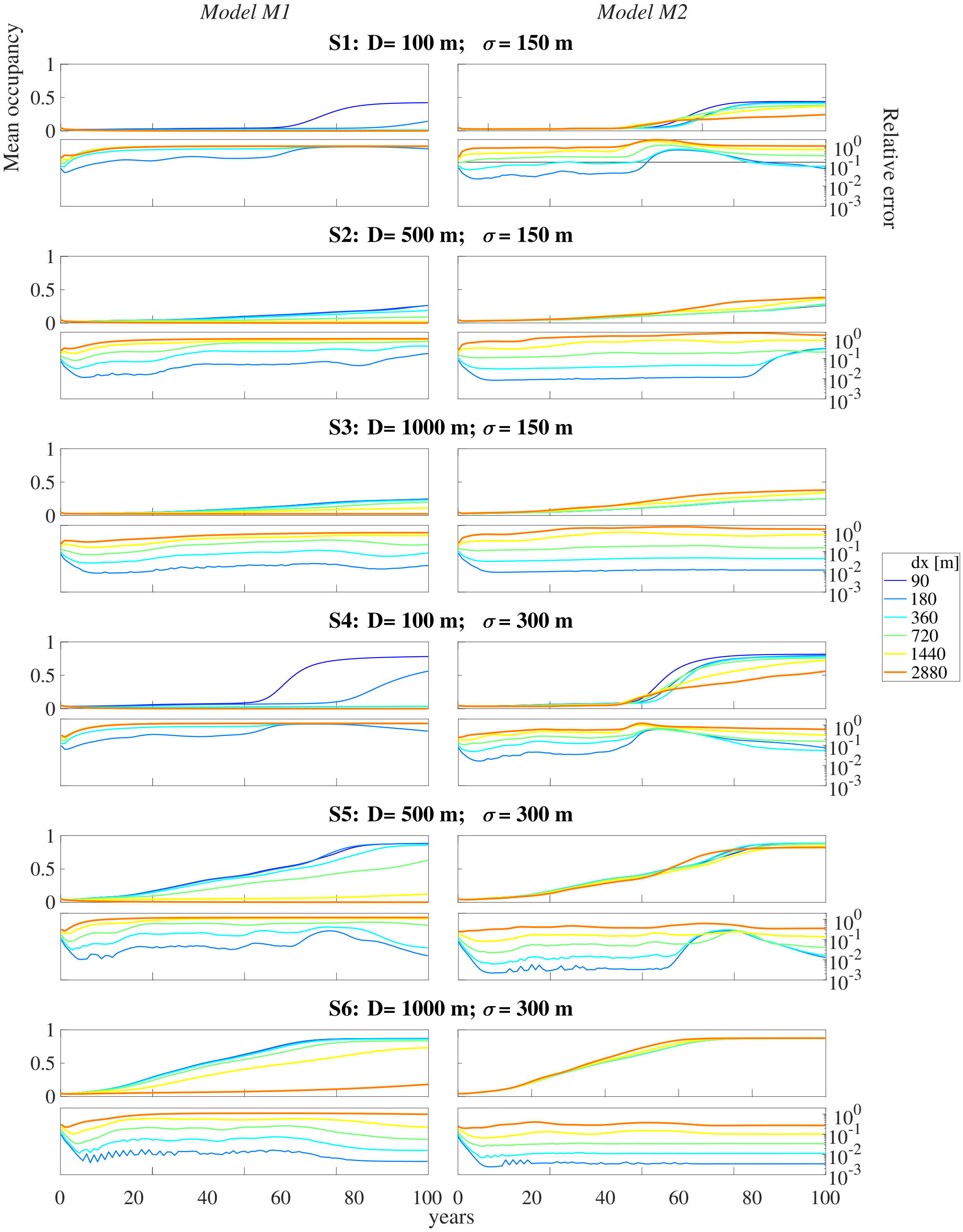

Figure 6: Behavior of the different model formulations under coarse-graining in a the GPNP domain. Landscape **L2**, GPNP domain. Temporal evolution of the probability of occupancy (averaged over the entire domain) and relative root mean squared error between an occupancy map and the corresponding reference solution obtained at the finest grain size (dx =90 m, dark blue line). Results are depicted for the six species and the six grain sizes considered. The two columns show the results for model **M1** (left), and model **M2** (right) (c.f. section 2.4).

to the errors observed above in the flat landscape (error in discretizing the dispersal kernel), but another part is explained by the loss of information on the landscape when coarse-graining it. This is especially visible when comparing species with the same dispersal but different niche width. Considering for instance species S3 and S6: although the dispersal distance is the same for both species, S6 converges to the same value for all considered grains. The difference between the two species is that S6 has a larger niche than S3, and loss in complexity with graining affects it less. Indeed, given the larger niche, species S6 is less sensitive to small losses in suitable area induced by the rescaling, i.e. the averaging of the cell. As such, the impact of graining is highest for species S1, where both dispersal distance and niche width are small, inducing the largest errors in numerical approximation of the dispersal volume and habitat suitability.

Fig. 7 shows an example of the occupancy of an invading species in space and time, and the associated difference.

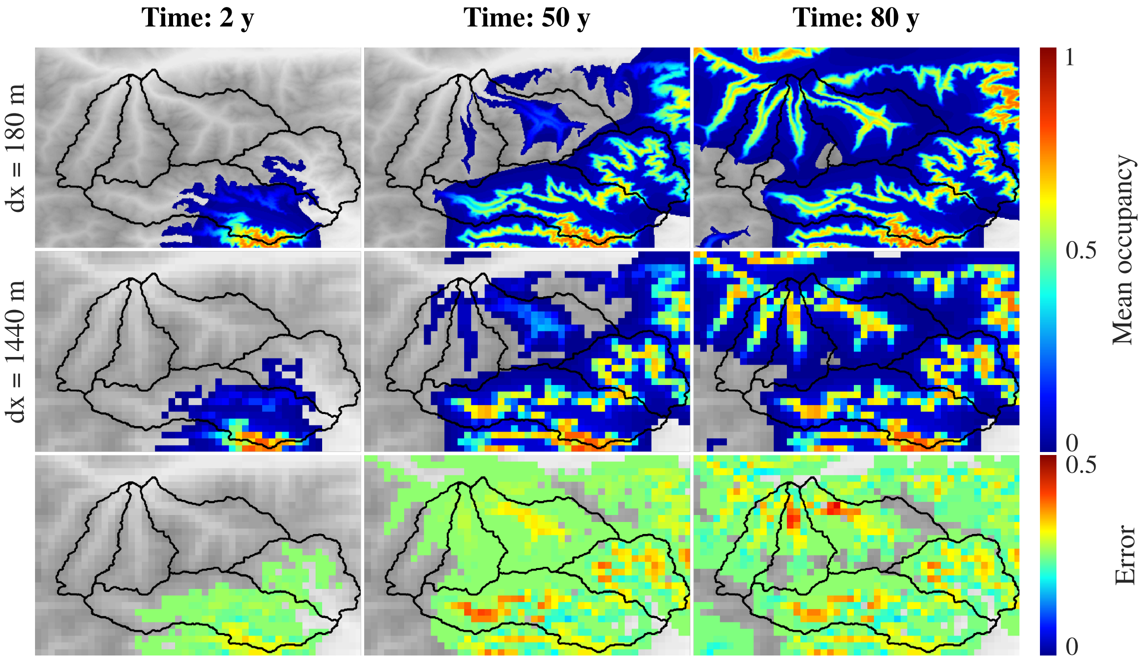

Figure 7: Spatial comparison of probability of occupancy in time. Probability of occupancy for the invading species S2 at three times (2 y, 50 y, 80 y) and for two grains (180 m and 1440 m). The last row shows the error computed when upscaling the occupancy from the 180 m resolution to the coarser resolution.

It is interesting to see that the errors in occupancy all seem to stem from overly complex areas, and mostly from an overestimation of the occupancy of the coarser grid, which is not always the case (c.f. Fig. 6 S4 for instance, where the coarser grid underestimates the occupancy). This overestimation could be due to the fact that the heterogeneous cells are averaged to a value closer to the optimal elevation of the focus species, and the inverse effect could mean that the averaging removes favorable area, although the underestimation of the coarse grid observed in species S1 and S4 could also be linked to the small dispersal value.

4 Discussion

In this work we propose an innovative formulation of the continuous metapopulation model for the simulation of the probability of presence of a focus species in a landscape represented by a regular raster of cells. The proposed model (indicated with **M2** differs from the traditional metapopulation formulation (indicated with **M0**) in three points:

1. The extinction and colonization rates of the species on a cell are independent from the area of the cell, but might depend on the habitat quality of the cell (e.g. elevation).

2. The coefficients of the discretized dispersal kernel are proportional to the cell areas, in order to preserve the integral of the kernel on the domain.

3. A term of self-colonization is introduced in the model equations in order to consider the dispersal processes that occur within the cell.

We showed that these properties naturally descend by asking that the model is consistent with coarse-granining, i.e. that the probability of occupancy of a coarse cell is equal to the average of the probabilities computed on the finer-level cells occupying the same surface.

Additionally, the model corresponding to **M2**, but without self-colonization was indicated as **M1**.

To show the benefits of the new metapopulation formulation when dealing with coarse-graining, Models **M0** and **M2** have been compared using numerical simulations on two landscapes, one depicting a flat domain and one depicting the DEM of the Gran Paradiso National Park.

The results described here show that:

- The proposed model **M2** strongly mitigates the numerical errors introduced in metapopulation models **M0** and **M1** when considering different grain sizes.

- Also when using model **M2**, there exists a threshold on the grid resolution which should not be exceeded. In fact, the discretization of the dispersal kernel on large cells might alter the colonization pressure thus affecting the simulation outcome. This is especially relevant for species having a short dispersal, as observed in Fig. 5 for species S1.

- Self-colonization appears to be an essential tool in reducing the loss in information occurring from coarse-graining, as evidently shown when comparing models **M1** and **M2** in Fig. 6.

- When using coarse grains (dx =1440 or $dx = 2880$) in heterogeneous landscapes, the loss of information in the data describing the cell quality introduces an additional source of error for the metapopulation solution. This error, added to the poor numerical approximation of the colonization pressure, is particularly relevant when species are characterized by short dispersals and small niche widths (see 6, species S1, and S2).

5 Conclusion

Depending on the species of interest, metapopulation models have been applied in the literature at different scales, considering both local and regional domains. However, when dealing with large domain sizes, only a coarse resolution might be adopted for computational limitations, which might corrupt the model results due to the loss of information on the landscape structure and the introduction of numerical errors. This study shows that, by using a metapopulation model with a graining-consistent formulation, it is possible to identify the grain sizes at which the loss on information becomes critical when simulating the species occupancy.

References

Jonathan Giezendanner. *Beyond the patch: on landscape-explicit metapopulation dynamics.* PhD thesis, EPFL, Lausanne, 2020. URL `https://infoscience.epfl.ch/handle/20.500.14299/164718`.

Bernhard Kirchheimer, Christoph C. F. Schinkel, Agnes S. Dellinger, Simone Klatt, Dietmar Moser, Manuela Winkler, Jonathan Lenoir, Marco Caccianiga, Antoine Guisan, Diego Nieto-Lugilde, Jens-Christian Svenning, Wilfried Thuiller, Pascal Vittoz, Wolfgang Willner, Niklaus E. Zimmermann, Elvira Hörandl, and Stefan Dullinger. A matter of scale: apparent niche differentiation of diploid and tetraploid plants may depend on extent and grain of analysis. *Journal of Biogeography*, 43(4):716–726, 4 2016. ISSN 03050270. doi: 10.1111/jbi.12663. URL `http://doi.wiley.com/10.1111/jbi.12663`.

Petr Keil, Jonathan Belmaker, Adam M Wilson, Philip Unitt, and Walter Jetz. Downscaling of species distribution models: A hierarchical approach. *Methods in Ecology and Evolution*, 4(1): 82–94, 2013. ISSN 2041210X. doi: 10.1111/j.2041-210x.2012.00264.x.

David A. Keith, H. Resit Akçakaya, and Nicholas J. Murray. Scaling range sizes to threats for robust predictions of risks to biodiversity. *Conservation Biology*, 32(2):322–332, 2018. ISSN 15231739. doi: 10.1111/cobi.12988.

Ilkka Hanski and Michael Gilpin. Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society*, 42(1-2):3–16, 1991. ISSN 10958312. doi: 10.1111/j.1095-8312.1991.tb00548.x.

J K Schnell, G M Harris, S L Pimm, and G J Russell. Estimating extinction risk with metapopulation models of large-scale fragmentation. *Conservation Biology*, 27(3):520–530, 2012.

Jonathan Giezendanner, Enrico Bertuzzo, Damiano Pasetto, Antoine Guisan, and Andrea Rinaldo. A minimalist model of extinction and range dynamics of virtual mountain species driven by warming temperatures. *PLoS ONE*, 14(3):1–19, 2019. ISSN 19326203. doi: 10.1371/journal.pone.0213775.

Ignacio Rodriguez-Iturbe and Andrea Rinaldo. *Fractal River Basins: Chance and Self-Organization.* Cambridge University Press, 2001.

Robert K. Colwell and George C Hurtt. Nonbiological Gradients in Species Richness and a Spurious Rapoport Effect. *The American Naturalist*, 144(4):570–595, 1994. URL `http://www.jstor.org/stable/2462939` `http://about.jstor.org/terms`.

Robert K. Colwell and David C. Lees. The mid-domain effect: Geometric constraints on the geography of species richness. *Trends in Ecology and Evolution*, 15(2):70–76, 2000. ISSN 01695347. doi: 10.1016/S0169-5347(99)01767-X.

Michael W Palmer and Peter S White. Scale Dependence and the Species-Area Relationship. *The American Naturalist*, 144(5):717–740, 1994.

Ilkka Hanski. Metapopulation dynamics. *Nature*, 396(6706):41–49, 11 1998. ISSN 00280836. doi: 10.1038/23876. URL `http://www.nature.com/doifinder/10.1038/23876`.

Lenore Fahrig and William K. Nuttle. Population Ecology in heterogeneous environments. In *Ecosystem function in heterogeneous landscapes*, chapter 6, pages 95–118. Springer, 2005.

Lenore Fahrig. Landscape heterogeneity and metapopulation dynamics. In Jianguo Wu and Richard J. Hobbs, editors, *Key topics and perspectives in landscape ecology.*, chapter 5, pages 78–89. Cambridge University Press, 2007.

Drew W. Purves, Miguel A. Zavala, Kiona Ogle, Fernando Prieto, and José M. Rey Benayas. Environmental Heterogeneity, Bird-mediated Directed Dispersal, and Oak Woodland Dynamics in Mediterranean Spain. *Ecological Monographs*, 77(1):77–97, 2 2007. doi: 10.1890/05-1923. URL `http://doi.wiley.com/10.1890/05-1923`.

Joel Rybicki and Ilkka Hanski. Species-area relationships and extinctions caused by habitat loss and fragmentation. *Ecology Letters*, 16:27–38, 2013. ISSN 1461023X. doi: 10.1111/ele.12065.

Jonathan Giezendanner, Damiano Pasetto, Javier Perez-Saez, Cristiana Cerrato, Ramona Viterbi, Silvia Terzago, Elisa Palazzi, and Andrea Rinaldo. Earth and field observations underpin metapopulation dynamics in complex landscapes: Near-term study on carabids. *Proceedings of the National Academy of Sciences*, 117(23):12877–12884, 2020. doi: 10.1073/pnas.1919580117. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1919580117`.

J. Giezendanner, P. Benettin, N. Durighetto, G. Botter, and A. Rinaldo. A Note on the Role of Seasonal Expansions and Contractions of the Flowing Fluvial Network on Metapopulation Persistence. 57(11):e2021WR029813, 2021. ISSN 0043-1397, 1944-7973. doi: 10.1029/2021WR029813. URL `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021WR029813`.

Enrico Bertuzzo, Francesco Carrara, Lorenzo Mari, Florian Altermatt, Ignacio Rodriguez-Iturbe, and Andrea Rinaldo. Geomorphic controls on elevational gradients of species richness. *Proceedings of the National Academy of Sciences of the United States of America*, 113(7):1737–1742, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1518922113. URL `http://www.pnas.org/lookup/doi/10.1073/pnas.1518922113`.

Ilkka Hanski and Mats Gyllenberg. Uniting Two General Patterns in the Distribution of Species. *American Association for the Advancement of Science*, 275(5298):397–400, 1997.

Damiano Pasetto, Salvador Arenas-Castro, Javier Bustamante, Renato Casagrandi, Nektarios Chrysoulakis, Anna F. Cord, Andreas Dittrich, Cristina Domingo-Marimon, Ghada El Serafy, Arnon Karnieli, Georgios A. Kordelas, Ioannis Manakos, Lorenzo Mari, Antonio Monteiro, Elisa Palazzi, Dimitris Poursanidis, Andrea Rinaldo, Silvia Terzago, Alex Ziemba, and Guy Ziv. Integration of satellite remote sensing data in ecosystem modelling at local scales: Practices and trends. *Methods in Ecology and Evolution*, 9(8):1810–1821, 2018. ISSN 2041210X. doi: 10.1111/2041-210X.13018.

Silvia Zaoli, Andrea Giometto, Jonathan Giezendanner, Amos Maritan, and Andrea Rinaldo. On the probabilistic nature of the species-area relation. *Journal of Theoretical Biology*, 462:391–407, 2019. ISSN 10958541. doi: 10.1016/j.jtbi.2018.11.032. URL `https://doi.org/10.1016/j.jtbi.2018.11.032`.

Enrico Bertuzzo, Ignacio Rodriguez-Iturbe, and Andrea Rinaldo. Metapopulation capacity of evolving fluvial landscapes. *Water Resources Research*, 51(4):2696–2706, 2015. ISSN 19447973. doi: 10.1002/2015WR016946.